Examining and exploiting docking domain-mediated carrier protein recognition in aeruginosin biosynthesis

Matthew John Beech

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Life Sciences

Supervisors: Prof. Gregory L. Challis and Prof. Józef R. Lewandowski

University of Warwick
Department of Chemistry
March 2021

I. Table of Contents

II. List of Figures

III. List of Schemes

IV. List of Tables

V. List of Abbreviations

Abbreviation	Definition
^{2}J	2-Bond coupling constant
^{3}J	3-Bond coupling constant
4-HPLA	4-Hydroxyphenyllactic acid
4-HPPA	4-Hydroxyphenylpropionic acid
AcOH	Acetic acid
ACP	Acyl carrier protein
AHCCA	(1S,3R,4S)-3-amino-4-hydroxycyclohexane-1-carboxylic acid
Amp	Ampicillin
ATP	Adenosine triophosphate
A.U.	Arbitrary units
B. ambifaria	*Burkholderia ambifaria*
βHDD	β-hairpin docking domain
BLI	Bio-Layer Interferometry
Boc	*tert*-Butoxycarbonyl
BSM	Basal salts medium
CoA	Coenzyme A
COM	Communication-mediating domain
Da	Daltons
DCM	Dichloromethane
DD	Docking domain
DHCCA	(1S,3R,4S)-3,4-dihydroxycyclohexane-1-carboxylic acid
DIPEA	Diisopropylethylamine (Hünig's base)
DMAP	4-Dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid

DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EIC	Extract ion chromatogram
ESI	Electrospray ionisation
EtOAc	Ethyl acetate
FAD	Flavin adenine dinucleotide
$FADH_2$	Reduced flavin adenine dinucleotide
FDA	(United States) Food and Drug Administration
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HOBt	1-Hydroxybenzotriazole
Hz	Hertz
IT-MS	Ion-trap mass spectrometry
Kan	Kanamycin
kb	Kilobases
kDa	Kilodaltons
M. aeruginosa	*Microcystis aeruginosa*
m/z	Mass-to-charge ratio
MeCN	Acetonitrile
MeOH	Methanol
NAC	*N*-acetyl cysteamine
NADH	β-Nicotinamide adenine dinucleotide
NADPH	β-Nicotinamide adenine dinucleotide 2'-phosphate
Nd:YLF	Neodynium-doped yttrium lithium fluoride
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NP	Natural product
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetase

OD	Optical density
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PK	Polyketide
PKS	Polyketide synthase
ppm	Parts per million
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
Q-TOF-MS	Quadripole-time-of-flight mass spectrometry
RT	Room temperature
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SLiM	Short linear motif
THF	Tetrahydrofuran
Tris	Trisaminomethane
UHPLC	Ultra-high-pressure liquid chromatography
UV-Vis	Ultraviolet-visible
WT	Wild-type

VI. Acknowledgements

Firstly, I would like to thank my supervisors Prof. Greg Challis and Prof. Józef Lewandowski for their advice and patience throughout the course of my studies. It has been a privilege to work on such interesting and diverse science and learn from such fantastic mentors.

I would also like to take the opportunity to thank Dr. Matthew Jenner, Dr. Lona Alkhalaf and Dr. Douglas Roberts who have all provided incredibly helpful feedback, valuable expertise and plenty of encouragement throughout the course of my master's and PhD studies. Thanks are also owed to Dr. Lijiang Song and Dr. Ivan Prokes for their helpful discussions regarding mass spectrometry and NMR spectroscopy, respectively.

I would like to thank members of the Challis and Lewandowski groups past and present for making my time here both productive and fun: Daniel Bahia, Dr. Joshua Cartwright, Dr. Maz Costa, Dr. Yousef Dashti, Dr. Trent Franks, Dr. Angelo Gallo, Dr. Richard Gibson, Dr. Daniel Griffiths, Dr. Arun Gupta, Dr. Joleen Masschelein, Ioanna Nakou, Fang Pang, Munro Passmore, Dr. Sławomir Potocki, Dr. Jade Ronan, Jacob Sargeant, Dr. Rakesh Saroay, Dr. Alma Svatoš, Dr. Dani Zabala and Dr. Jinlian Zhao. I am particularly grateful to Dr. Christopher Fage, Dr. Christian Hobson, Dr. Chuan Huang, Dr. Xinyun Jian, Dr. Simone Kosol, Dr. Chris Perry, and Dr. Shanshan Zhou for their incredible guidance and friendship throughout the years.

Thanks are owed to the Midlands Integrative Biosciences Training Partnership (BBSRC) for funding.

Special thanks go to Helen. She has been there to cheer me on when the going has been good, pick up the pieces when it has not and made me smile without fail every single day.

Finally, I would like to thank my Grandad, my Nan and my Dad and Anne, who have supported me throughout.

VII. Declaration

All original experimental work presented in this thesis was conducted by the author, unless otherwise stated in the text, in the Department of Chemistry, University of Warwick between July 2016 and March 2021. No material disclosed has been submitted in support of any other degree or at any other institution. Work by other authors has been referenced throughout the text.

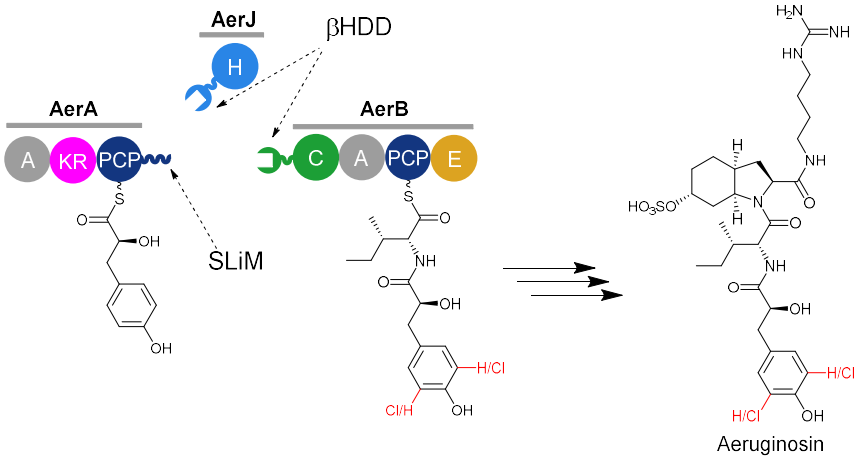
Date: 30/03/21

(Matthew John Beech)

VIII. Abstract

Nonribosomal-peptide synthetases (NRPSs) are multimodular megasynthetases that produce bioactive secondary metabolites in an assembly line-like manner. Selective protein-protein interactions between subunits in NRPS systems ensure biosynthetic fidelity and are typically mediated by recognition between C- and N-terminal docking domains. One example is that between beta-hairpin docking domains (βHDDs) and glutamate-rich short linear motifs (SLiMs).

Recent bioinformatics analysis has uncovered a novel βHDD-SLiM interface in NRPS subunits responsible for the biosynthesis of aeruginosin in *Microcystis aeruginosa* NIES-98 (scheme 1), which produces a mixture of non-, mono- and di-chlorinated aeruginosin congeners. *In vitro* reconstitution of the system has shown the AerA carrier protein domain bearing a C-terminal SLiM can interact productively with two distinct catalytic domains bearing N-terminal βHDDs, appended to the AerB condensation domain and a flavin-dependent halogenase, AerJ. Additionally, it has been demonstrated than competition between the two βHDDs for binding to the carrier protein SLiM is responsible for the mixture of chlorination patterns observed in the natural product.



Scheme 1 Partial domain architecture demonstrating the SLiM-βHDD interface in the initial steps of aeruginosin biosynthesis.

Protein-protein interactions have been shown to be a key factor directing halogenase activity in the aeruginosin biosynthetic pathway. βHDD-bearing enzymes from evolutionarily divergent species can crosstalk productively with AerA(PCP) to generate novel molecules with desirable structural complexity. This work provides a new platform for pathway engineering by harnessing the potential of protein-protein interactions for the recruitment of carrier protein-tethered substrates to direct natural product biosynthesis.

Chapter 1: Introduction

1.1. Microbial natural products as a source of bioactive molecules

Natural products are secondary metabolites originating from plants, animals, bacteria and eukaryotes. They possess remarkable structural diversity and complexity and have been vital in the development of pharmaceuticals throughout the history of medicine. 37.7% of all newly approved drugs in the past forty years are either natural products, natural product derivatives or are inspired by natural product pharmacocores.[1]

Natural products from microbial sources in particular have been indispensable in the pursuit of drug candidates. They possess an extensive range of activities, with microbial natural products comprising many potent antifungal, anticancer, immunosuppressant, anti-inflammatory and antiparasitic agents. Most importantly, however, since the discovery of β-lactam antibiotic penicillin G (**1**) microbial natural products have been central to antimicrobial therapeutics. As of 2016, 69% of all FDA-approved antibiotics were of natural product origin, of which 51% and 46% originate from bacteria and fungi, respectively.[2] Important examples include streptomycin (**2**),[3,4] the first antibiotic used to treat *Mycobacterium tuberculosis*; gentamicin (**3**),[5] deemed by the World Health Organization to be an essential medicine and critical to human life; and vancomycin (**4**),[6] a glycopeptide active against multidrug-resistant pathogens (figure **1.1**).

Despite their prominence in the history of pharmaceuticals, more recently the use of microbial natural products as lead compounds has heavily declined, replaced by synthetic fragment-based high-throughput screening approaches.

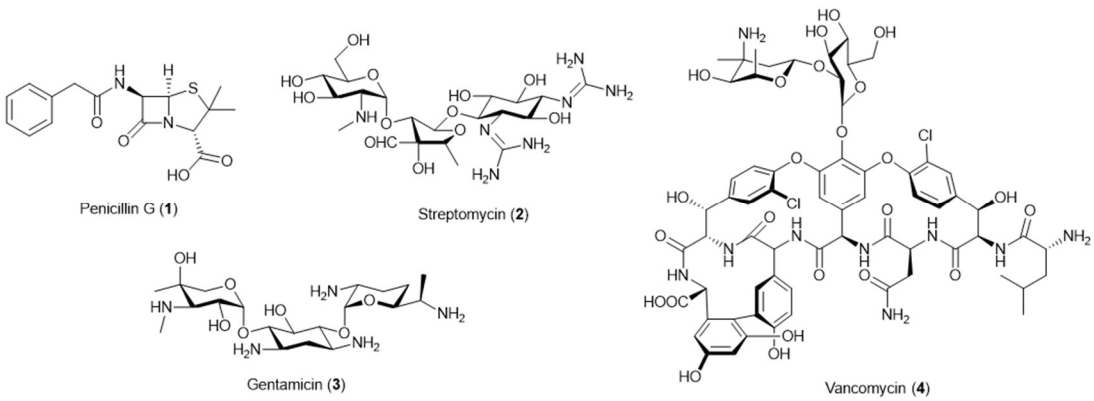


Figure 1.1 Examples of microbial natural products in use as antibiotic treatments.

However, in the face of impending antibiotic and antifungal resistance crises,[7-9] the discovery of new natural products and their analogues with novel chemistry and bioactivities is now more imperative than ever. Renewed interest in microbial natural products can be attributed to recent advances in bioinformatics, which has enabled discovery and utilisation of silent gene clusters to access previously uncharacterised molecules.[10] Similarly, ongoing discovery and characterisation of novel bacterial strains, especially those from marine and other underutilised habitats,[11-14] has expanded the chemical space occupied by natural products. As the number of characterised natural products and their derivatives increases, development of large, diverse bacterial extract and compound libraries will improve the viability of high-throughput approaches.[15]

1.2. Nonribosomal peptide biosynthesis

1.2.1. Nonribosomal peptides

Nonribosomal peptides (NRPs) are a class of natural products produced as secondary metabolites of bacteria and fungi. They exhibit a wide range of biological activities, serving as a source of antimicrobial, anticancer and immunosuppressant compounds with pharmaceutical application (figure 1.2).[12,16] They also comprise a number of relevant antifungal and insecticidal agrochemicals.

Although NRPs remain a valuable source of medicinally-important compounds, their clinical use is inhibited by a number of considerations. Notably, their pharmacokinetic properties are typically outside of the drug design paradigms employed by the pharmaceutical industry. These may include compromised bioavailability, toxicity, stability and metabolism. As a result, discovery of novel NRP derivatives has become paramount in the search for new drugs. However, accessing NRP derivatives *via* total synthesis or semi-synthesis is often challenging owing to the inherent structural complexity of NRPs, which often harbour many stereocentres, high degrees of functionality and sensitive moieties. Developing understanding of the biosynthesis of NRPs, therefore, remains of utmost importance as it provides

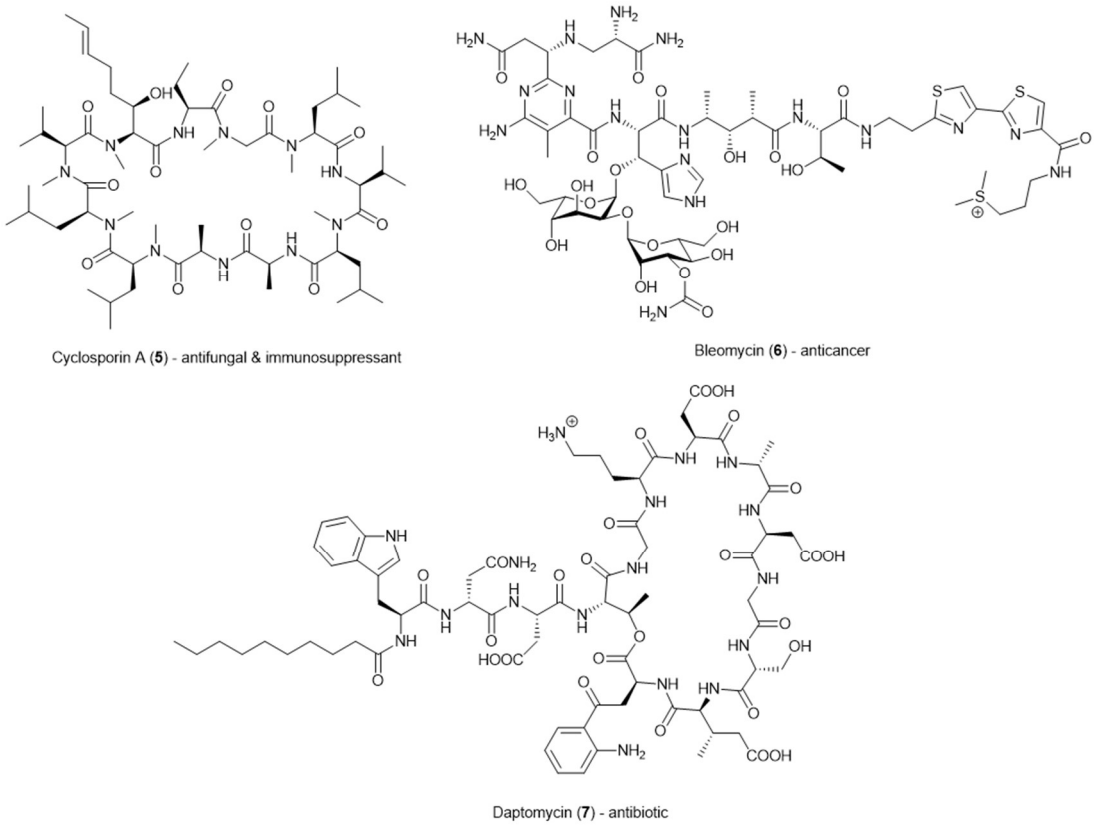


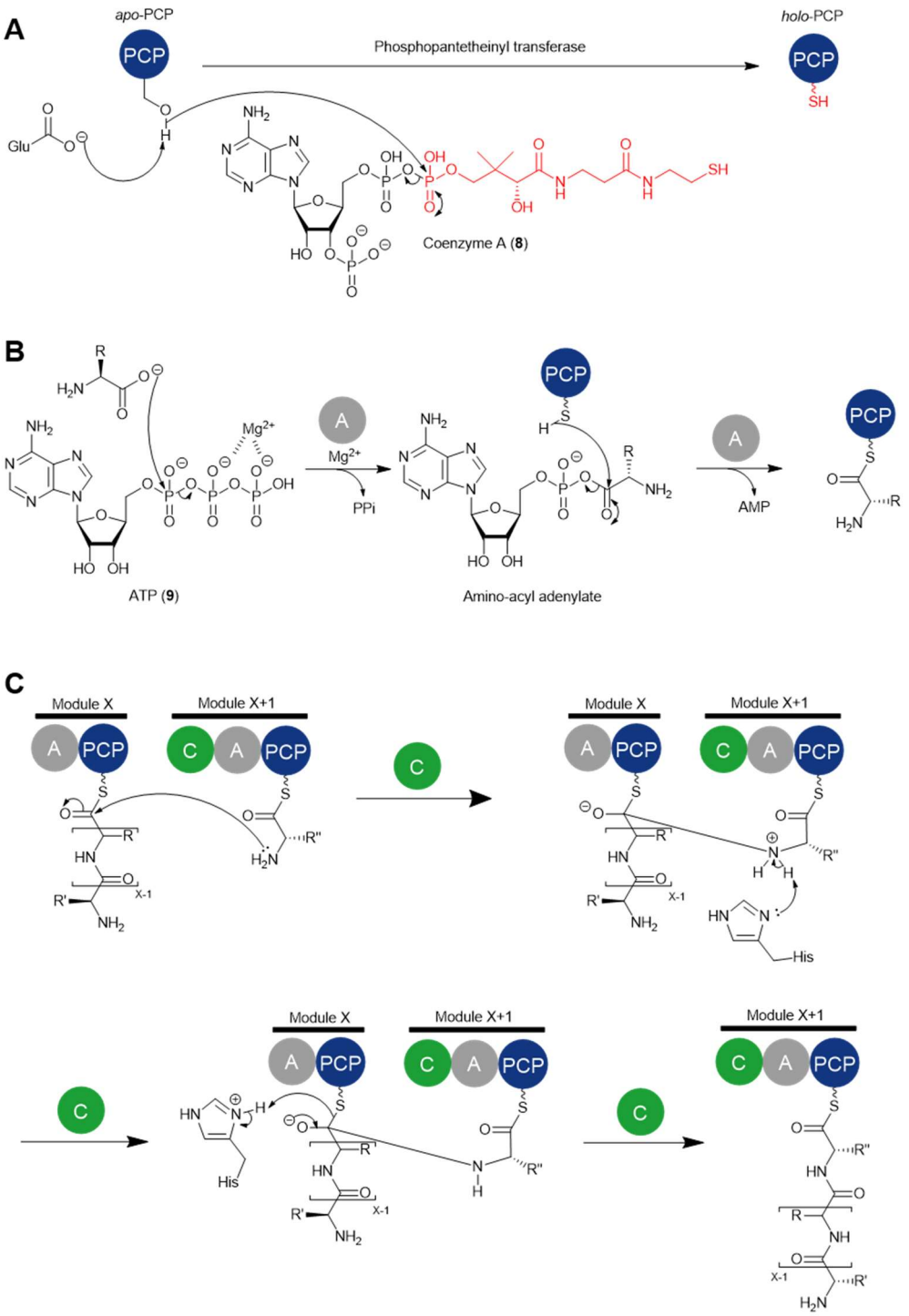
Figure 1.2 Examples of non-ribosomal peptide natural products with pharmaceutical application.

a basis for the rational engineering of their biosynthetic machinery for the development of new NRP analogues.

1.2.2. Nonribsomal peptide synthetases

NRPs are biosynthesised by nonribosomal peptide synthetases (NRPSs). They are large, multi-functional protein complexes that assemble peptides in an assembly line-like manner. In contrast to peptides from ribosomal synthesis, NRPSs are not limited to incorporation of the twenty proteinogenic amino acids, and can integrate D-amino acids, non-proteinogenic amino acids, carboxylic acids and glycosyl groups into the peptidyl product. An NRPS assembly line consists of a number of modules, each of which is usually responsible for incorporation of a single monomer. Each module is comprised of several domains which provide catalytic activity for processing of the amino acid or amino acid-like building blocks.

The canonical architecture of an NRPS module consists of an adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation (C) domain (scheme **1.1**). First, the PCP domain must be post-translationally

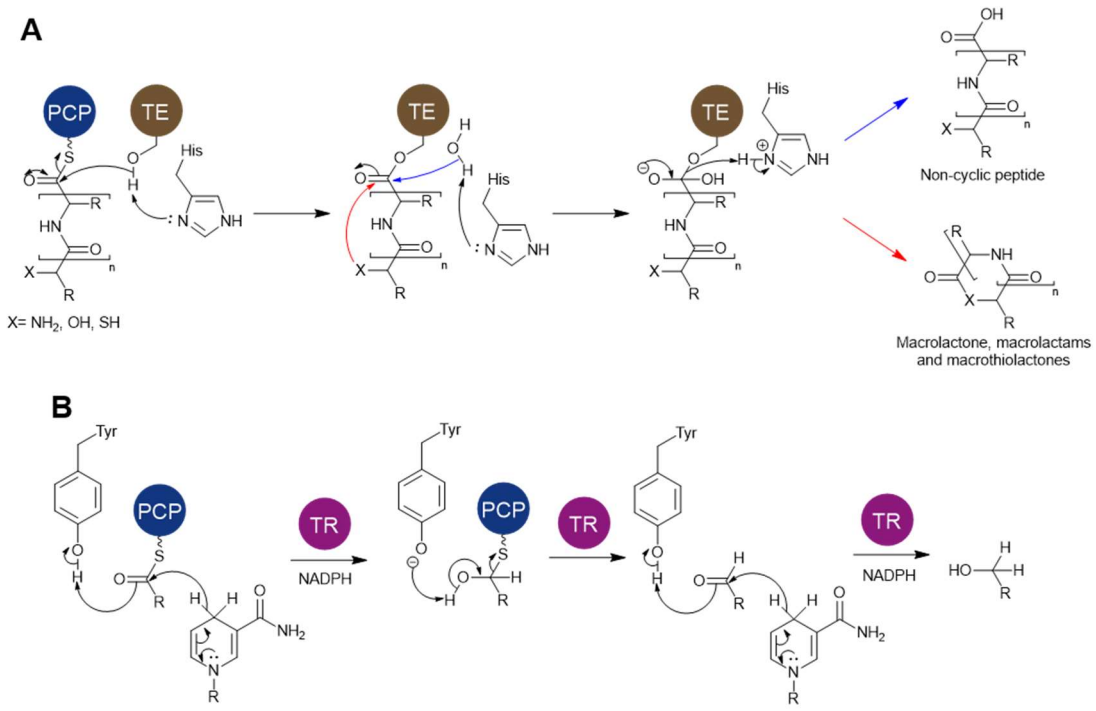


Scheme 1.1 Reaction mechanisms demonstrating: (**A**) phosphopantetheinylation of an *apo*-PCP domain with coenzyme A by a phosphopantetheinyl transferase; (**B**) adenylation of an amino acid and subsequent acylation of a *holo*-PCP catalysed by an adenylation domain; (**C**) condensation domain-catalysed peptide bond formation between the aminoacyl-PCP and the upstream peptidyl-PCP domain. Domain abbreviations: C = condensation domain; A = adenylation domain; PCP = peptidyl carrier protein domain.

modified with a 4'-phosphopantetheine arm from coenzyme A (**8**) by a phosphopantetheinyl transferase (scheme **1.1 A**).[17] This yields the catalytically active *holo*-form of the carrier protein domain. The A domain determines amino acid specificity, recruiting an amino acid building block and converting it to the corresponding aminoacyl-adenylate by reaction with ATP (**9**) (scheme **1.1.B**). 10 amino acid residues in the binding pocket of the A domain, often referred to as the Stachelhaus code, are proposed to confer substrate selectivity.[18] These residues can be used to reliably predict a given A domain's selectivity for common amino acid substrates *in silico*.[19,20] The A domain subsequently catalyses nucleophilic attack of the adenylate by the thiol group of the post-translationally-appended 4'-phosphopantetheinyl prosthetic arm of the downstream PCP domain to generate an aminoacyl thioester. This long, flexible prosthetic arm allows the covalently-tethered substrate to be shuttled between domain active sites within the module. The C domain catalyses selective condensation between the α-amino group of the aminoacyl PCP and the thioester bond of the nascent peptide chain tethered to the PCP domain of the module directly upstream.[21] The second histidine of a conserved HHxxxDG motif found in C domain active sites is thought to be essential to facilitate proton transfers during the reaction (scheme **1.1 C**).[22]

As well as these three core domains, an NRPS must contain a mechanism for product release, typically facilitated by a C-terminal thioesterase (TE) domain or thioester reductase (TR) domain. TE domains catalyse chain release by hydrolysis or macrocyclization of the terminal peptidyl thioester (scheme **1.2 A**),[23-25] whilst TR domains catalyse two- or four-electron reduction of the thioester to aldehydes or alcohols respectively in an NAD(P)H-dependent process (scheme **1.2 B**).[26] There are now also several reported examples of C domain-catalysed condensation with a small molecule acyl acceptor acting as the termination mechanism for NRPS and hybrid PKS-NRPS pathways.[27-29]

In addition to their core domain architectures, NRPSs can also contain optional structure tailoring enzymes, either as embedded domains or acting *in trans*. These domains contribute to the functionalisation of the final NRPS product, conferring new chemical moieties which are often critical for bioactivity. Such

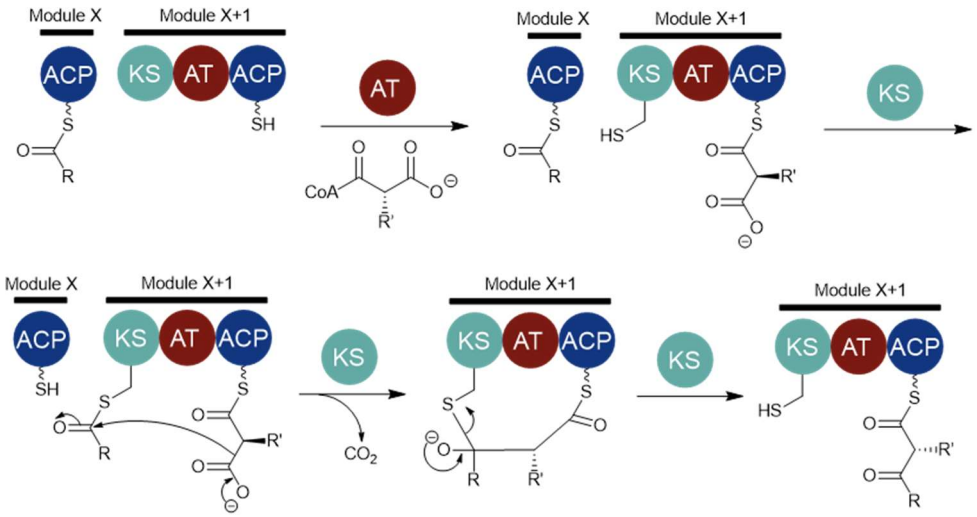


Scheme 1.2 Reaction mechanisms demonstrating chain release of peptide thioesters by: (**A**) thioesterase domain-catalysed hydrolysis or macrocyclisation to generate non-cyclic peptides, macrolactams, macrolactones and macrothiolactones; (**B**) thioreductase domain-catalysed reduction to give *C*-terminal aldehydes and alcohols. Domain abbreviations: TE = thioesterase domain; TR = thioester reductase domain.

tailoring enzymes include: cyclisation domains,[30-32] which replace condensation domains and catalyse thiazoline and oxazoline formation from cysteine and serine or threonine residues respectively; epimerisation domains,[33] which catalyse conversion of L-amino acids to unnatural D-amino acids; *N*-methyl transferases[34]; *C*-methyl transferases[32]; formylation domains[35]; *O*-methyltransferases[36]; oxidases[37]; reductases[38]; halogenases[39,40]; and glycosyl transferases.[41]

1.3. Polyketide biosynthesis

In addition to nonribsomal peptides, bacteria also synthesise polyketide natural products, a diverse class of compounds encompassing numerous potent antibiotics, immunosuppressants, anticancer drugs and agrochemicals. Non-iterative type I polyketide synthases (PKSs) are a class of megasynthases which biosynthesise polyketide secondary metabolites from acyl-CoA building blocks.[42,43] Similar to NRPS assembly lines, PKSs are arranged into modules, where each module is responsible for introduction of one monomeric unit and

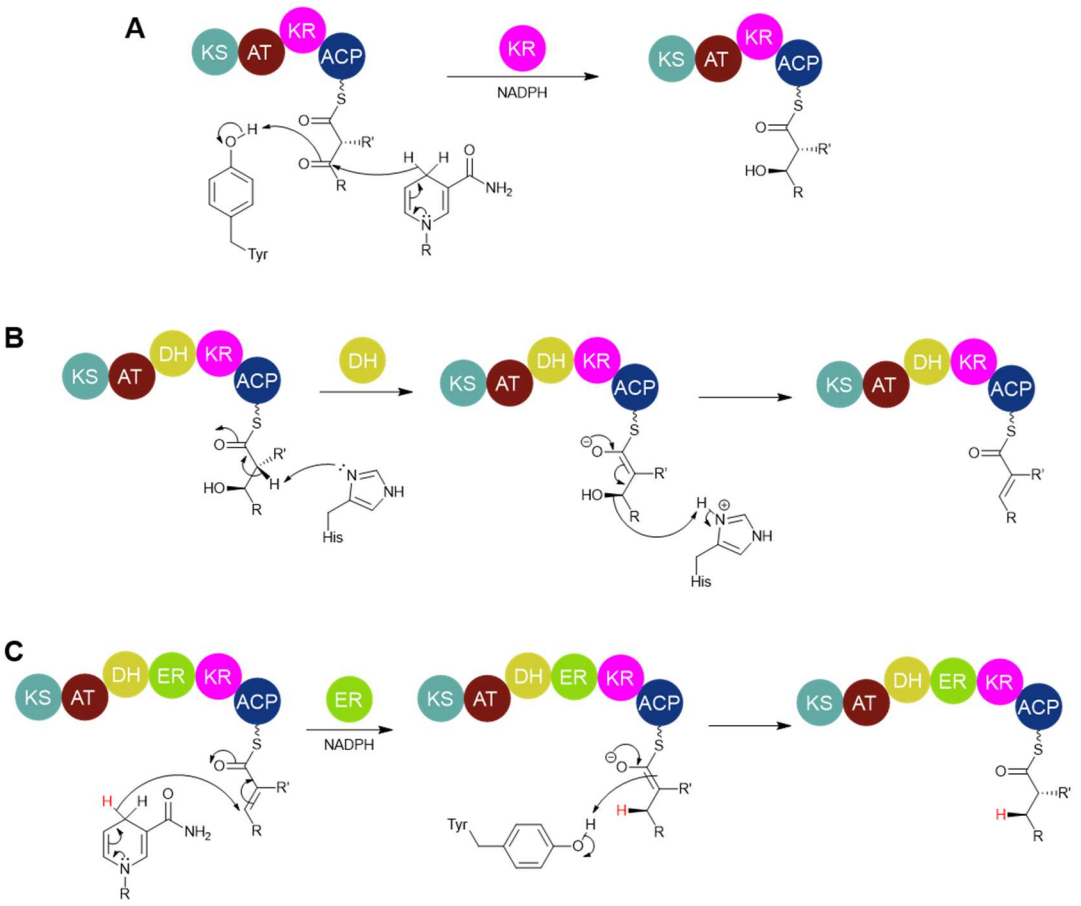


Scheme 1.3 Mechanism of polyketide synthase chain elongation by a minimal PKS module. Domain abbreviations: KS = ketosynthase domain, AT = acyltransferase domain, ACP = acyl carrier protein domain.

consist of several catalytic domains to affect chain elongation and tailoring. Unlike NRPS megasynthetases, however, PKS assembly lines are canonically homodimeric.[44,45]

A minimal modular PKS module consists of an acyl-carrier protein (ACP) domain, an acyltransferase (AT) domain and ketosynthase (KS) domain.[46-48] Much like PCP domains, ACP domains must be post-translationally modified with a 4'-phophopantetheinyl arm to be active.[17] The AT domain catalyses transfer of a malonyl extender unit from malonyl-CoA onto the terminal thiol of the 4'-phosphopantetheine arm. The KS domain harbours an active site cysteine residue which can act as a nucleophile towards the 4'-phosphopantetheinyl arm of the upstream module and be acylated with the growing polyketide chain. The acylated KS can then catalyse a decarboxylative Claisen condensation of the nascent polyketide with the downstream malonyl extender unit (scheme **1.3**).

In addition to the minimal domains required for chain elongation, PKS modules may optionally utilise a cassette of accessory domains to further functionalise the α- and β-carbons of the ACP-bound thioester intermediate.[46,48] Ketoreductase (KR) domains stereoselectively reduce the formed β-ketone to an alcohol utilising an NADPH cofactor (scheme **1.4 A**).[49] Dehydratase (DH)



Scheme 1.4 Reaction mechanisms demonstrating: (**A**) ketoreductase domain-catalysed reduction of a β-keto thioester; (**B**) dehydratase domain-catalysed dehydration of a β-hydroxy thioester; (**C**) enoyl reductase-catalysed reduction of α,β-enoyl thioester. Domain abbreviations: DH = dehydratase domain, KR = ketoreductase domain; ER = enoyl reductase domain.

domains dehydrate the β-alcohol formed by the KR domain to give an α,β-unsaturated intermediate (scheme **1.4 B**).[50] Enoyl reductases reduce the α,β-unsaturated bond generated by the DH domain using NADPH as a cofactor to afford a saturated thioester (scheme **1.4 C**).[42,51] Other common structure-tailoring domains include *C*-methyltransferases,[52] *O*-methyltransferases[53] and pyran synthases.[54] Combinations of these domains present in each module give rise to the large diversity of monomeric units found in polyketide natural products. Like NRPS systems, PKSs require a termination mechanism, usually hydrolysis or macrolactonisation catalysed by a thioesterase domain.[24,55]

Modular PKSs can broadly be grouped into two phylogenetically distinct classes; *cis*- and *trans*-AT. In *cis*-AT PKS systems, an AT domain is found within each module and the modules typically strictly adhere to the co-linearity rule, where the arrangement of domains directly correlates to the structure of

the polyketide product.[47] In *trans*-AT PKS systems, one or more AT domains are encoded on a gene distinct from those of the PKS subunits and provide AT activity *in trans*.[56] Furthermore, trans-AT PKSs can often recruit standalone ER domains *in trans*.[57] As a consequence, *trans*-AT PKS systems usually do not obey the co-linearity principle, and reliable prediction of the products of *trans*-AT biosynthetic gene clusters *in silico* remains challenging.

1.4. Halogenases in natural product biosynthesis

Halogen atoms are common structural features found in a wide range of microbial natural products (figure **1.3**).[58,59] In many examples, halogenation of the natural product scaffold has been demonstrated to be critical for maintaining potent biological activity.[60-62] Halogenation is provided by a class of enzymes known as halogenases. These enzymes most commonly install chlorine or bromine atoms[63,64] but can also catalyse fluorinations[65] and iodinations.[66] Furthermore, halogenases exhibit extraordinary regio- and stereochemical control, and can functionalise C-H bonds that are not easily activated by synthetic methodologies. As a consequence, halogenases are of great interest as biocatalysts for chemoenzymatic reactions and for the semi-synthesis of natural products.[67-69]

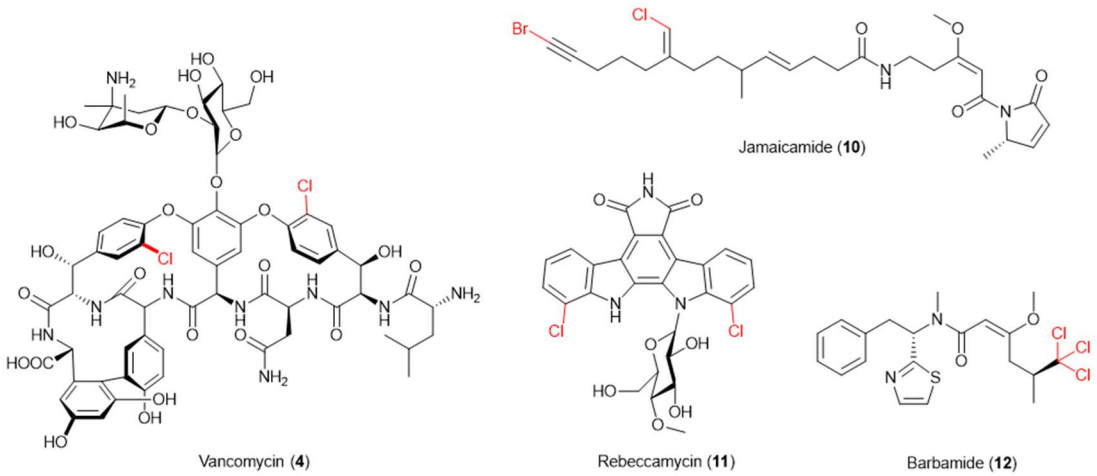
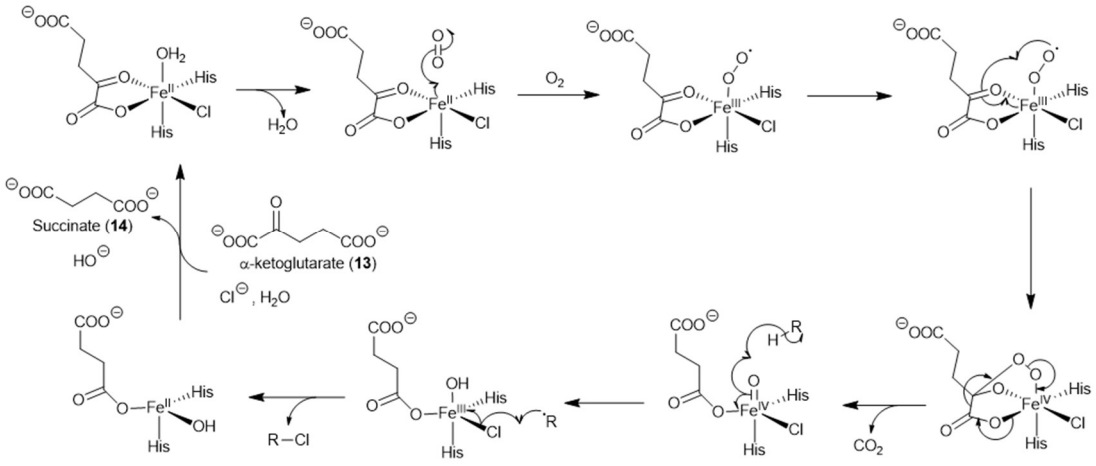


Figure 1.3 Examples of halogenated microbial natural products.

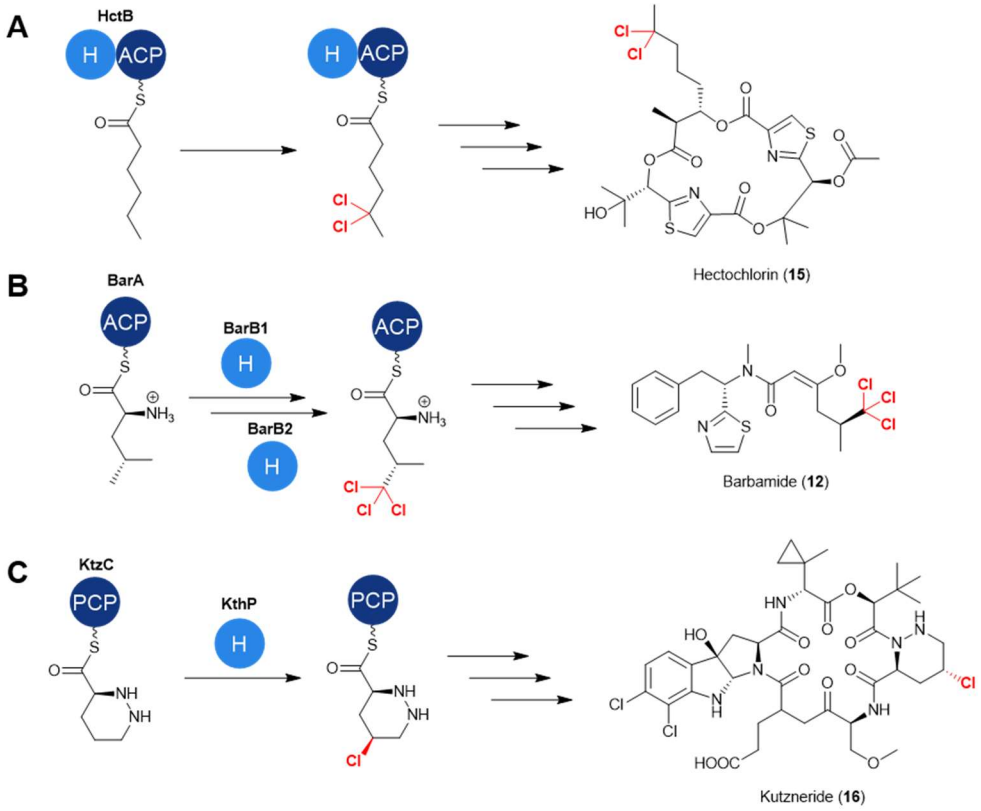
1.4.1. Non-heme Fe(II)-α-ketoglutarate-dependent halogenases

α-Ketoglutarate-dependent halogenases utilise a non-heme Fe(II) prosthetic group to install halogens, typically chlorine, at inactivated carbon centres. They



Scheme 1.5 Catalytic cycle of non-heme Fe(II)-α-ketoglutarate-dependent halogenases.

are highly analogous to α-ketoglutarate-dependent hydroxylases, and can even be converted to hydroxylases by mutagenesis and vice versa.[70-72] Mechanistically, they are proposed to be similar to their hydroxylase counterparts, with conversion of an α-ketoglutarate-bound Fe(II) species to an Fe(IV)-oxido species *via* binding of molecular oxygen and decarboxylation of α-ketoglutarate (**13**).[70] The Fe(IV)-oxido complex can then catalyse



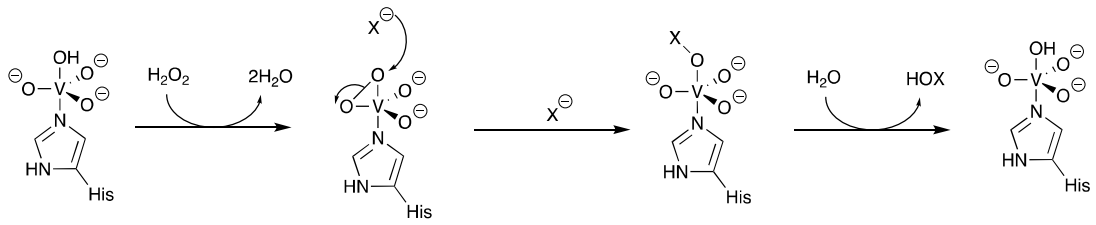
Scheme 1.6 Example reactions catalysed by non-heme Fe(II)-α-ketoglutarate-dependent halogenases in the biosyntheses of natural products (**A**) hectochlorin, (**B**) barbamide and (**C**) kutzneride. H = halogenase.

abstraction of a proton from the substrate by a radical mechanism. Reaction of the formed substrate radical with Fe-bound chloride generates the halogenated species. Loss of succinate (**14**) and binding of chloride regenerates the Fe(II) complex for further catalytic cycles (scheme **1.5**). It is indeterminate exactly how hydroxylase activity is inhibited relative to halogenation, however, proximity and orientation of the chloride relative to the substrate likely imparts chemoselectivity.

α-Ketoglutarate-dependent halogenases are known to act primarily upon carrier protein-tethered substrates, often on amino acid-derived side chains in NRPS biosyntheses,[73-78] although limited examples of enzymes that halogenate PKS[79] and non-megasynth(et)ase substrates[80] are known (scheme **1.6**). Their narrow substrate scope limits their use as traditional biocatalysts. Moreover, these halogenases are not practical for use *in vitro* reactions as they require inert conditions to prevent degradative autooxidation of the Fe(II) cofactor. However, their apparent preference for carrier protein-tethered amino acids and ability to chlorinate synthetically intractable carbon centres make them attractive candidates for biosynthetic engineering of NRPSs.

1.4.2. Haloperoxidases

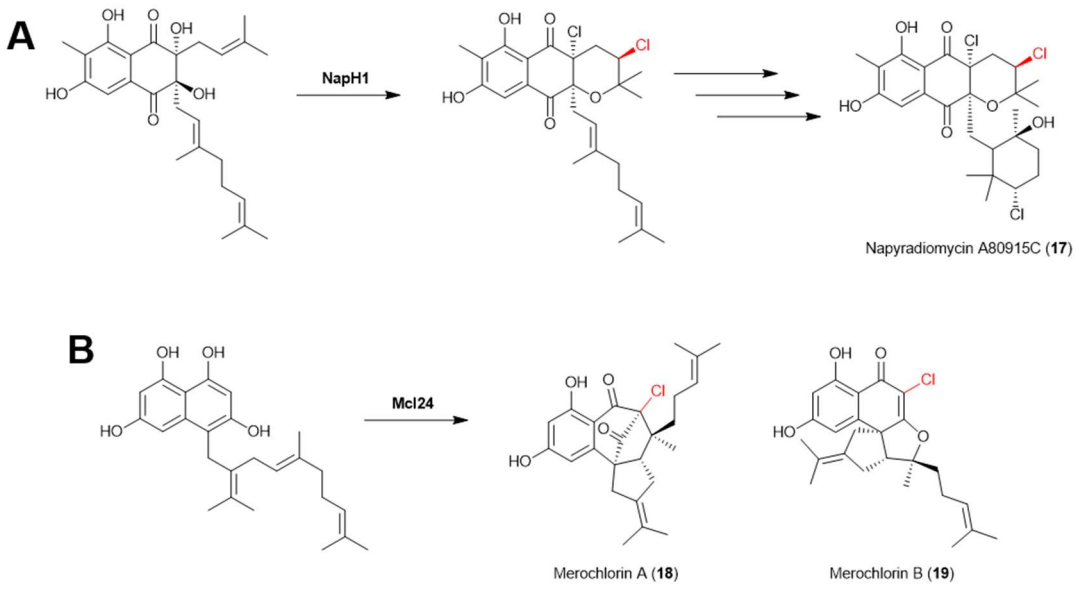
Haloperoxidases are metalloenzymes that utilise hydrogen peroxide and halide ions to generate hypohalous acids as an electrophilic source of halogen. They are known to activate chloride, bromide and iodide in this manner, and can accept a wide range of electron-rich substrates. Haloperoxidases can be divided into two classes defined by the metal cofactor they bind and utilise for catalysis; either a heme or vanadate prosthetic group. Bacterial heme-dependent haloperoxidases are known, however the scope of their involvement in secondary metabolite production remains limited.[81]



Scheme 1.7 The proposed mechanism of halogenation by vanadium dependent haloperoxidases.

Vanadium-dependent haloperoxidases (VHPOs), conversely, have been implicated and characterised in a number of microorganisms, most recently in marine bacteria. They are structurally homologous to bacterial acid phosphatases, which bind organophosphates to catalyse phosphate ester bond hydrolysis.[82] Their vanadate-deficient *apo*- forms have been shown to harbour mild phosphatase activity and their haloperoxidase activity can be inhibited in the presence of high phosphate concentrations.[82,83] Their proposed catalytic mechanism proceeds *via* oxidation of the active site-bound vanadate by hydrogen peroxide. Nucleophilic attack of a halide ion then generates hypohalous acid as a halogenating agent (scheme **1.7**).[83-85]

Several VHPOs have been identified in the biosynthesis of bacterial terpenoid natural products. VHPO NapH1 from the biosynthesis of terpene antibiotic napyradiomycin A80915C (**17**) in *Streptomyces* sp. CNQ-525 catalyses a stereoselective chlorination reaction.[86] Similarly, biosynthesis of merochlorins A (**18**) and B (**19**) in *Streptomyces* sp. CNH-189 utilise VHPO Mcl24 to chlorinate a napthalenyl intermediate.[87] In both cases, the VHPO chlorination is proposed to promote an oxidative cyclisation reaction (scheme **1.8**). The potential of VHPOs to introduce chlorine, bromine and iodine to wide range of electron-rich substrates and their stability towards oxidative inactivation makes them appealing candidates as prospective industrial biotechnology catalysts.



Scheme 1.8 Reactions catalysed by vanadium-dependent haloperoxidases: (**A**) NapH1 in the biosynthesis of Napyradiomycin A80915C; (**B**) Mcl24 in the biosynthesis of merochlorins A and B.

1.4.3. Flavin-Dependent Halogenases

Flavin-dependent halogenases (FDHs) are a class of two-component monooxygenase enzymes. They halogenate electron-rich carbon centres *via* generation of electrophilic halogen sources utilising the redox chemistry of the flavin cofactor. Similar to two-component monooxygenases, they utilise a reduced flavin adenine dinucleotide ($FADH_2$, **24**) cofactor, which is provided from flavin adenine dinucleotide (FAD, **23**) by a flavin-reductase partner enzyme. Typical substrates are pyrroles, phenols and indole moieties (figure **1.4**).[88-90]

Crystal structures of FDHs have revealed that the flavin-binding and substrate-binding pockets are distinct and connected by an approximate 10 Å tunnel

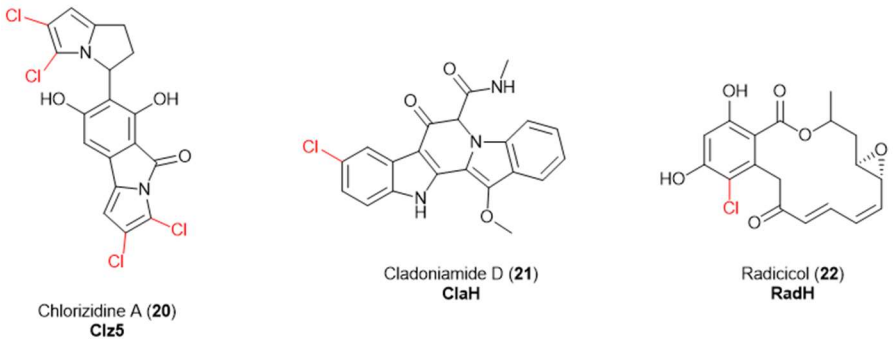


Figure 1.4 Examples of chlorinated pyrrole, indole and phenol-containing natural products and the flavin-dependent halogenases responsible for their incorporation.

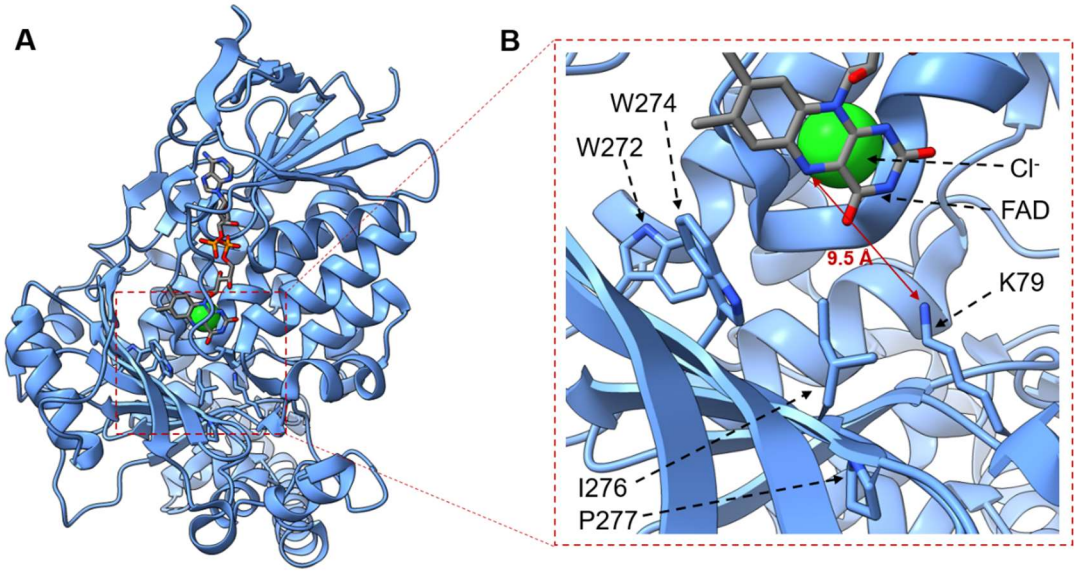
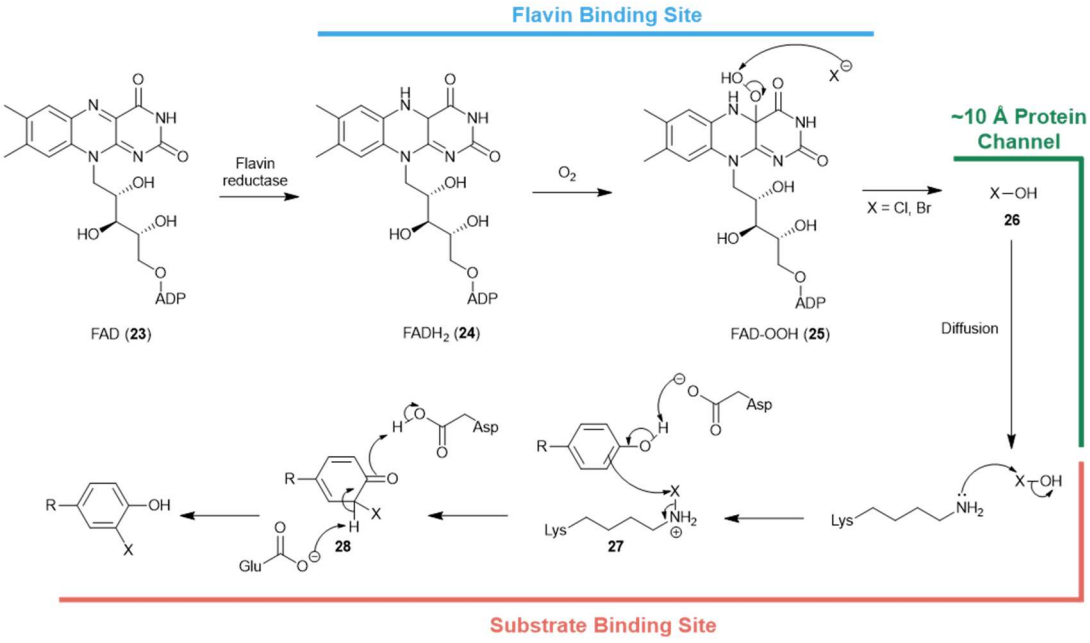


Figure 1.5 Crystal structure of tryptophan flavin-dependent halogenase, PrnA (PDB: 2ARD). (**A**) Overall crystal structure of the halogenase. (**B**) Positioning of the FAD isoalloxazine, chloride ion, conserved WxWxIP motif and catalytic lysine residue in the active site tunnel.

(figure **1.5**).[91] This is in contrast to monooxygenases, where the substrate is suggested to react directly with the peroxyflavin species. Monooxygenase activity is postulated to be prevented by a conserved WxWxIP motif positioned on one face of the substrate tunnel, which prevents proximal binding of the substrate to the cofactor.[91,92] Based on these observations, the current hypothesised mechanism for FDH catalysis involves the generation of hypohalous acid by nucleophilic attack of a chloride ion on peroxyflavin (**25**) (scheme **1.9**). This is proposed to diffuse through a 10 Å protein channel to the substrate binding site where chlorination can occur. A conserved serine residue is suggested to be implicated in forming a hydrogen bond network facilitating hypohalous acid formation and diffusion,[91,93,94] and its presence has been shown to have implication for activity from mutagenesis studies.[94,95]

A conserved lysine residue in the substrate binding site has been identified from structural and mutagenesis studies to be essential for halogenase activity.[90,91,96] Further to this, Walsh and co-workers demonstrated that the rebeccamycin halogenase RebH retained activity following priming with reduced flavin cofactor and subsequent desalting, suggesting a long-lived reactive intermediate.[96] These combined experiments have provided evidence for reaction of hypohalous acid with the conserved lysine residue to generate



Scheme 1.9 Mechanism of halogenation of a phenolic compound by a flavin-dependent halogenase. The putative sites of the protein where each reaction occurs is highlighted.

a more stable chloramine adduct (**27**) (scheme **1.9**). It is unclear whether the lysine chloramine adduct directly acts as a source of electrophilic chlorine or as a storage mechanism from which hypohalous acid can be regenerated as the reactive species.[93] In either case, positioning of the substrate relative to this catalytic lysine residue likely controls the regioselectivity of the flavin-dependent halogenase. Upon substrate binding, an electrophilic aromatic substitution mechanism can occur with the electrophilic chlorine source, where the Wheland intermediate (**28**) can be rearomatized through deprotonation by a conserved glutamate or serine residue (scheme **1.9**).[91,97] FDHs that halogenate phenol substrates possess a conserved aspartate residue positioned in close proximity to the phenolic proton. This is likely to be involved in deprotonation of the phenol to further activate the substrate towards electrophilic aromatic substitution.[90]

1.5. Docking domains in megasynth(et)ases

A single megasynth(et)ase assembly line typically consists of several protein subunits which canonically must communicate with their appropriate partners in a specific linear order to ensure biosynthetic fidelity. Selective protein-protein interactions between C- and N-terminal docking domains (DDs) have commonly been found to be critical in mediating inter-subunit communication to facilitate substrate transfer between successive modules found on separate proteins.[98–102] Docking domains have become an increasingly relevant area of study, as their ability to mediate interactions between protein subunits could make them powerful tools towards novel biosynthetic pathway engineering approaches.

1.5.1. Polyketide Synthase Docking Domains

Polyketide synthase docking domains can broadly be defined by the PKS system they originate from; either *cis*-AT or *trans*-AT. There have been two identified classes of docking domains from *cis*-AT PKSs, denoted class 1 and class 2. Class 1 PKS docking domains are found in actinobacterial PKS

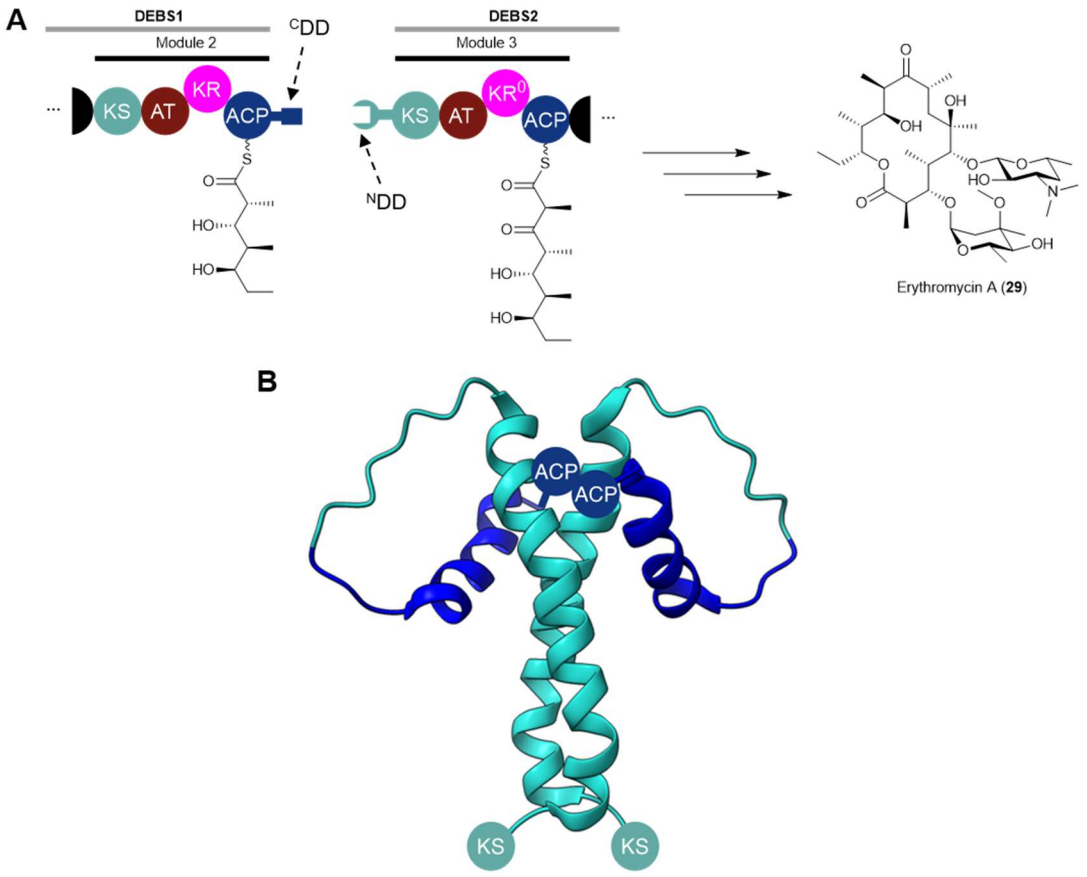


Figure 1.6 The class 1 docking domain-mediated interface between module 2 and module 3 of the DEBS polyketide synthase responsible for the biosynthesis of erythromycin A. (**A**) Partial domain architecture of the DEBS PKS highlighting the docking domain interface. (**B**) The dimeric solution-state NMR structure of fused docking domains of DEBS2 and DEBS3 (PDB: 1PZR).

assembly lines at inter-subunit ACP-KS domain boundaries. They were first discovered at the interface between DEBS modules 2 and 3 from erythromycin A (**29**) biosynthesis in *Saccharopolyspora erythraea* (figure **1.6 A**),[99,103,104] but have since been characterised at the PikAIII-PikAIV modular junction in the pikromycin PKS from *Streptomyces venezuelae* and the Bamb_5920-Bamb_5919 junction from enacyloxin IIa biosynthesis in *Burkholderia ambifaria*.[45,105,106] NMR and crystallographic structures have determined that class 1 PKS docking domains constitute an α-helix at the N-terminus of the KS domain, which acts as an interface facilitating both homodimeric assembly and inter-subunit communication with short C-terminal α-helices tethered to the ACP domains of the upstream module homodimer to create a four-helix bundle (figure **1.6 B**).[103,105] Recognition between helices has been demonstrated from docking domain-exchange and point mutation experiments to be mostly

dependent upon compatible electrostatic residues at the helix-helix interface.[107,108] Khosla, Cane and co-workers established these domains as biosynthetic engineering tools, using the DEBS2-DEBS3 DD pair to direct biosynthesis in 13 unique chimeric polyketide synthases.[109] Docking domains of this class from bioinformatics analysis appear to be ubiquitous in actinobacterial PKSs,[110-112] but characterisation of further examples remains underexplored.

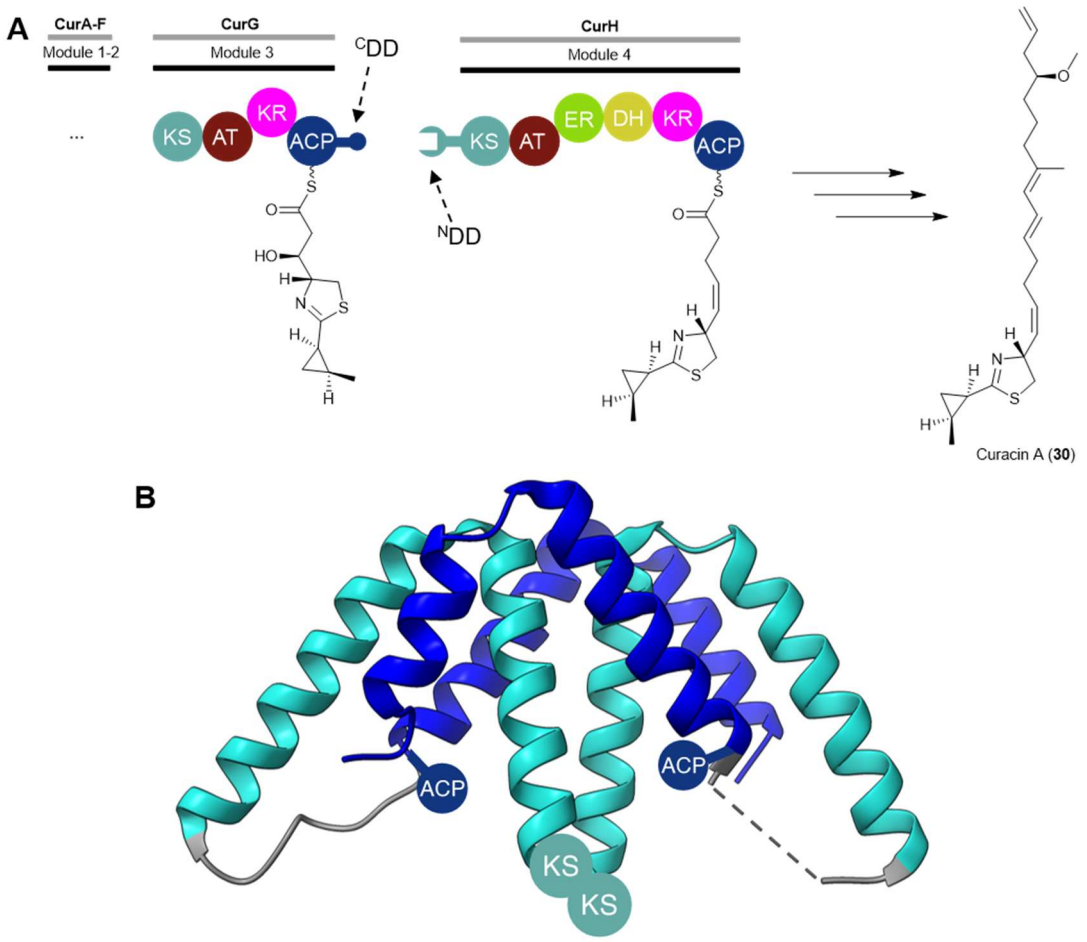


Figure 1.7 The class 2 docking domain-mediated interface between module 3 and module 4 of the polyketide synthase responsible for curacin A biosynthesis. (**A**) Partial domain architecture of the curacin PKS highlighting the class 2 DD interface. (**B**) The 1.7 Å resolution crystal structure of the fused class 2docking domains found at the CurG-CurH interface (PDB: 4MYY). The grey loop represents a synthetic $(GGGS)_2$ linker used to tether the docking domains.

Class 2 docking domains have been identified in cyanobacterial type I PKS systems. They were first elucidated at the CurG-CurH and CurK-CurL interfaces in the curacin A (**30**) biosynthetic pathway from *Moorea producens* (figure **1.7 A**).[102] Like class 1 docking domains, class 2 interfaces are found at ACP-KS domain interfaces. The N-terminal docking domain (^{N}DD) of each KS

domain monomer consists of a two α-helices, the first of which forms a dimerization interface with another KS domain ^{N}DD *via* formation of a coiled-coil motif (figure **1.7 B**). The two helices of the ^{C}DDs appended to the upstream ACP domain can then interact with the coiled-coil motif through a combination of hydrophobic and selectivity-conferring electrostatic interactions.[102] Replacement of the aforementioned class 1 domains in the PikAIII-PikAIV junction with class 2 domains was shown to maintain a productive interaction between the module, highlighting their portability and potential to be used for combinatorial biosynthesis. Moreover, recent strategies using these domains to promote recognition between non-cognate subunits from different assembly lines have shown promise in the *de novo* generation of assembly lines.[113]

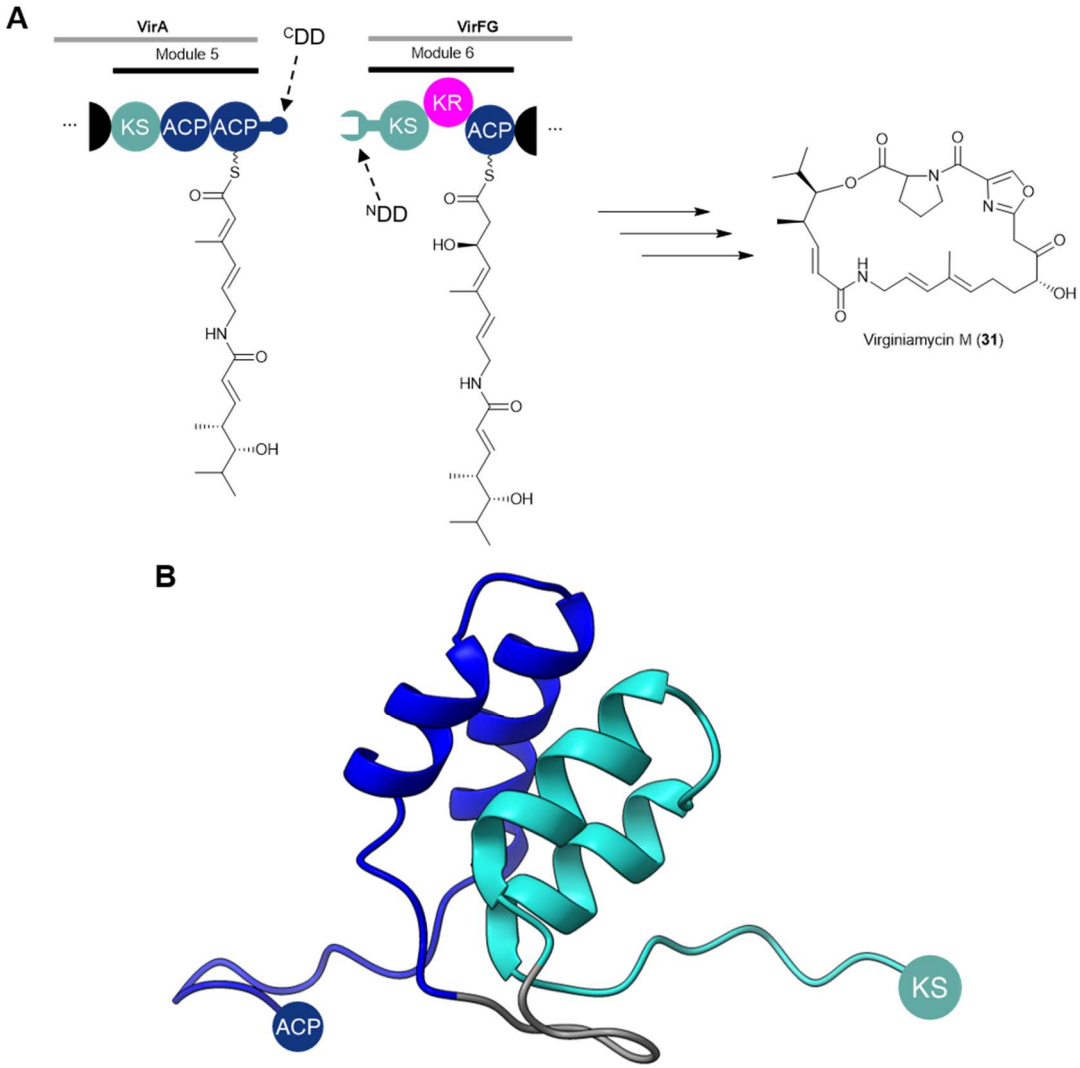


Figure 1.8 The four-helix-bundle docking domain-mediated interface between module 5 and module 6 of the trans-AT polyketide synthase responsible for virginiamycin M biosynthesis. (**A**) Partial domain architecture of the viginiamycin PKS. (**B**) NMR solution structure of fused docking domains found at the VirA-VirFG interface. The grey loop represents a $(GGGS)_2$ linker used to tether the docking domains.

Trans-AT PKS docking domains remain underexplored in comparison to their *cis*-AT counterparts. Nevertheless, some docking domain paradigms have been uncovered. The NMR solution structure of the ACP-KS intermodular subunit boundary between VirA and VirFG from the viginiamycin M (**31**) PKS from *Streptomyces virginiae* revealed the first *trans*-AT docking domain archetype (figure **1.8 A**). Here, monomeric docking domains at the N-terminus of VirA and C-terminus of VirFG each adopt helix-turn-helix motifs in complex to form a four-helix bundle (figure **1.8 B**).[114] Unlike class I docking domains from *cis*-AT PKSs, these docking domains do not act as dimerization elements. A small-angle X-ray scattering (SAXS) structure of module 5 of VirA revealed that there are two distinct copies of the ^{C}DD in the homodimer,[115] suggesting two separate docking domain interfaces are formed. These docking domains have been identified by bioinformatics analysis to occur at a number of intersubunit boundaries, mediating interactions between KS-DH, KS-KR, DH-KR and ACP-ER domain boundaries as well as ACP-C domain boundaries in hybrid PKS-NRPS biosynthetic systems.[116] They appear to form highly specific interaction interfaces, with selectivity putatively conferred by pairwise interaction of charged residues.[114] Only recognition between cognate docking domains pairs have thus far been demonstrated.[116]

A new *trans*-AT docking domain has also been identified at a KS-DH intramodular subunit interface in the gladiolin (**32**) PKS from *Burkholderia gladioli* (figure **1.9**).[117] This domain, coined a dehydratase-docking (DHD) domain, consists of a long, highly-disordered region appended to the N-terminus of the KS domain. This interacts directly with residues on the surface of the cognate DH domain partner.[118] Although uncertainty remains regarding the specificity-conferring residues and the structure formed in complex, from bioinformatics analysis DHD domains appear to be ubiquitous amongst KS-DH intersubunit boundaries in *trans*-AT PKSs. Therefore, they may make unique handles for biosynthetic engineering.[118]

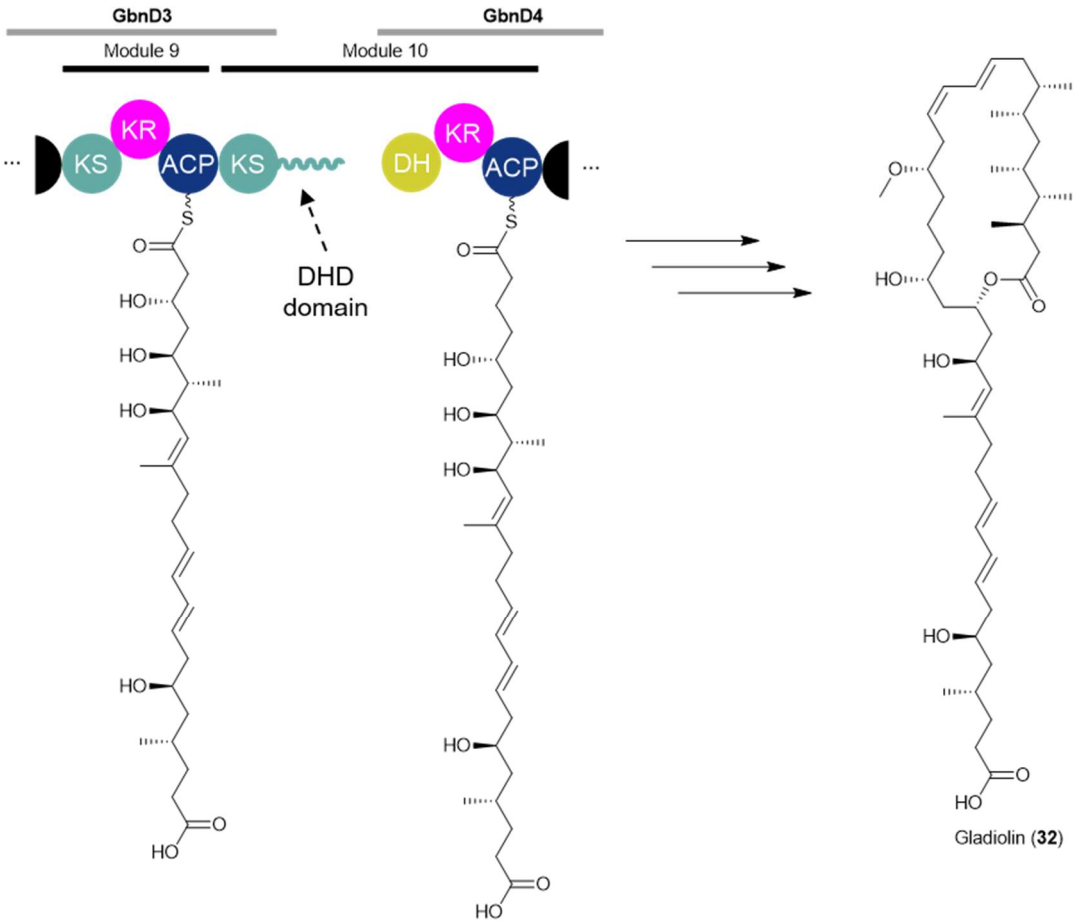


Figure 1.9 The intramodular dehydratase-docking domain interface in module 10 of the polyketide synthase responsible for gladiolin biosynthesis.

1.5.2. Non-Ribosomal Peptide Synthetase Docking Domains

1.5.2.1. COM domains

The first characterised class of docking domains from NRPS systems were the so-called communication-mediating (COM) domains. These were initially discovered at TycA-TycB and TycB-TycC E-C domain intersubunit interfaces from the biosynthesis of tyrocidine A (**33**) in *Bacillus brevis* (figure **1.10 A**).[100,119] Hahn and Stachelhaus demonstrated through *in vitro* and *in vivo* characterisation of N- and C-terminal docking domain deletion mutants that COM docking domains are necessary for productive protein-protein interactions.[100]

The structure of a COM domain pair in complex has not yet been published. In lieu of this, a crystal structure of the SrfA-C subunit from surfactin A biosynthesis was found to have an N-terminal COM domain interacting with its

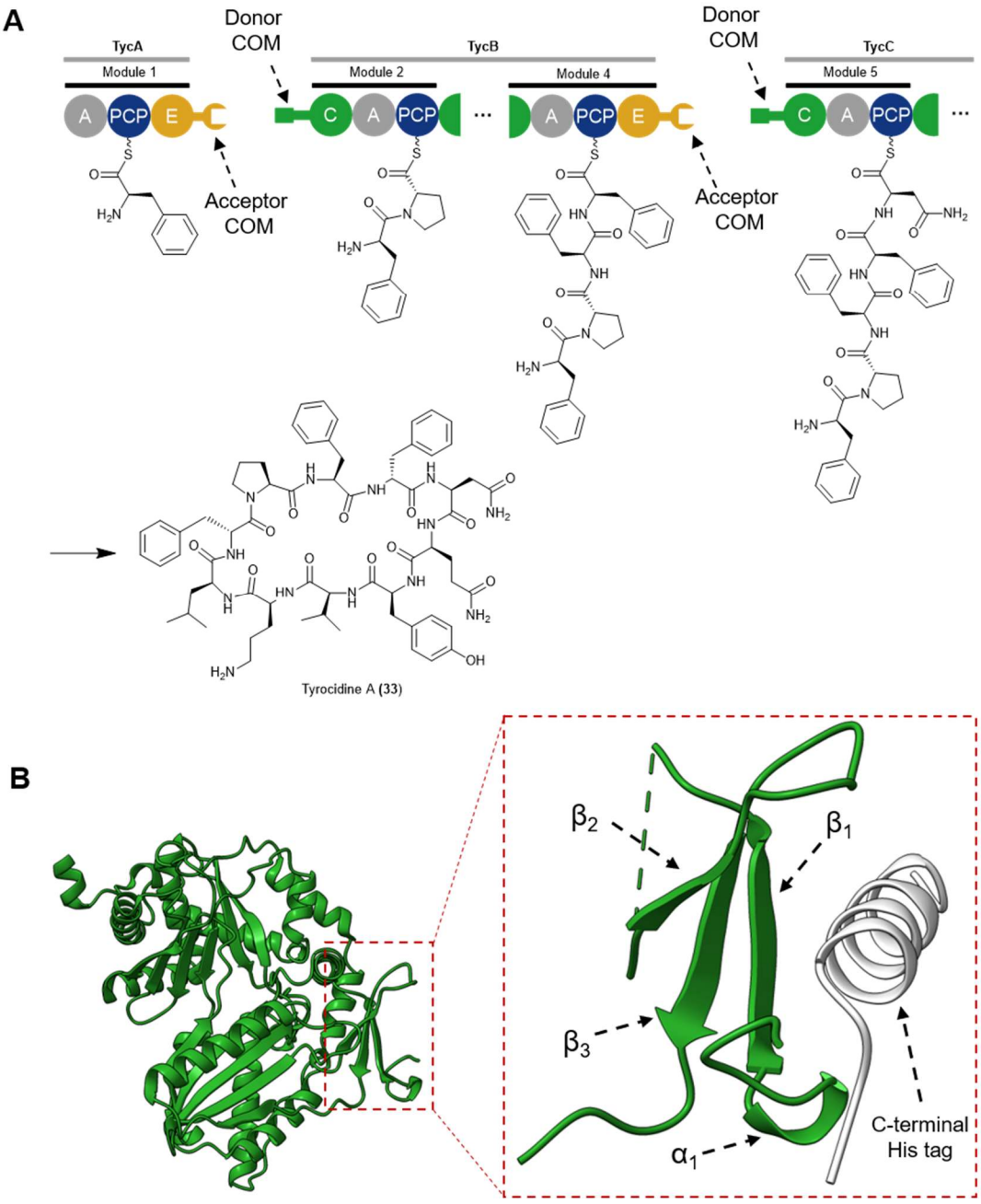


Figure 1.10 (**A**) The biosynthesis of tyrocidine A highlighting the TycA-TycB and TycB-TycC COM domain interfaces. (**B**) The X-ray crystal structure of the SrfA-C condensation domain bearing an N-terminal COM domain (PDB accession: 2VSQ). *Inset*: structure of the N-terminal COM domain in complex with the SrfA-C constructs His-tag.

own C-terminal His-tag (figure **1.10 B**). This His-tag shares moderate sequence similarity to the C-terminal COM domain from the SrfA-B subunit and was found to form a helix on crystallisation.[120] Interestingly, the N-terminal COM domain was found to encompass both the proposed N-terminal portion of the condensation domain (α_1 and β_1), but also two additional beta strands

on the surface of the catalytic domain (β_2.and β_3). These form a pocket into which the His-tag helix docks, termed a 'helix-hand' structure.

Recent studies of COM domain interfaces have focused upon their use as combinatorial biosynthesis tools. Stachelhaus and co-workers demonstrated reordering of subunit processing of the surfactin biosynthetic system *in vivo* by COM domain swapping for modulation of the lipopeptide product spectrum.[121] More recently, reprogramming of the lipopeptide product spectrum of the plipastatin biosynthetic pathway from *Bacillus subtilis* using COM domain engineering strategies has been described.[122] Although more elaborate rational engineering with COM domains remains elusive, these initial studies highlight their potential as tools for bioengineering. Further structural validation of the key interacting residues between donor and acceptor COM domains will be vital to facilitate further efforts to engineer megasynth(et)ase assembly lines.

1.5.2.2. Beta-Hairpin Docking Domains

A novel archetype of NRPS docking is the interaction between beta-hairpin docking domains (βHDDs) and glutamate-rich short linear motifs (SLiMs). SLiMs are typically found at the C-terminus of a carrier protein domain, whilst the corresponding βHDD is found at the N-terminus of the downstream subunit, typically appended to an elongation-catalysing condensation or cyclisation domain. However, putative examples of βHDDs appended to TE, TR, N-methytransferases and halogenases have been identified in hidden Markov model searches of the MiBIG and GenBank databases.[123]

1.5.2.2.1. Structures of β-hairpin docking domains

The first βHDD to be structurally characterised was the TubC βHDD found at in the TubB-TubC subunit boundary in the hybrid PKS-NRPS biosynthetic system of tubulysins (**34**) from *Angiococcus disciformis* An d48 (figure **1.11 A**).[124,125] Richter *et al*. determined the first structure of a βHDD by solution-state NMR spectroscopy of the excised TubC domain βHDD (figure **1.11 B**).

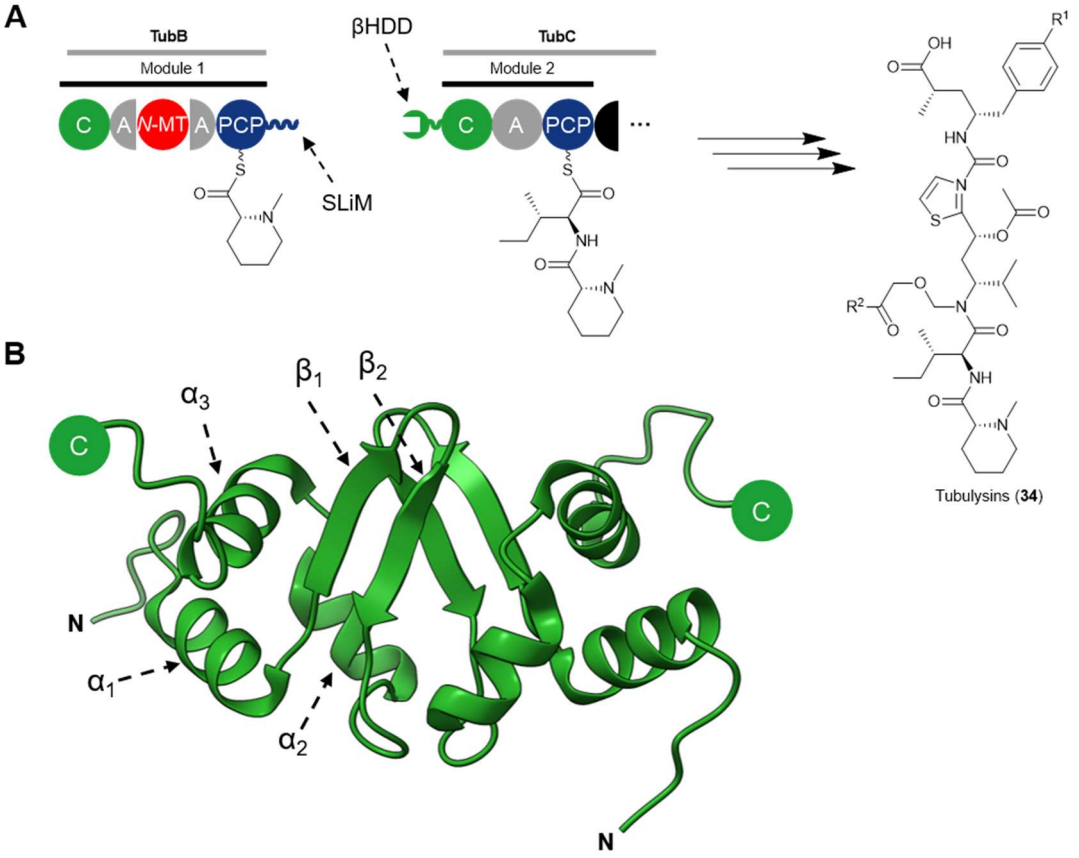


Figure 1.11 (**A**) Partial domain architecture of the tubulysin biosynthetic pathway highlighting the TubB-TubC SLiM-BHDD interface. (**B**) The NMR solution structure of the homodimer of the TubC-βHDD (PDB: 2JUG), displaying the conserved αααββα fold.

This revealed that the βHDD adopts an αββαα fold.[101] This structure has proved to be conserved across all subsequently reported βHDDs, with high structural homology maintained irrespective of the degree of sequence homology. From a bioinformatics perspective, therefore, it is productive to consider predicted secondary structure as the key determinant of the identification of a βHDD rather than high sequence similarity. However, despite the overall topology of each βHDD being highly conserved, the spatial arrangement of the secondary structural elements is somewhat variable, exemplified by the 4.7 Å α-carbon RMSD between the TubC βHDD structure and that determined for the βHDDs found in the rhabdopeptide-producing NRPS in *Xenorhabdus stockiae* Kj12.1.[126] The excised TubC βHDD structure was also found to be homodimeric, although this dimerisation now appears to be an anomaly as it has not been observed in subsequent structural studies of other βHDDs, which have all been found to be monomeric.[123,126,127]

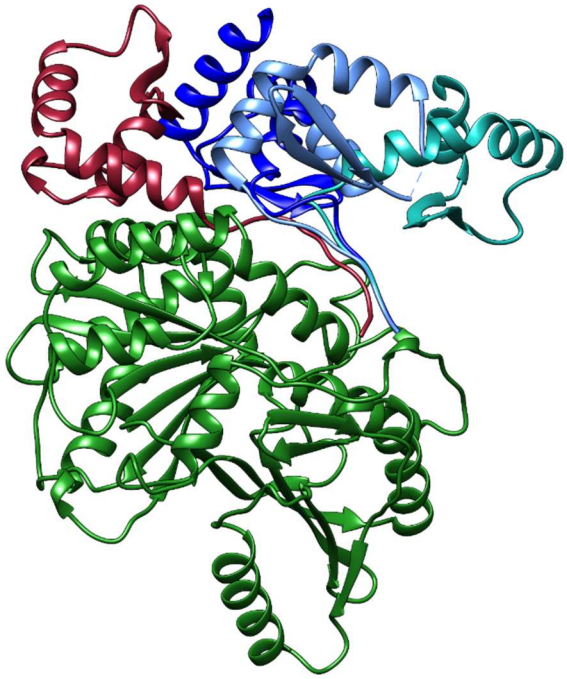


Figure 1.12 Four conformations of the βHDD relative to the catalytic domain identified in crystal structures of the EpoB cyclisation domain and Bamb_5915 C domain (PDB: 5T81, 5T7Z and 6CGO). The EpoB Cy catalytic domain is depicted in green, whilst the observed βHDD conformations are depicted in shades of blue (EpoB) and red (Bamb_5915).

Homodimerisation of the TubC βHDD construct is therefore likely an artefact of domain excision and is not representative of the NRPS subunit, which canonically are monomeric.[128]

The first crystal structure of a βHDD was provided by Dowling *et al* at the *N*-terminus of the EpoB Cy domain from epithilone biosynthesis.[127] Notably, in different asymmetric units of the crystal structure the βHDD was found in distinct conformations with respect to the catalytic domain whilst retaining a rigid structure (figure **1.12**). This indicates an innate flexibility of the linker between the catalytic domain and the βHDD, allowing the βHDD to mobilise around the surface of the Cy domain. Intrinsically disordered linkers between the catalytic domain and βHDD appear to be a conserved structural feature across known proteins, exemplified by the crystal structure of the standalone C domain Bamb_5915 from enacyloxin IIa (**35**) biosynthesis which is found in an additional conformation relative to the catalytic domain.[123] Motion of the βHDD may be highly significant in permitting the βHDD to sample the space around protein. This increases the effective available surface area of the

binding site of the βHDD, and ultimately may be critical in facilitating the recognition of a compatible SLiM partner and promoting carrier protein recruitment.

1.5.2.2.2. Understanding SLiM-β -hairpin docking domain binding

Residue level understanding of the interactions mediating docking domain binding will be critical to their progression as biosynthetic tools. To this end, several studies have provided valuable information allowing the development of a docking paradigm for future engineering of SLiM-βHDD interactions.

Low micromolar affinity binding of the excised TubC βHDD and TubB SLiM was demonstrated by complementary surface plasmon resonance (SPR), isothermal titration calorimetry (ITC) and solution-state NMR titration experiments. NMR titration studies also revealed key residues implicated in interface formation, displaying large chemical shifts for the residues of the positively-charged β2 strand. Corroborating this, mutation of two positively-charged arginine residues on the β2 strand to neutral alanine residues was shown to significantly reduce the binding affinity between the two excised DDs. Single mutations of these two residues increased the K_D by approximately 3- and 10-fold, providing preliminary evidence for the critical role of electrostatic interactions in SLiM-βHDD binding. The working hypothesis arising from this work was that SLiM-βHDD docking is predicated on a specific 'electrostatic code', which may be engineered to tune docking affinities.

Recently Hacker *et al.* published the NMR solution structure of a covalently-tethered SLiM-βHDD construct using docking domains from rhabdopeptide-producing NRPS in *Xenorhabdus stockiae* KJ12 (figure **1.13 A-B**).[126] By covalently tethering the SLiM-βHDD pair, the structure they form in complex could be determined. Interestingly, only the five amino acid at the terminus of the SLiM appear to bind to the βHDD, and they lie in the groove between the β2 strand and α2 helix. From TalosN calculations, it was predicted that these terminal amino acids adopt a β-sheet secondary structure in complex which orients antiparallel to the β2 strand. Charged glutamate and arginine residues on the face of the β2 strand of the βHDD were identified to contribute to salt

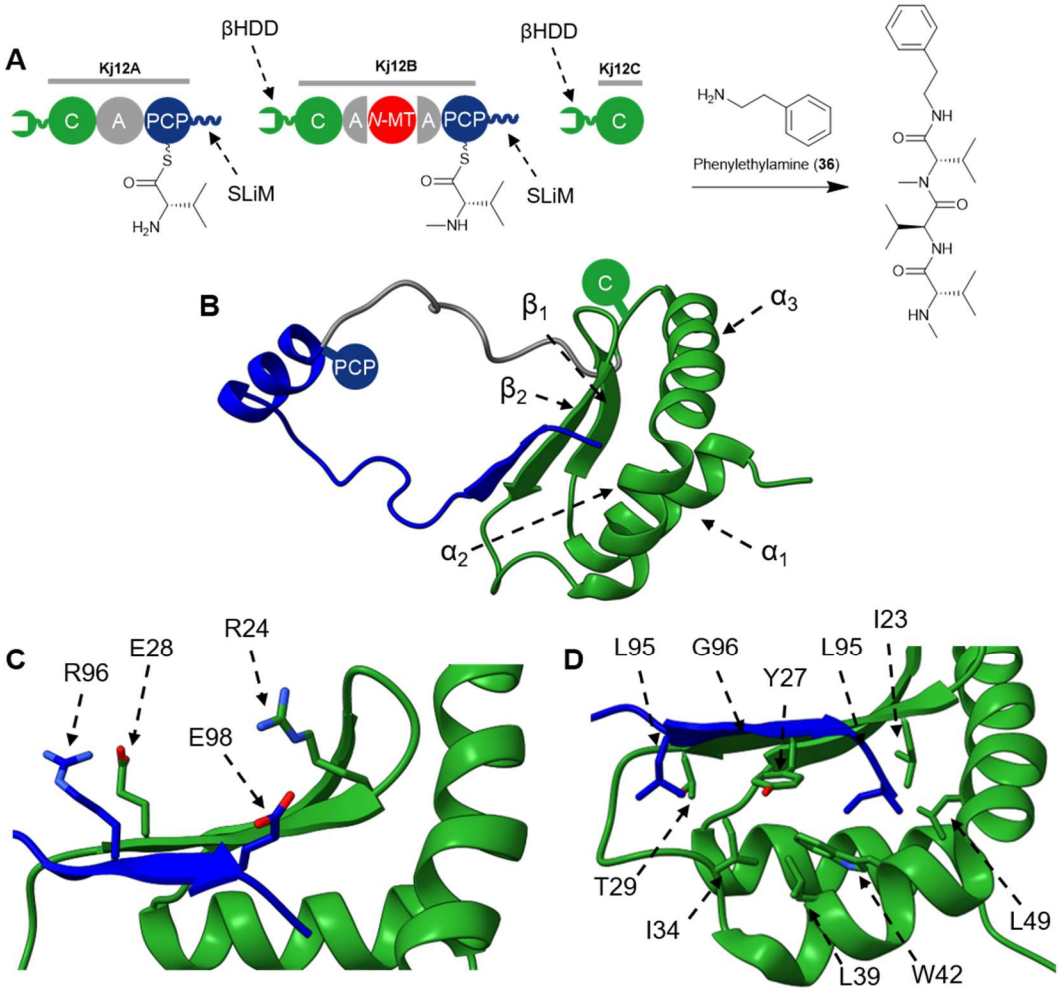


Figure 1.13 (**A**) Domain architecture of the rhabdopeptide-producing NRPS from *Xenorhabdus stockiae* KJ12.1, displaying three βHDDs and two SLiMs. An example rhabdopeptide produced by this system is provided. (**B**) The NMR solution structure of a covalently-tethered SLiM-βHDD construct incorporating the Kj12B SLiM and Kj12C βHDD (PDB accession code: 6EWV). The grey loop represents a $(GS)_6$ linker used to link the docking domain. (**C**) Salt bridges between glutamate and arginine residues on the SLiM (blue) and β2 strand of the βHDD (green). (**D**) Hydrophobic aggregation of leucine and isoleucine residues of the SLiM into the hydrophobic core of the βHDD.

bridge formation, forming pairwise interactions with an arginine and glutamate on the SLiM, respectively (figure **1.13 C**). Isothermal titration calorimetry (ITC) measurements of affinities between the excised βHDD and SLiM demonstrated charge reversal of either charged residues on the SLiM decreased binding affinity by a factor of 4-fold or higher. This highlights the requirement for complementary pairwise charges for efficient binding. Two large hydrophobic amino acids, leucine and isoleucine, flanking the charged residues of the SLiM were shown in the same NMR structure to bury into a

hydrophobic pocket of the βHDD between the α2 and α3 helices and the loop linking the α2 helix and the β2 strand (figure **1.13 D**). Mutations of these two residues to alanines increased the dissociation constant by a factor greater than 7-fold. Whilst these results corroborate the 'electrostatic code' model concluded from the TubC βHDD studies, it is likely that binding affinity is determined by a balance of electrostatic and hydrophobic interactions. Such a balance of interactions may explain the variety of charge patterns identified in both characterised and putative SLiM-βHDD pairs, as well as the high occurrence of leucine and isoleucine residues in predicted SLiM amino acid sequences.[123]

Analysis of the interaction between the standalone C domain Bamb_5915 and the SLiM-bearing Bamb_5917 PCP domain from the unique polyketide release mechanism in the enacyloxin IIa (**35**) biosynthetic pathway from *B. ambifaria* (figure **1.14 A**) both corroborated the findings of Hacker *et al*. and provided additional information concerning the contribution of the SLiM-βHDD interface to overall protein-protein interactions. Solution-state NMR titrations of the Bamb_5917 PCP domain with both the excised Bamb_5915 βHDD and full-length C domain were undertaken. From local K_a calculations, the final 10 amino acids of the PCP domain were shown to be involved in binding (figure **1.14 B**). Titration experiments also identified several residues on the surface of the α_2 helix of the PCP domain, adjacent to the 4'-phosphopantetheine attachment site, likely deriving from binding of the globular portion of the C domain. High local K_a values were determined for residues on the surface of the α_1 helix. In addition to NMR titrations, carbene footprinting experiments, a mass-spectrometry-based technique which gives indications of changes of solvent accessibility associated with complex formation, provided supplementary evidence for similar binding regions on the Bamb_5917 PCP domain. For the βHDD, however, a decrease in solvent accessibility was only determined for the α1 helix, which is not thought to participate in interface formation. This was attributed to the conformational flexibility of the βHDD and associated linker region potentially reducing the degree of labelling of the βHDD when not in complex with the SLiM.

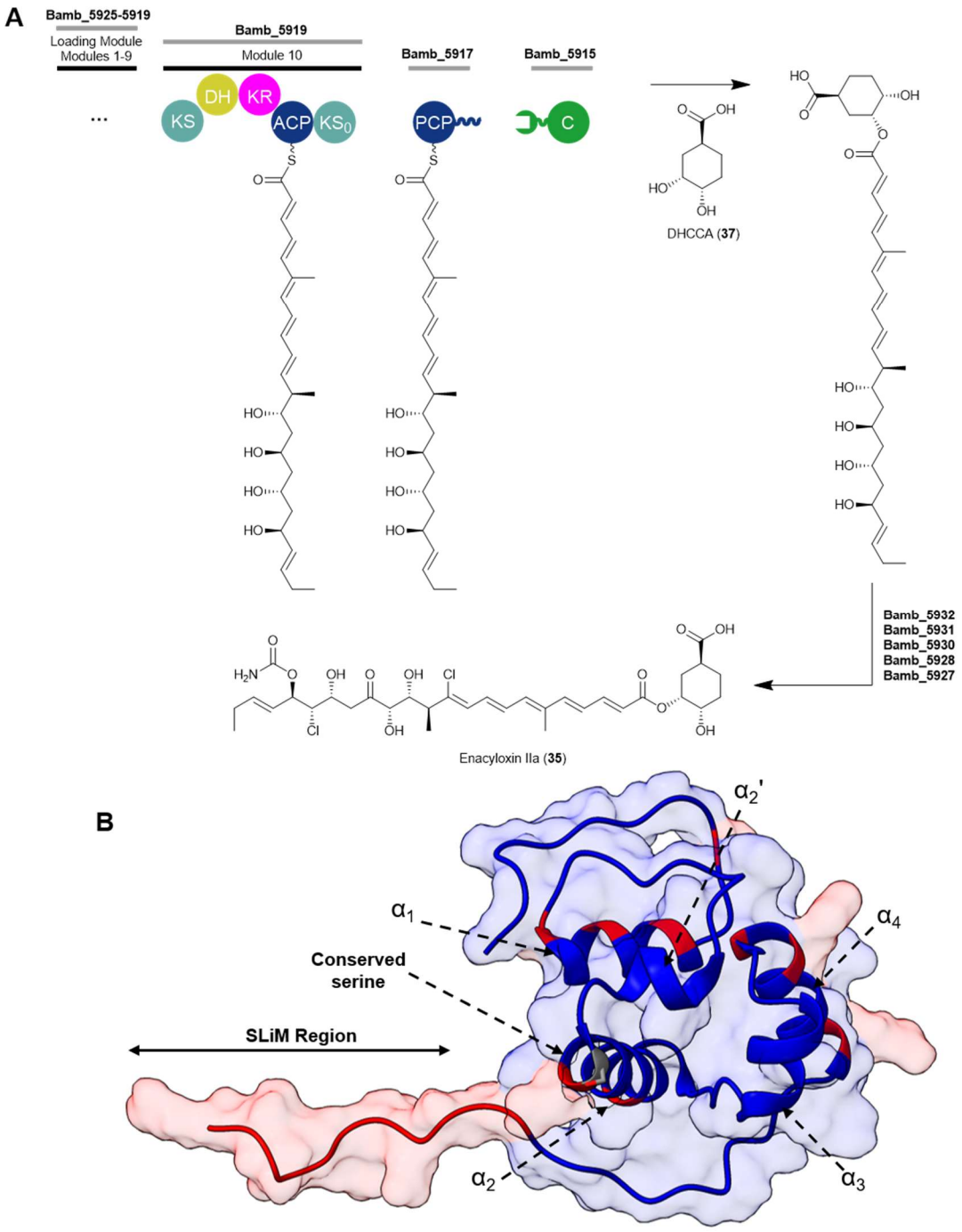
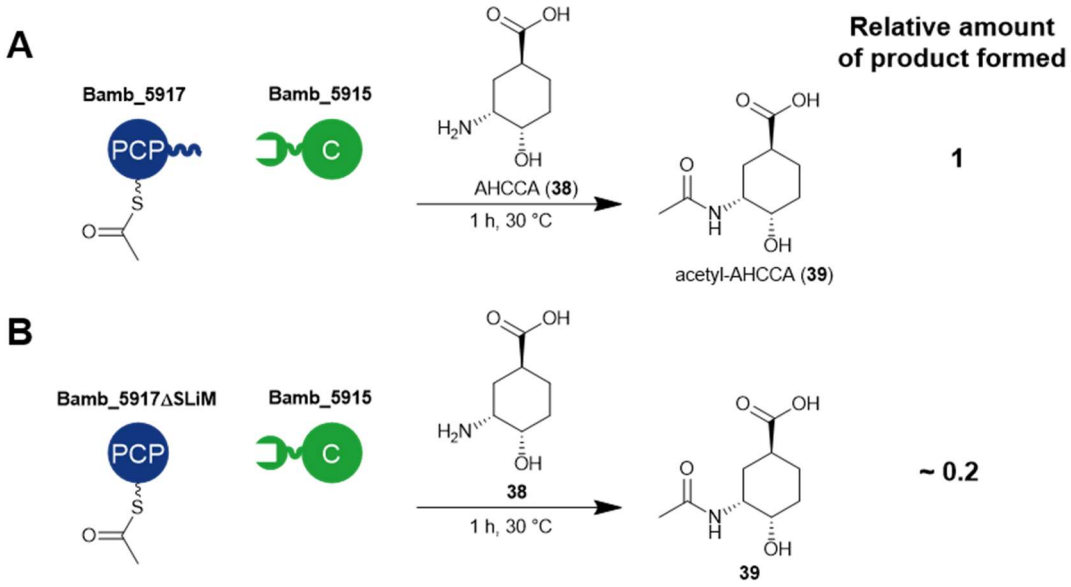


Figure 1.14 (**A**) The partial domain architecture of the enacyloxin IIa biosynthetic pathway in *Burkholderia ambifaria*. A SLiM-βHDD interface is found at the Bamb_5917-Bamb_5917 subunit junction, which constitute the release mechanism of the polyketide. (**B**) NMR solution structure of Bamb_5917(PCP) from enacyloxin IIa biosynthesis (PDB accession code: 5MTI). Highlighted in red are residues with a local K_a higher than average when titrated with interaction partner Bamb_5915 in NMR experiments. The conserved serine which acts as the 4'-phosphopantetheine arm attachment site is depicted in grey.

One of the key findings with regards to SLiM-βHDD interactions has been that their association is not only beneficial to protein-protein binding, but can indeed be crucial for productive intermodular interactions. Removal of the



Scheme 1.10 *In vitro* reactions demonstrating: (**A**) Bamb_5915 catalysing reaction of acetyl-Bamb_5917(PCP) and DHCCA (**37**) analogue, AHCCA (**38**); (**B**) reaction as in (**A**) with Bamb_5917 lacking the C-terminal SLiM. Relative product formation was determined by comparison of peak areas of LC-MS analysis of the reaction mixture.

SLiM region from the Bamb_5917 PCP domain was shown to significantly inhibit the condensation of an acetyl-PCP thioester and an amine analogue of the native small molecule (1S,3R,4S)-3,4-dihydroxycyclohexane-1-carboxylic acid (DHCCA, **37**) substrate, (1S,3R,4S)-3-amino-4-hydroxycyclohexane-1-carboxylic acid (AHCCA, **38**) catalysed by the C domain Bamb_5915 *in vitro* (scheme **1.10**). Meanwhile, titrating either the synthetic Bamb_5917 SLiM peptide or excised Bamb_5915 βHDD into the *in vitro* reaction was demonstrated to appreciably decrease the amount of condensation product formed.[123]

1.5.2.2.3. Engineering biosynthetic pathways using β-hairpin docking domains

Biosynthetic engineering using SLiM-βHDD interfaces remains relatively underexplored in comparison to their PKS counterparts. However, some preliminary experiments to determine their utility to engineer biosynthetic pathways and modulate product formation have been undertaken. Thus far, this has typically been attempted via two distinct methodologies; point mutation of existing docking domains or introduction of non-cognate docking domains to enhance or diminish the degree of productive protein-protein interactions within a biosynthetic system, and utilising docking domain-bearing

proteins from different biosynthetic systems to achieve productive enzymatic cross-talk.

The first attempt at βHDD-derived biosynthetic engineering utilised the myxothiazol A (**40**) biosynthetic pathway from *Stigmatella aurantiaca* DW4/3-1 (figure **1.15 A**).[129,130] Sequence alignment with the TubC docking domain had previously revealed 3 putative SLiM-βHDD interfaces in this biosynthetic

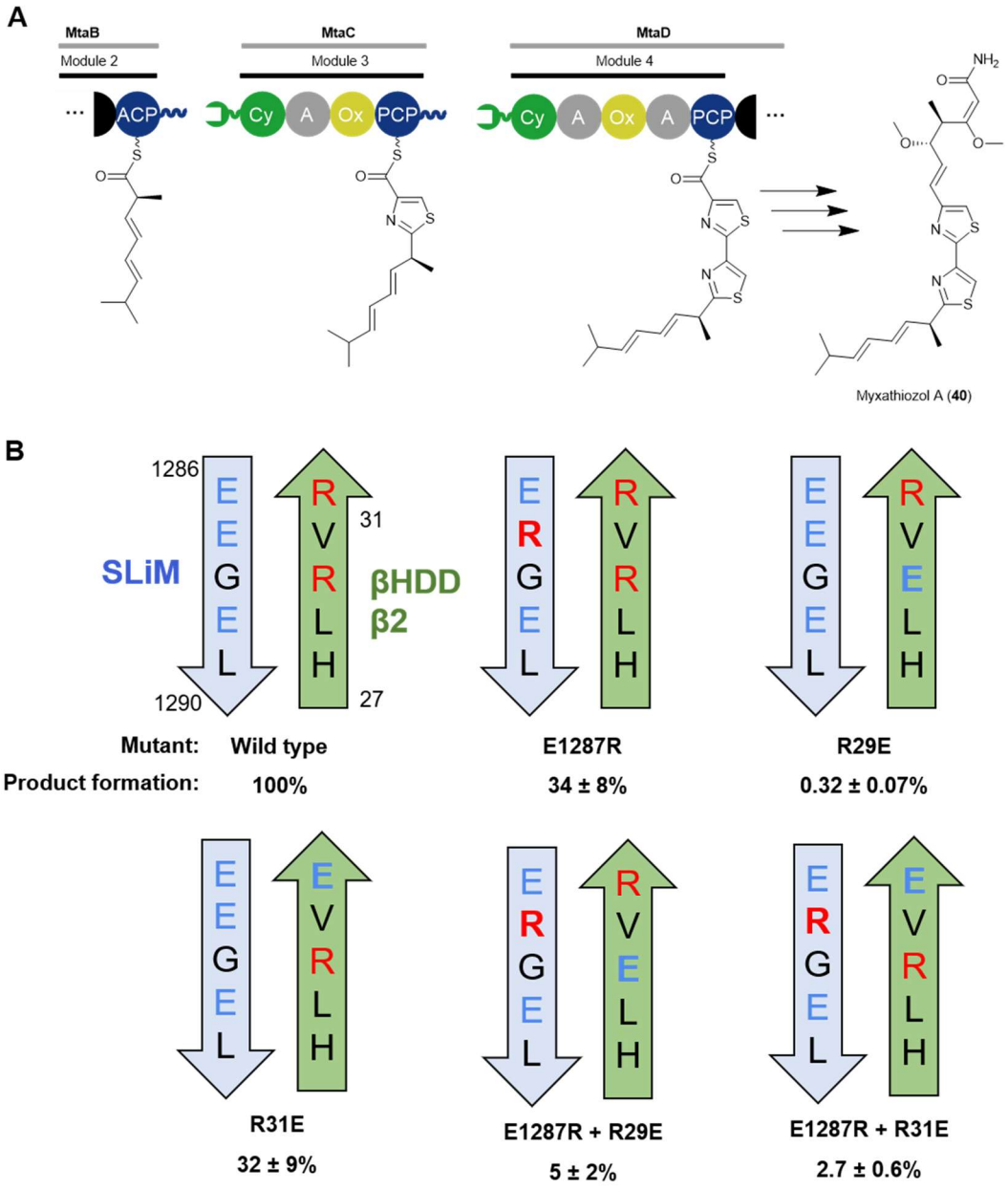


Figure 1.15 (**A**) Partial domain architecture of the myxathiazol A hybrid PKS-NRPS biosynthetic pathway demonstrating the MtaB-MtaC and MtaC-MtaD SLiM-βHDD interfaces. (**B**) Schematic representations of the MtaC-MtaD SLiM-βHDD interface and exemplar point mutations conducted by Li *et al*. Levels of heterologous production of myxothiazol A in *Pseudomonas putida* with these engineered interfaces is presented as a percentage relative to wild type production. Positively charged residues are coloured red, whilst negatively charged residues are coloured blue.

system. The MtaC-MtaD interface was chosen as a model interface for introducing docking domain mutations. Heterologous expression of the myxothiazol biosynthetic gene cluster in *Pseudomonas putida* had previously been achieved and production of the natural product in the heterologous host observed.[131] Point mutations in residues in both the βHDD and SLiM were achieved and introduced to the heterologous host in order to understand the effect on product formation. It was found that reversing the charge of key charged residues in either the SLiM or βHDD β2 strand led to inhibition of myxothiazol production in the heterologous host, indicating biosynthetic fidelity had been diminished (figure **1.15 B**). However, reversing a residue on the βHDD whilst systematically reversing the oppositely charged residues of the putative interaction pair on the SLiM did not provide complete restoration of production. This suggests that the docking interface may be more complex than simple consideration of charge-charge interactions. It also remains to be confirmed whether reduction in production levels was entirely due to change in the electrostatic profile of the docking domains or if mutagenesis of the residues caused structural changes which inhibit productive protein-protein interactions. Nevertheless, these experiments clearly demonstrate that altering the fidelity of biosynthesis in an *in vivo* context by alteration of docking domain affinities is possible, providing scope to continue to develop SLiM-βHDD pairs as future engineering tools.

A more elaborate example of biosynthetic engineering using SLiM-βHDD has been described for the rhabdopeptide-producing NRPS.[126] Guided by *in vitro* docking domain affinities, the interaction between the Kj12A subunit and standalone C domain Kj12C was artificially improved by mutagenesis of the Kj12A SLiM. *In vivo*, this led to an increase in production of long-chain rhabdopeptides, owing to the stronger interaction between Kj12A and Kj12C outcompeting the interaction between Kj12B and Kj12C. This prevents chain termination and instead promotes interaction of Kj12B with another Kj12B subunit to catalyse multiple rounds of elongation (figure **1.16 A**). The same shift in peptidic product profile was observed when the Kj12B βHDD was engineered for tighter binding to the Kj12B SLiM or when the Kj12C SLiM was engineered for weaker binding to the Kj12B SLiM, both promoting chain

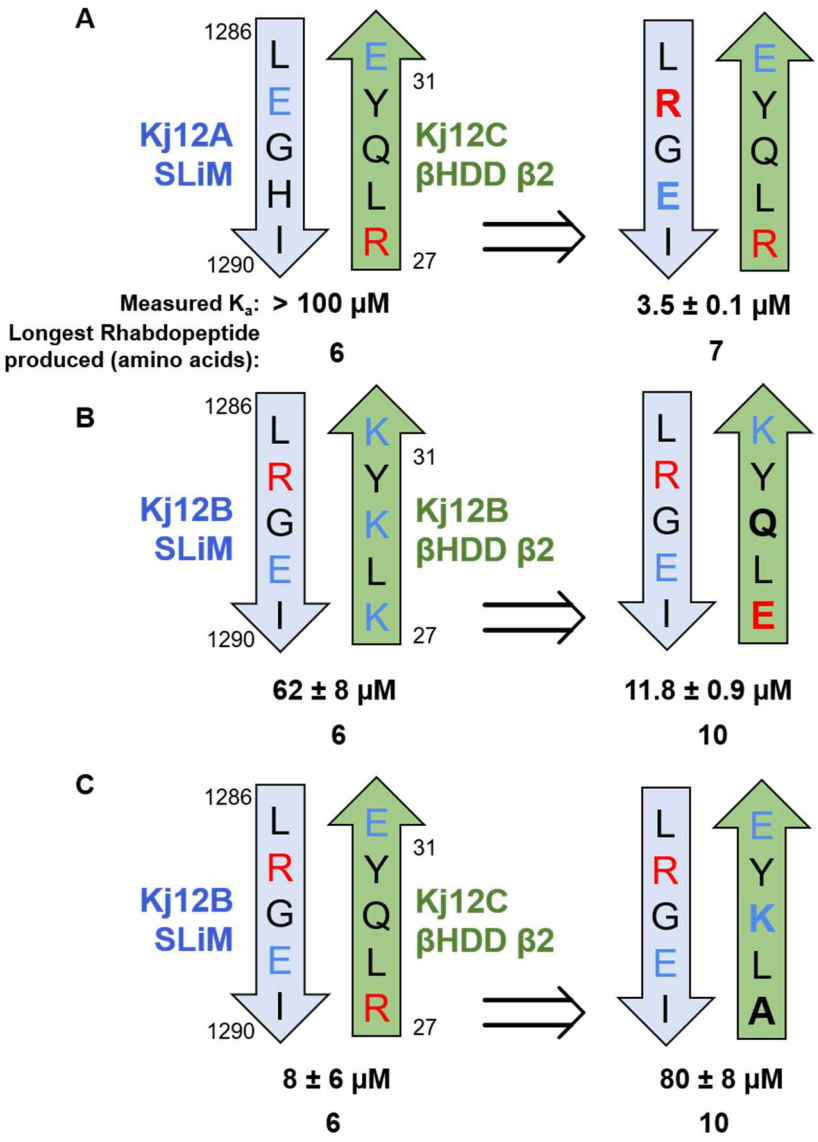


Figure 1.16 Engineering of the rhabdopeptide-producing NRPS *via* exchange of the wild-type SLiM-βHDD interfaces with heterologous SLiM-βHDD pairs from the xenoamicin-producing NRPS in *Xenorhabdus innexi* DSM 16336. (**A**) Wild-type system. (**B**) Replacement of the Kj12A-B DD pair with the XabA-XabB DD pair. (**C**) Sequential replacement of the Kj12B-C DD pair with the XabB-XabC DD pair. The rhabdopeptide produced by these engineered pathways is provided. mV = N-methyl valine, V = valine, PEA = phenylethylamine.

elongation (figure **1.16 B-C**). In a later study, grafting of homologous docking domain pairs from the xenoamicin-producing NRPS from *Xenorhabdus innexi* DSM 16336 to replace the native docking domains of the rhabdopeptide-producing NRPS was shown to permit productive protein-protein interactions. Perhaps more interestingly, implementing orthologous SLiM-βHDD pairs at the Kj12A-Kj12B and Kj12B-Kj12C boundaries could convert this usually iterative NRPS system into a near-linear system, shifting the product profile dramatically towards shorter-chain peptides (figure **1.17**).[132]

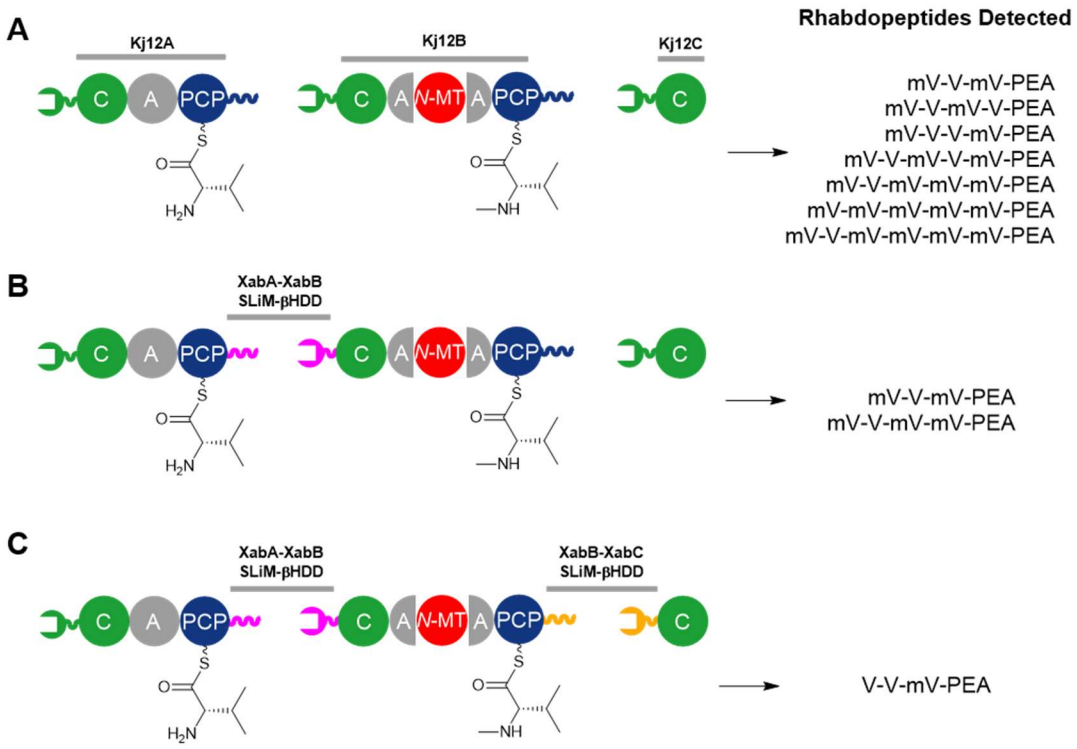
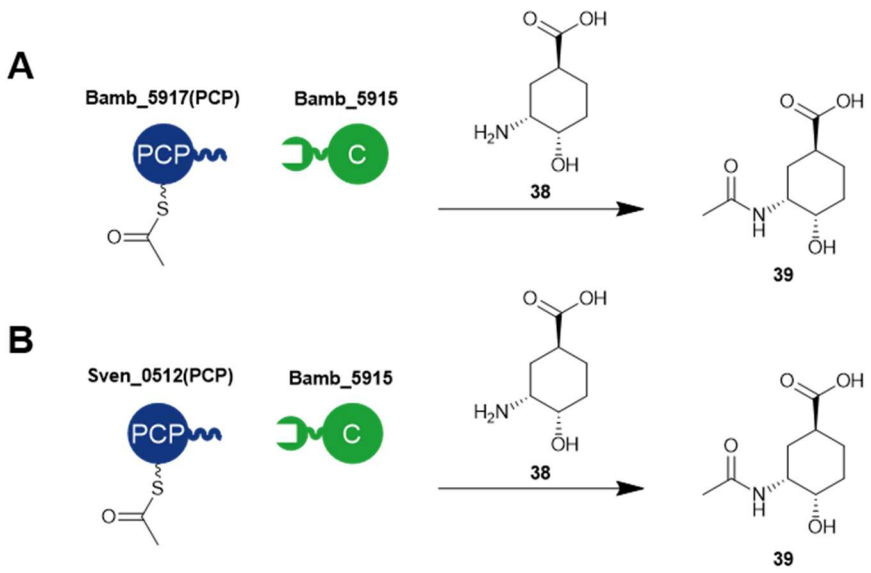


Figure 1.17 Engineering of the rhabdopeptide-producing NRPS by introduction of mutations at the SLiM-βHDD interfaces between Kj12A, Kj12B and Kj12C. The wild-type interfaces are displayed on the left and engineered interfaces are displayed on the right.

Exploiting crosstalk of non-cognate interaction partners has been shown in PKS and NRPS systems to be a rapid and effective method for exploring the amenability of docking domains towards engineering approaches.[133] Bamb_5915 and the Sven_0512 PCP domain from watasemycin biosynthesis in *Streptomyces venezuelae*, which bears a SLiM with high sequence identity to the Bamb_5917 SLiM, were determined to be compatible by solution-state NMR. Subsequently, *in vitro* experiments demonstrated that a productive condensation reaction between acetyl-Sven_0512(PCP) and AHCCA (**38**) catalysed by Bamb_5915 could be observed (scheme **1.11**).[134] This is despite the Sven_0512 PCPs native interaction partner, the cyclisation domain of Sven_0517, having an appended βHDD that only shares 35% sequence homology with that of Bamb_5915. Titration of the excised Sven_0512 SLiM or βHDD could inhibit the reaction, demonstrating the importance of the SLiM-βHDD interface in mediating productive crosstalk. These preliminary results



Scheme 1.11 *In vitro* assays demonstrating acylation of synthetic DHCCA (**37**) analogue AHCCA (**38**) by: (**A**) cognate interaction of the Bamb_5917 PCP domain with standalone C domain Bamb_5915, both from enacyloxin IIa biosynthesis; (**B**) crosstalk of the Sven_0512(PCP) from watasemycin biosynthesis with Bamb_5915.

exemplify the ability of these docking domains to recruit non-native and evolutionarily divergent NRPS subunits. This could be a valuable characteristic of these domains in future attempts to engineer *de novo* NRPS assembly lines to produce novel natural product analogues.

1.5.2.3. Novel Docking Domains from PAX Peptide Biosynthesis

Wöhnert and co-workers have recently identified a novel docking domain archetype in peptide-antimicrobial-Xenorhabdus (PAX) peptide (**41**) biosynthesis in *Xenorhabdus bovienii* (figure **1.18**).[135,136] These docking domains were identified at the C-terminus of the PCP domain of PaxB and the N-terminus of the downstream C domain of PaxC (figure **1.18 A**). Solution-state NMR structures of the two excised DDs connected by a flexible 25 amino acid linker revealed that the two domains in complex adopt a unique three-α-helix bundle structure (figure **1.18 B**). Similarly to βHDDs, a number of polar residues have been identified as important for electrostatic interactions and salt bridge formation between the DDs and may constitute a comparable 'electrostatic code'. However, hydrophobic interactions also appear to contribute substantially to interface formation. At least 17 of these docking domains have been identified by bioinformatics analysis in a diverse range

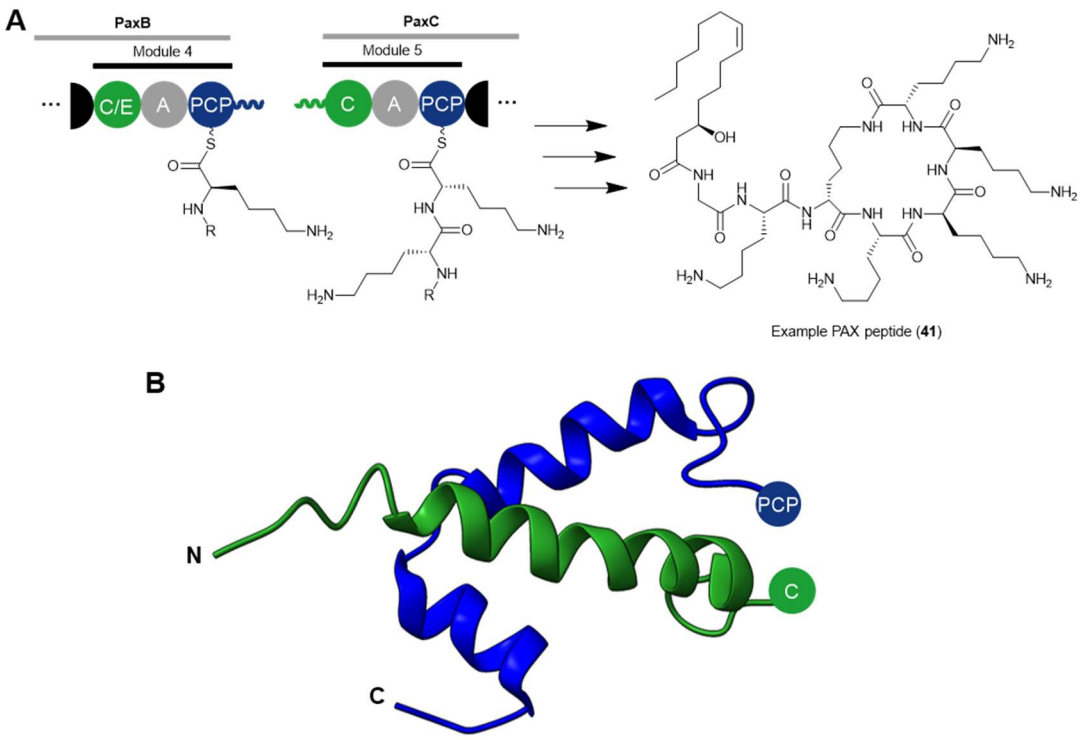


Figure 1.18 (**A**) Partial domain architecture of the PAX-producing NRPS highlighting the docking domain interface between PaxB and PaxC. (**B**) NMR solution structure of the PaxB-PaxC docking domains in complex (PDB accession code: 6TRP).

NRPS pathways from evolutionary-divergent bacteria, lending promise to future attempts to use these domains for combinatorial biosynthesis applications. Their capacity to mediate communication across the same PCP-C domain boundaries as SLiM-βHDDs may allow these two archetypes to act as orthogonal docking domains in biosynthetic engineering.

1.6. Aeruginosins

Aeruginosins are a family of linear peptide compounds isolated from marine sponges and cyanobacteria that display inhibitory activity against trypsin-specific serine protease inhibitors, and are of interest as potential anticoagulant agents.[137] They are noteworthy compounds due to their uncommon structural features, including the 2-carboxy-6-hydroxyoctahydroindole (Choi) moiety, N-terminal 4-hydroxylphenyllactic acid (4-HPLA) derivatives and C-terminal arginine derivatives (figure **1.19**).

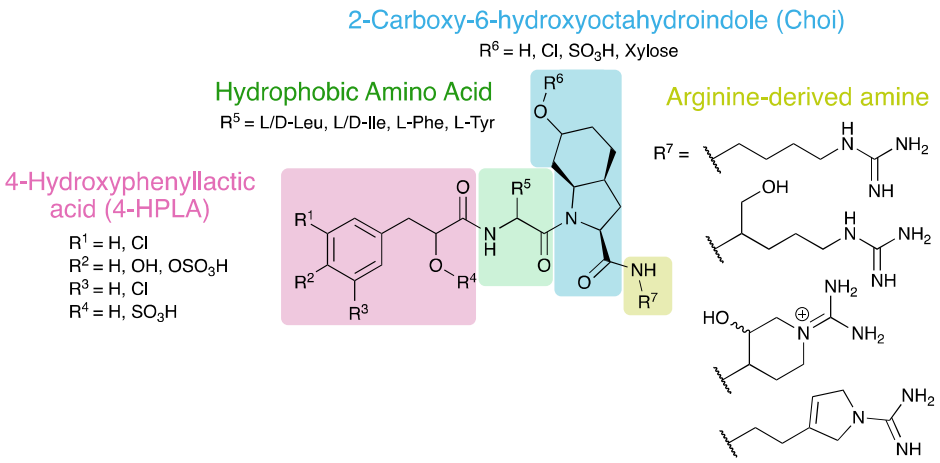


Figure 1.19 The structure of aeruginosin congeners isolated from marine sponges and cyanobacteria.

1.6.1. Biosynthesis of aeruginosin in *Microcystis aeruginosa* NIES-98

Microcystis aeruginosa NIES-98 produces three congeners of aeruginosin: 98A (**42**), 98B (**43**) and 101 (**44**), which differ only by chlorination pattern of their 4-HPLA motif (figure **1.20**).[138,139] The 25 kb aeruginosin biosynthetic gene cluster[139,140] consists of a hybrid PKS-NRPS gene, *aerA*, and three NRPS genes, *aerB*, *aerG1* and *aerG2*. 4-hydroxyphenylpyruvate (**45**), an intermediate in the biosynthesis of tyrosine from chorismate,[141] is predicted to be the starter unit. This is adenylated by the AerA A domain. This can be subsequently reduced by the ketoreductase domain encoded in the *aerA* gene to yield the 4-HPLA moiety. The AerB condensation domain can then catalyse the condensation of 4-HPLA with L-isoleucine, which is epimerised at the

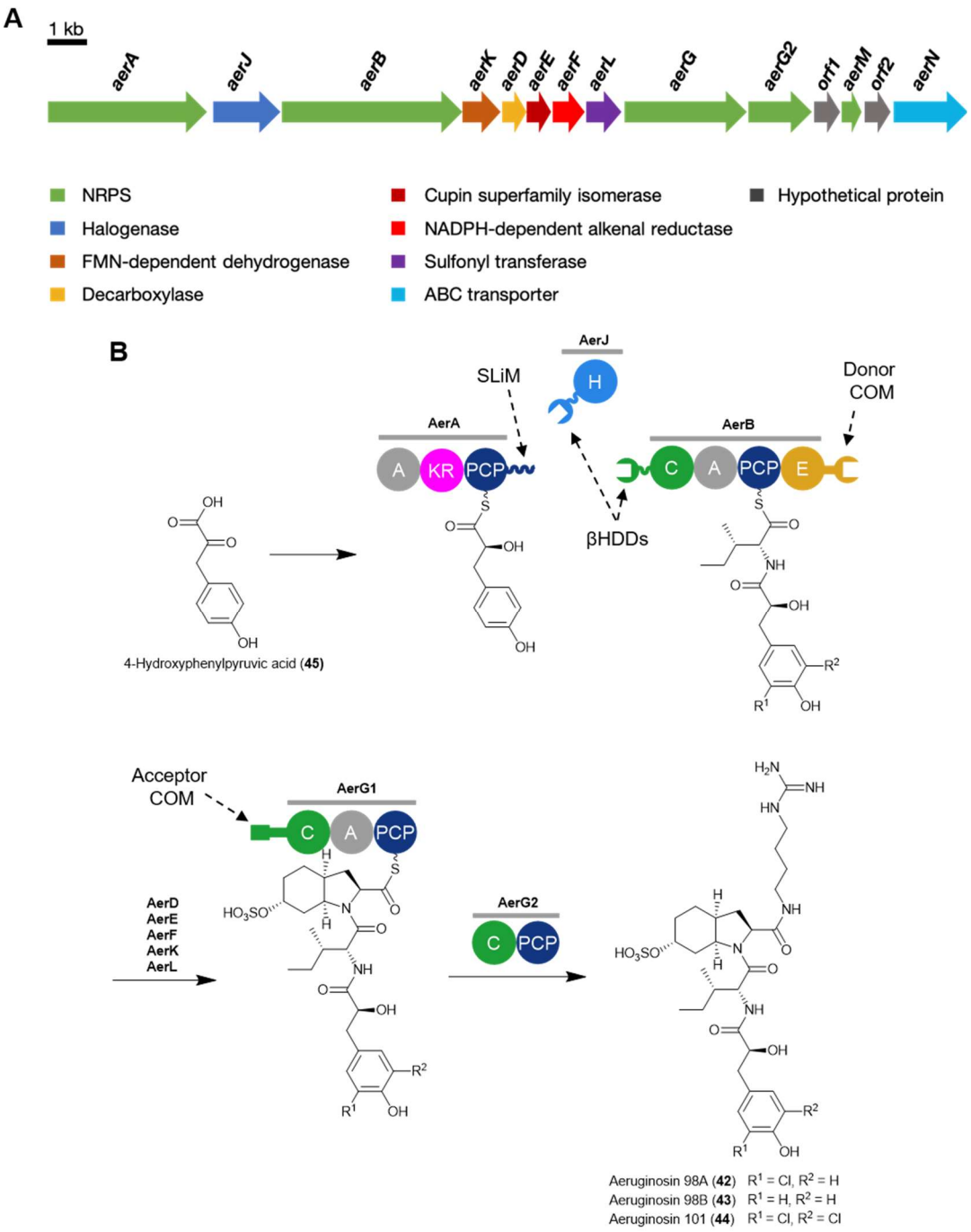
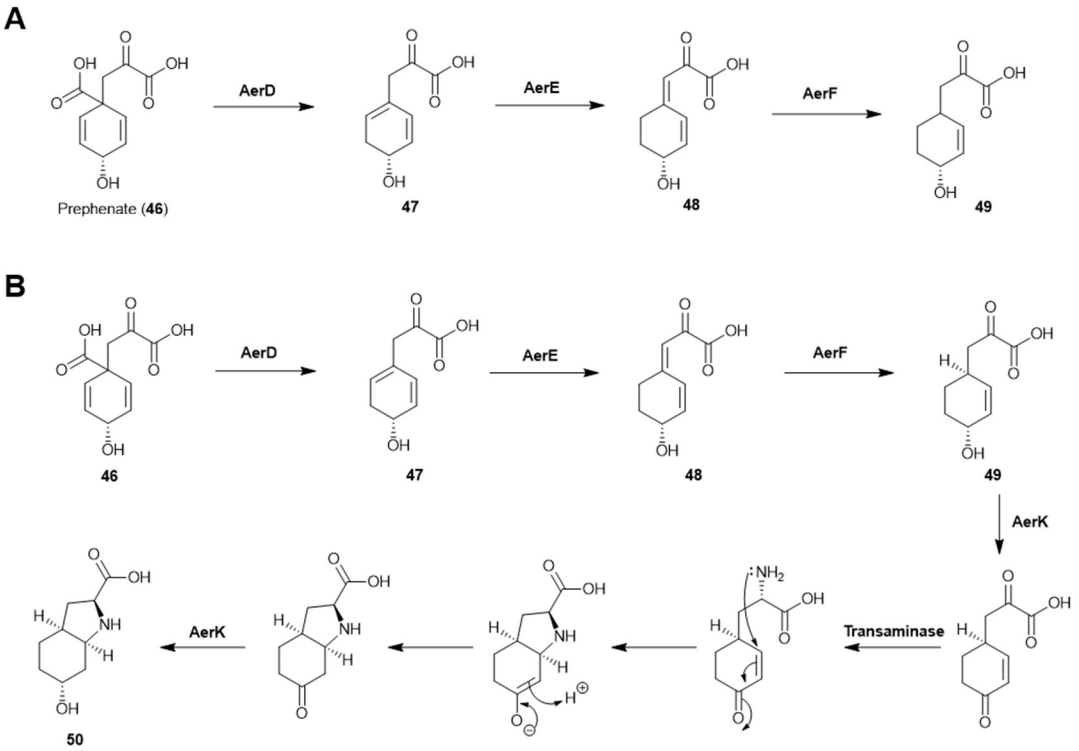


Figure 1.20 (**A**) The biosynthetic gene cluster of aeruginosin in *M. aeruginosa* NIES-98. The putative function of each gene is provided. (**B**) The proposed biosynthesis of aeruginosin congeners produced by *M. aeruginosa* NIES-98.

α-carbon to D-allo-isoleucine by the module's epimerisation domain. The substrate of AerG1's A domain cannot be reliably predicted *in silico*. However, studies in the homologous aeruginoside 126A and 126B biosynthetic pathways in *Planktothrix agardhii* CYA126/8 have suggested the substrate could be similar to arogenate, a tyrosine biosynthesis intermediate.[141]

Following loading of this cryptic residue onto the AerG1 PCP, the C domain catalyses elongation. Chain release, in the absence of a putative thioesterase or thioester reductase domain, is proposed to be catalysed by condensation of the nascent peptide with a molecule of arginine-derived agmatine by the condensation domain of AerG2. However, this release mechanism has yet to be characterised. More curiously, AerG2 appears to house a PCP domain without a corresponding A domain in the gene cluster and its biochemical function is yet to be determined.

Genes *aerD-F* have been shown *via* mutational studies by Ishida *et al.* to be critical to synthesis of the Choi moiety in the biosynthesis of aeruginosides 126A and 126B.[142] AerD is a decarboxylase with sequence homology to BacA from *Bacillus subtilis* and converts prephenate (**46**) from tyrosine and phenylalanine biosynthesis to dihydro-4-hydroxyphenylpyruvate (**47**) (scheme **1.12 A**).[143,144] An AerE orthologue from *M. aeruginosa* PCC7806 has been shown to act as an isomerase, catalysing 1,3-allylic isomerisation of **47** to give compound **48** (scheme **1.12 A**).[143] AerF is homologous to the



Scheme 1.12 (**A**) The conversion of tyrosine precursor prephenate (**46**) to **49** by AerD, AerE and AerF as determined by Mahlstedt *et al.* and Yan and co-workers. (**B**) A proposed biosynthesis of the choi moiety from prephenate. It is unknown whether the enzymes act upon the free acid or a thioester-tethered substrate *in vivo*.

dihydroanticapsin 7-dehydrogenase BacC from bacilysin biosynthesis in *Bacillus subtilis* and has been shown to act as an NADPH-dependent alkenyl reductase, reducing the 3,4-double bond of **48** (scheme **1.12 A**).[145] *aerK* has been demonstrated by gene knockout experiments to be critical for aeruginosin production in *M. aeruginosa* PCC 7806.[139] From basic local alignment search tool (BLAST) analysis[146] of *aerK*, the encoded protein is potentially a type II isopentenyl-diphosphate delta-isomerase, a reduced-flavin mononucleotide (FMN)-dependent dehydrogenase-like protein. Genes with high percentage identity to *aerK* appear to be ubiquitous in *Microcystis* strains, even in those which do not contain other genes from the aeruginosin biosynthetic cluster,[147] but the reason for this is unclear. Based upon these observations and proposed functions, a postulated pathway for Choi moiety formation from prephenate **46** is provided in scheme **1.12 B**. However, it is not known whether these proteins act on the free substrate or a carrier protein-tethered thioester. Moreover, conversion of the α-keto group to an amino group is necessary for this mechanism, yet the aeruginosin cluster does not appear to contain a gene encoding a transaminase. However, it is plausible that a transaminase encoded elsewhere in the genome is recruited for this purpose. Closer mechanistic examination of the biosynthetic proteins involved in Choi moiety formation is necessary to understand exactly how this unique aminoacyl group is synthesised.

The aeruginosin gene cluster also contains genes encoding structure-tailoring and processing enzymes. The *aerJ* gene encodes a flavin-dependent halogenase, which is proposed to catalyse *ortho*- chlorination of the 4-HPLA phenol. Due to the presence of an N-terminal βHDD, AerJ is highly likely to act on the AerA carrier protein-tethered substrate (see section 1.6.3). *aerL* encodes a putative sulfotransferase, proposed to sulfonate the hydroxyl group of the Choi moiety.[139] *aerN* encodes an ATP-binding cassette transporter-like protein.[148] The roles of other genes in the cluster are enigmatic. Fragments of an *aerM* orthologue, an NRPS-encoding gene found in aeruginosin biosynthetic clusters of other *Microcystis* strains, is present but is predicted to be non-functional.[139] Genes annotated *orf1* and *orf2* code for a putative NAD(P)H-dependent oxidoreductase that is ketoreductase-like and a

hypothetical protein with conserved domains aligning with the acyltransferase superfamily of proteins, respectively. The biosynthetic significance of these genes remains currently unclear.

1.6.2. Docking Domains in aeruginosin biosynthesis

A profile hidden Markov (HMM) model by Kosol *et al.* searching for proteins with N-terminal βHDDs identified a unique βHDD-SLiM interface in aeruginosin biosynthesis.[123] Two βHDDs, appended to the N-terminus of the standalone flavin-dependent halogenase AerJ and NRPS subunit AerB, were identified. However, only one SLiM could be identified at the C-terminus of AerA. As a result, both AerJ and AerB are predicted to interact with AerA and their βHDDs may be competing for binding to the AerA SLiM. This competition could be responsible for the mixture of chlorinated congeners isolated from *M. aeruginosa* NIES-98.[139] The apparent affinity of both βHDDs for this single SLiM is further exemplified by their high homology and comparable predicted secondary structures (figure **1.21**).

Interestingly, a putative COM domain pair is found at the AerB-AerG1 interface and is presumably orthogonal to the SLiM-βHDD mediated interface between

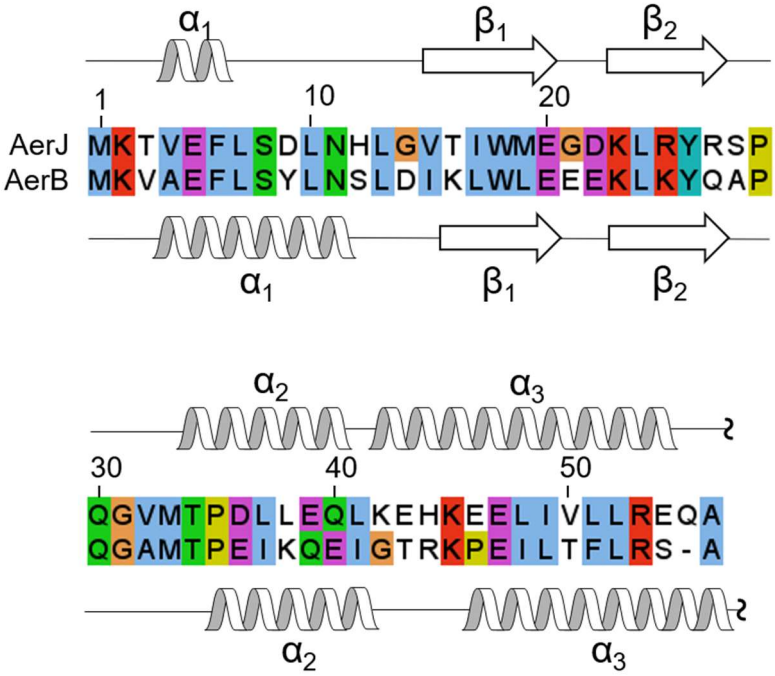


Figure 1.21 Comparison of the amino acid sequences of the AerJ and AerB βHDDs. The secondary structural elements predicted by the PsiPred 4.0 webserver are given for each sequence.

AerA and AerB. However, this docking domain interface currently remains uncharacterised.

The aeruginosin biosynthetic system offers a unique foundation for expanding the versatility of SLiM-βHDD interfaces. Although competition between βHDDs for SLiM binding has described for rhabdopeptide biosynthesis, as discussed in section 1.5.2.2, the aeruginosin NRPS more closely resembles a canonical, linear biosynthetic pathway. This suggests that observations from this system may be more universally applicable to NRPS megasynthetases. Moreover, this system represents a remarkable first example of a tailoring enzyme, rather than a chain-elongating condensation or cyclisation domain, utilising a βHDD for binding to a carrier protein domain. This provides promise for biosynthetic engineering applications, as βHDDs could potentially be appended to a diverse range of catalytic enzymes, facilitating recruitment to carrier protein-tethered substrates. This could allow hijacking of pathways during early biosynthetic steps to install simple chemistries, enabling development of libraries of natural product analogues.

1.7. Project aims and objectives

The overall aim of this project was to probe the two SLiM-βHDD interfaces in the initial steps of aeruginosin biosynthesis by *in vitro* reconstitution to prove the hypothesised biosynthetic origins of the mixture of chlorination patterns in the isolated natural product. Further to this, it was envisioned that the docking interfaces could be probed by complementary biochemical and biophysical techniques to develop our understanding of the key residues mediating SLiM-βHDD binding.

Further to this initial goal, understanding the ability of the AerA PCP domain to crosstalk with the βHDD-bearing C domain Bamb_5915 from enacyloxin IIa (**35**) biosynthesis was desirable. We aimed to investigate this by both *in vitro* assays and *in vivo* engineering of enacyloxin-producing *B. ambifaria* strains. This would help to develop understanding of SLiM-βHDD interfaces as methods to engineer biosynthetic pathways to generate new molecular entities with potential bioactivity and drug-like properties.

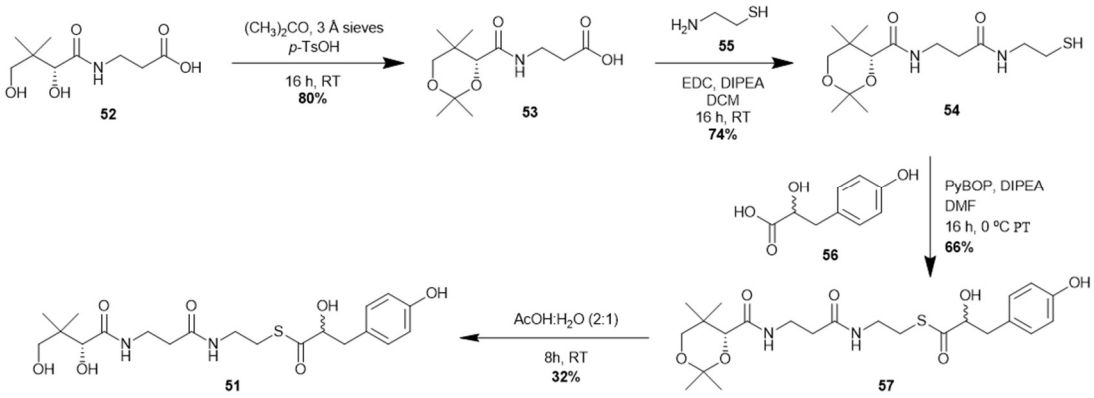
Chapter 2: *In vitro* reconstitution of the initial steps of aeruginosin biosynthesis

2.1. *In vitro* reconstitution of AerJ activity

Reconstitution of flavin-dependent halogenase activity towards small molecule substrates *in vitro* is well-established.[67,149] To investigate the recruitment of flavin dependent halogenase AerJ to the AerA PCP domain by docking domain-mediated interactions, an assay that monitors halogenation on a PCP domain-tethered substrate was desirable. Cryle and co-workers have shown halogenation of carrier protein tethered substrates can be monitored by methylamine cleavage of the pantetheine thioester prior to single ion monitoring or multiple reaction monitoring by UHPLC-MS.[150] However, this method cannot monitor the degree of substrate loading and thioester hydrolysis during the reaction. This is important for AerJ, as the expected 4-hydroxylphenyllactic acid substrate is similar in structure to tyrosine, and amino acyl-*S*-pantetheines are often less hydrolytically stable than polyketide-like derivatives. Phosphopantetheine ejection experiments, in which the PCP domain-tethered product can be by monitored by observation of the elimination product following collision-induced dissociation of the protein,[151] are also commonly used to detect substrates appended to PCP domains resulting from *in vitro* experiments.[21,152,153] However, in initial trials of the experiments described herein, this elimination product was not detected by UHPLC-ESI-Q-TOF-MS analysis, and further optimisation of collision energy would likely be required to facilitate this approach. An intact protein mass spectrometry-based assay was therefore chosen to detect halogenase activity.

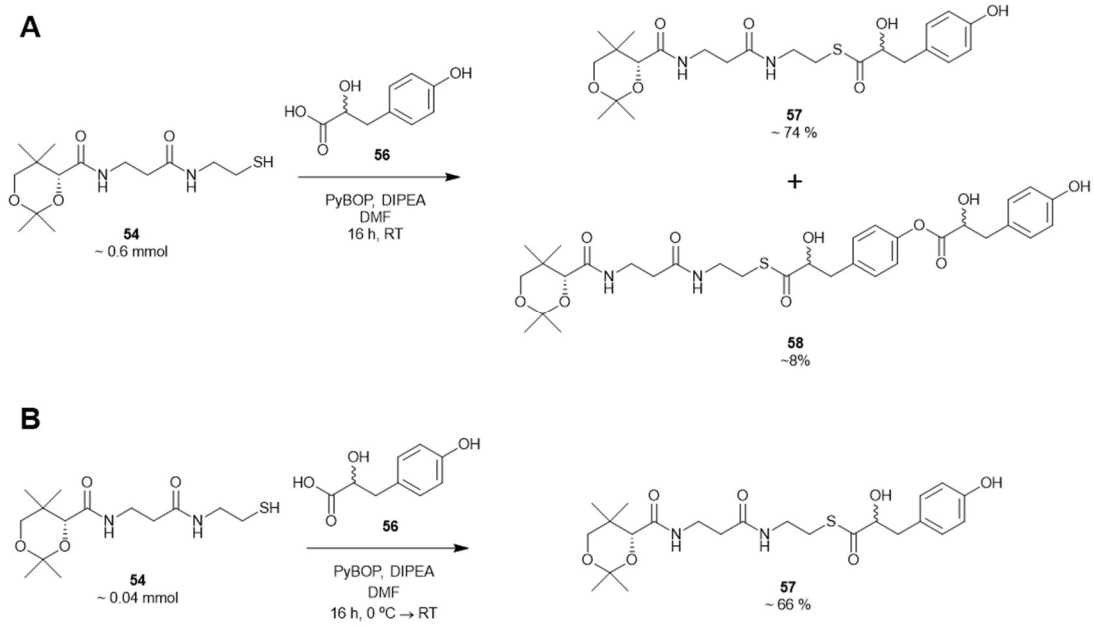
2.1.1. Synthesis of 4-hydroxyphenyllactic acid pantetheine thioester (51)

To establish whether AerJ was able to chlorinate AerA(PCP) tethered substrates, a 4-hydroxyphenyllactic acid pantetheine thioester (**51**) was synthesised for loading onto *apo*-AerA(PCP) (scheme **2.1**). Acetonide-protected pantetheine (**53**) was synthesised from pantothenic acid (**52**) *via* the procedure outlined by Agarwal *et al*.[154] **53** was then converted to acetonide-protected pantetheine (**54**) by coupling to cysteamine (**55**).[154] Commercially available DL-4-hydroxyphenyllactic acid (**56**) was then coupled to the



Scheme 2.1 Synthesis of 4-HPLA-*S*-pantetheine (**51**).

pantetheine thiol of **54** using the phosphonium-based peptide coupling reagent PyBOP®.[155] From trialling multiple reaction solvents, *N,N*-dimethyl formamide (DMF) was found to be essential for solvation of the poorly-soluble acid **56**. Initially, a significant impurity was observed resulting from coupling of a second molecule of **56** to the terminal phenol of **57** to give **58** (scheme **2.2 A**) due to the acidity of phenol and mild nucleophilicity of the generated phenoxide anion. This was found to be inseparable from the desired product by silica gel chromatography. The identity of this by-product was confirmed by low resolution MS and ^{1}H-NMR. Conducting the reaction with low concentrations



Scheme 2.2 (**A**) Formation of an undesirable side product resulting from coupling of the 4-HPLA phenol with the carboxylic acid of a second molecule. Ratios of product formation were determined by comparison of ^{1}H NMR integrals. (**B**) Synthetic optimisations to yield exclusively the singly-coupled product.

of the starting materials and addition of reagents at low temperatures proved necessary to supress formation of this by-product (scheme **2.2 B**). Removal of the acetonide protecting group of **57** under mild acidic conditions afforded the 4-HPLA-*S*-pantetheine thioester (**51**) as a mixture of inseparable diastereomers. Racemic 4-hydroxyphenyllactic acid (**56**) was used due to the high cost of the enantiopure acids. Attempts were made to produce single enantiomers of the starting material via the stereoselective reduction of 4-hydroxyphenylpyruvic acid (**45**) with (+)- and (-)-diisopinocampheylchloroborane in accordance with literature procedures,[156,157] but this was not successful.

2.1.2. Overproduction and purification of recombinant proteins

The AerA PCP domain and AerJ were overproduced and purified by Ni-NTA affinity chromatography from pET24a-(+) constructs originally designed by Dr. Matt Jenner and Dr. Sławomir Potocki. Due to the specialised equipment required for cultivating cyanobacteria such as *Microcystis aeruginosa*, genomic DNA was not available for cloning of the desired genes. Instead, genes encoding AerJ and the AerA PCP domain (from GenBank accession FJ609416.1[139] and AM773660.1,[158] respectively) were synthesised and cloned into pET24a-(+) using NdeI and XhoI restriction sites by Epoch Life Sciences Inc. with no codon optimisation (figure **2.1**). The N-terminal domain boundary of the AerA PCP domain was determined with the aid of annotations

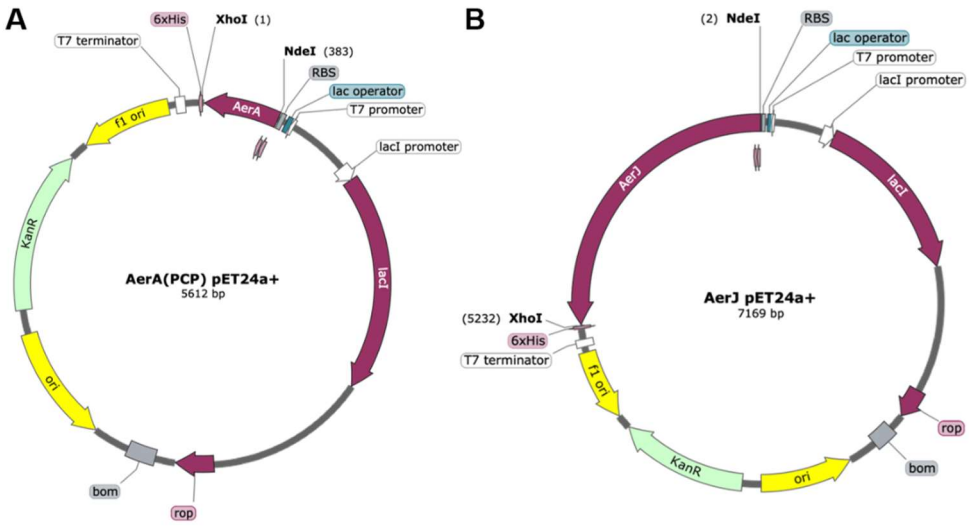


Figure 2.1 Plasmid maps of: (**A**) pET24a-(+)-AerA(PCP) and (**B**) pET24a-(+)-AerJ. Restriction sites used for cloning of the gene insert are provided.

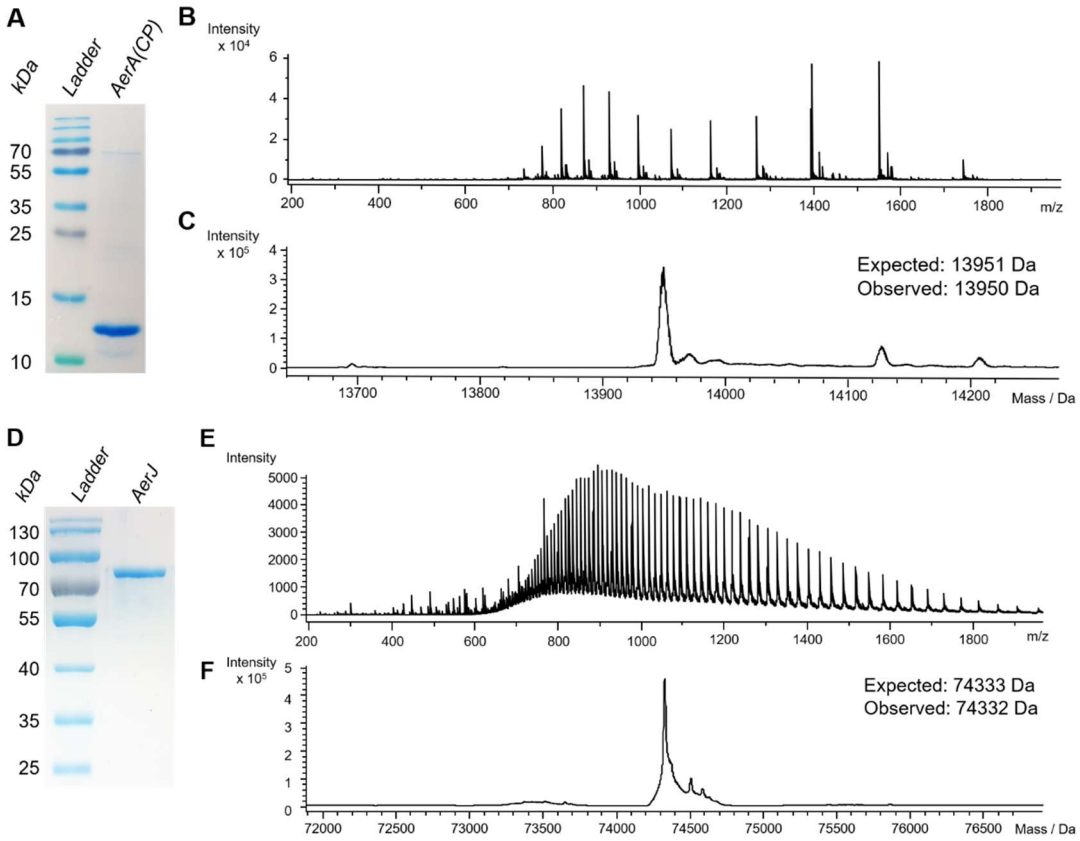


Figure 2.2 Confirmation of the identity of recombinant AerA PCP domain and AerJ. (**A**) SDS PAGE analysis of the AerA PCP domain resulting from Ni-affinity chromatography. (**B**) Mass spectrum of the AerA PCP domain from UHPLC-ESI-Q-TOF-MS analysis. (**C**) Charge state deconvoluted mass spectrum of the AerA PCP domain. (**D**) SDS PAGE analysis of AerJ resulting from Ni-affinity chromatography. (**E**) Mass spectrum of AerJ from UHPLC-ESI-Q-TOF-MS analysis. (**F**) Charge state deconvoluted mass spectrum of AerJ. Expected masses were calculated from the amino acid sequence of the protein expected to be expressed by the relevant plasmid construct.

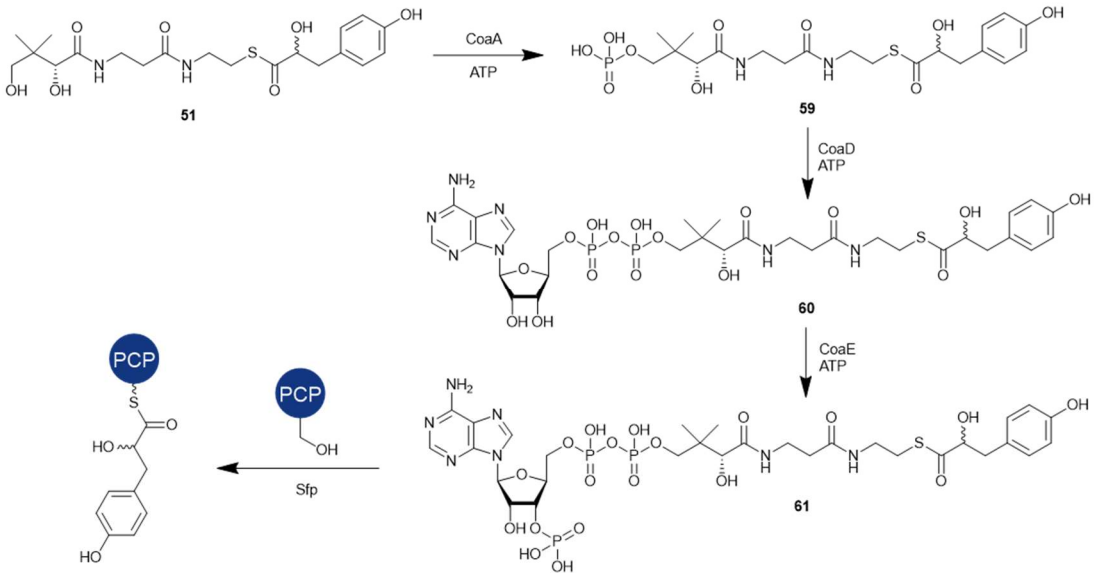
provided by the BLAST-P webserver using the Conserved Domain Database.[159,160] Constructs contained an *N*-terminal His_8-tag allowing nickel affinity purification and a Tobacco Etch Virus (TEV)-protease-cleavable site, allowing removal of the His_8-tag for future structural biology applications. The pET24a-(+) vector contains a T7 promoter, which is recognised by T7 RNA polymerase. T7 RNA polymerase is typically delivered by lysogenic λ prophage DE3-infected *E. coli* under the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible lacUV5 promoter, allowing controlled expression of the protein of interest.[161]

The plasmids were used to transform *E. coli* BL21(DE3) cell lines. Following growth of cells in LB broth to OD_{600} 0.6-0.8 at and overnight induction with IPTG at 15 °C, cells were lysed, clarified by centrifugation and the desired

proteins purified by nickel affinity chromatography. The identity of the purified proteins was confirmed by SDS-PAGE and UHPLC-ESI-Q-TOF-MS analyses (figure **2.2**). A minor amount of proteins of masses 14128 and 14208 Da were observed for the AerA PCP domain (figure **2.2 C**). These mass shift of 178 and 256 Da from the expected protein are consistent with posttranslational phosphogluconylation in the *E. coli* strain.[162]

2.1.3. Chemoenzymatic synthesis of 4-HPLA-*S*-AerA(PCP)

Loading of pantetheine thioesters onto carrier proteins can be carried out using a well-established chemoenzymatic procedure.[154,163-165] The pantetheine thioester is first converted *in situ* to the corresponding acyl-coenzyme A using recombinant enzymes CoaA, CoaD and CoaE from the biosynthesis of coenzyme in *E. coli*. These enzymes are responsible for sequential phosphorylation, adenylation and phosphorylation respectively (scheme **2.3**). Adenylation by CoaD is an ATP-dependent reaction. pET vector expression constructs for these enzymes were provided by the group of Dr. Manuela Tosin, University of Warwick.[166] The acyl coenzyme A derivative (**61**) generated can then be loaded onto the conserved serine of an *apo*-PCP by promiscuous phosphopantetheinyl transferase Sfp from *Bacillus subtilis* in a Mg^{2+}-dependent process (scheme **2.3**).[154,167] A pET vector expression construct for Sfp was also provided by the group of Dr. Manuela Tosin,



Scheme 2.3 Conversion of 4-HPLA-*S*-pantetheine to a coenzyme A derivative by CoaA, CoaD and CoaE from *Escherichia coli* and subsequent loading onto a carrier protein by the phopsphopantetheinyl transferase Sfp from *Bacillus subtilis*.

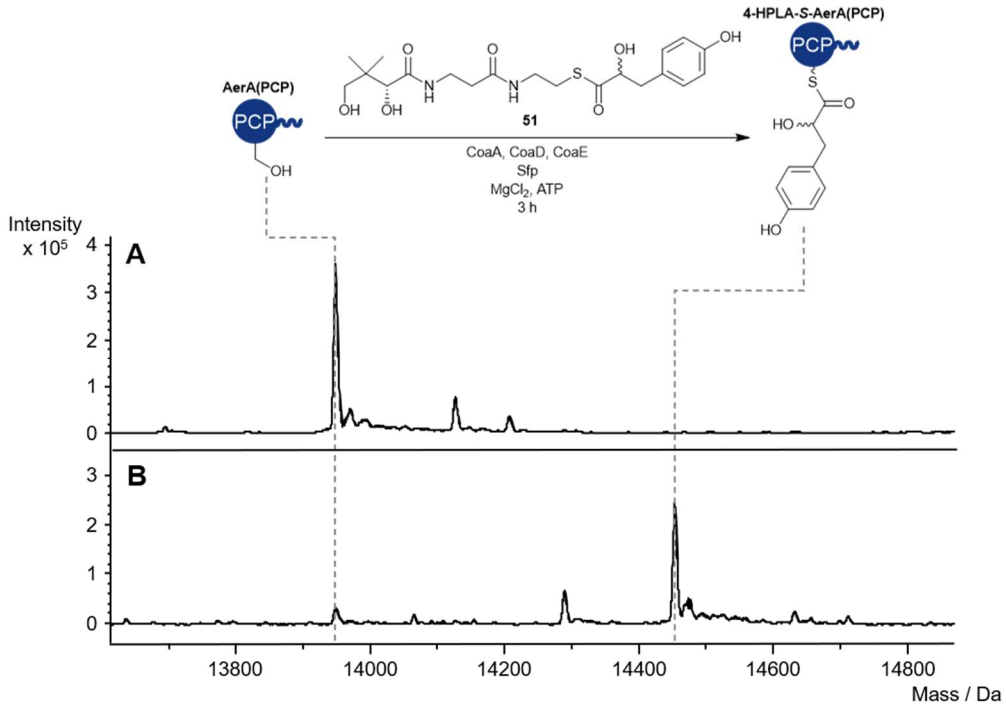


Figure 2.3 Charge state deconvoluted spectra from UHPLC-ESI-Q-TOF-MS analysis demonstrating chemoenzymatic conversion of *apo*-AerA(PCP) to 4-HPLA-S-AerA(PCP) using 4-HPLA-*S*-pantetheine (**51**), enzymes CoaA, CoaD, CoaE, Sfp and relevant cofactors. Dashed lines indicate the calculated mass for each species. The blue star (*) indicates the peak corresponding to *holo*-AerA(PCP) resulting from thioester hydrolysis.

University of Warwick. This one-pot methodology circumvents the need to synthesise the CoA thioester derivatives directly, which can often be expensive owing to the high cost of commercial coenzyme A as well as challenging due to the intrinsic instability of CoA derivatives.[168] All four loading enzymes were overproduced from the appropriate pET vector introduced into *E. coli* BL21(DE3) cells and purified as His_6-tagged proteins by nickel affinity chromatography.

Chemoenzymatic synthesis of 4-HPLA-*S*-AerA(PCP) was performed using this methodology. Following incubation for 3 h at ambient temperature, loading of the desired substrate was confirmed by UHPLC-ESI-Q-TOF-MS analysis (figure **2.3**). A 504 Da shift in mass of the AerA PCP domain was observed, consistent with the covalent addition of 4-HPLA-*S*-phosphopantetheine (**59**). Only minor hydrolysis of the thioester linkage to give *holo*-AerA(PCP) was displayed, identified by a 164 Da reduction in mass. This loading reaction was used for all subsequent production of 4-HPLA-*S*-AerA(PCP). Attempts to store the loaded AerA PCP domain resulted in accumulation of *holo*-AerA(PCP)

resulting from hydrolysis of the thioester linkage. As a result, 4-HPLA-*S*-AerA(PCP) was prepared immediately before use in subsequent assays.

2.1.4. Demonstrating AerJ halogenase activity towards 4-HPLA-*S*-AerA(PCP)

After demonstrating that 4-HPLA-*S*-AerA(PCP) could be generated *in situ* as a substrate for the flavin-dependent halogenase AerJ, reconstituting halogenase activity was explored. To generate the $FADH_2$ cofactor required for halogenase activity, the reaction mixture was supplemented with FAD, NADH and a flavin reductase. Initially, flavin reductase SsuE from *E. coli* K12, which had been used in previous *in vitro* reconstitutions of flavin-dependent halogenases,[40] was trialled.[169] The expression construct for SsuE was acquired from former Challis group member Dr. Paulina Sydor. However, following overproduction of SsuE from BL21(DE3) cell lines (figure **2.4 A**), the protein was found to be highly insoluble in typical 20 mM Tris reaction buffers in the pH range 6.8-8.0, and only low concentrations could be recovered after Ni-affinity purification. Instead, flavin reductase Fre from *E. coli* K12 (pET expression construct provided by Dr. Binuraj Menon, University of Warwick) was overproduced as a soluble protein in *E. coli* BL21(DE3), purified by Ni-NTA affinity chromatography (figure **2.4 B-D**) and utilised to reduce FAD in subsequent experiments.[170,171]

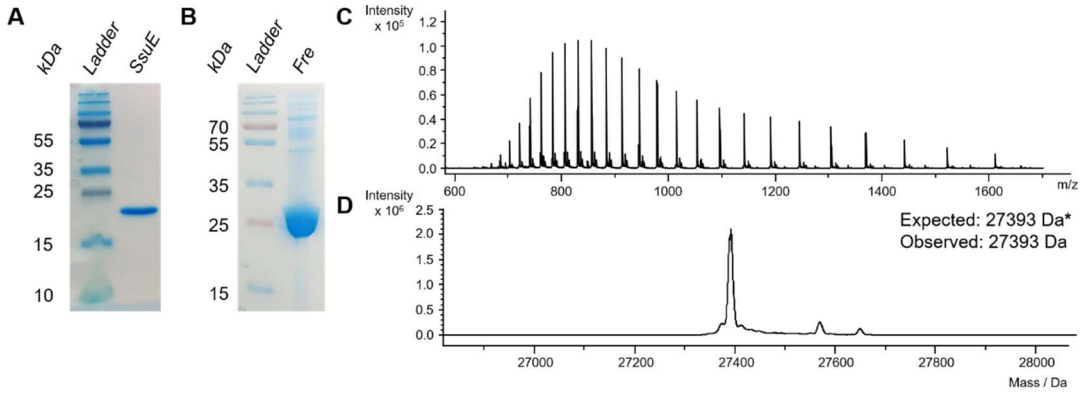


Figure 2.4 Confirmation of identity of recombinant flavin reductases. (**A**) SDS PAGE analysis of SsuE following Ni-affinity chromatography. (**B**) SDS PAGE analysis of Fre following Ni-affinity chromatography. (**C**) Mass spectrum of Fre from UHPLC-ESI-Q-TOF-MS analysis. (**D**) Charge state deconvoluted mass spectra of Fre. Note that the expected mass value corresponds to the mass calculated for the expected protein derived from the Fre expression construct with the initial methionine removed. N-terminal methionine cleavage in *E. coli* is known to be catalysed by methionine aminopeptidases, of which one can be found in the *E. coli* BL21 genome sequence (Genbank accession number: CP053601.1).

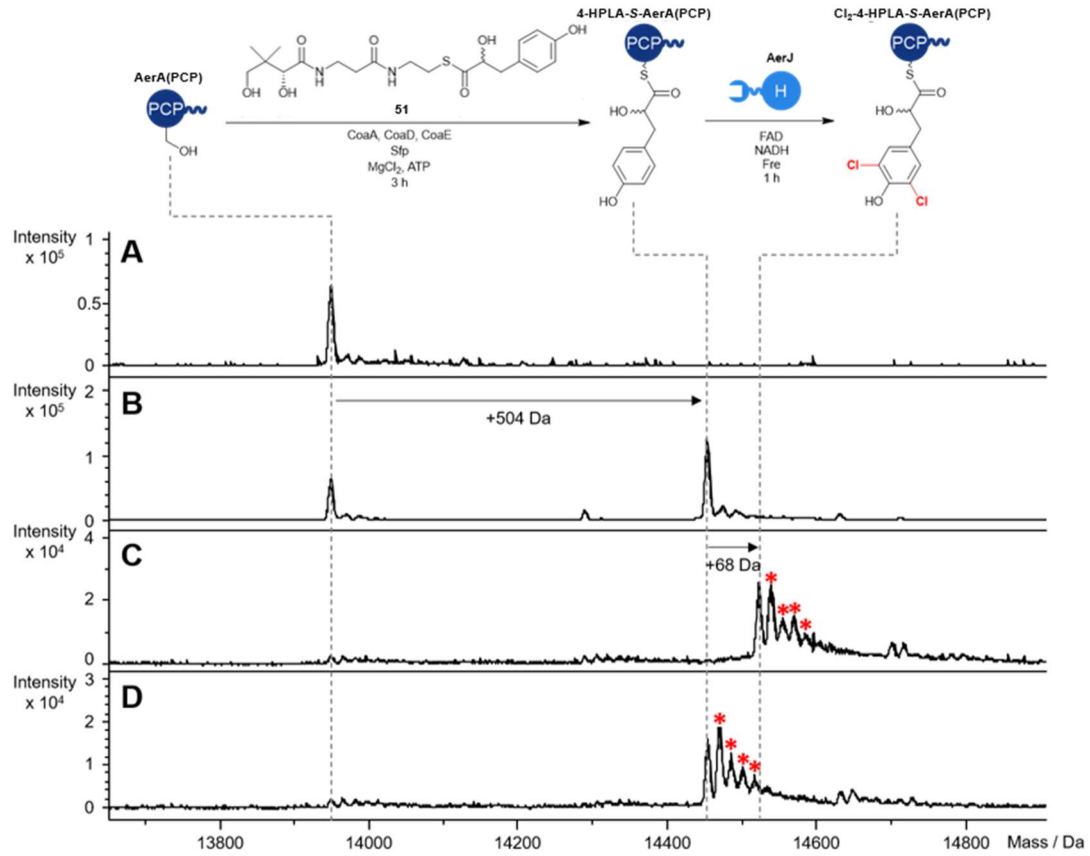


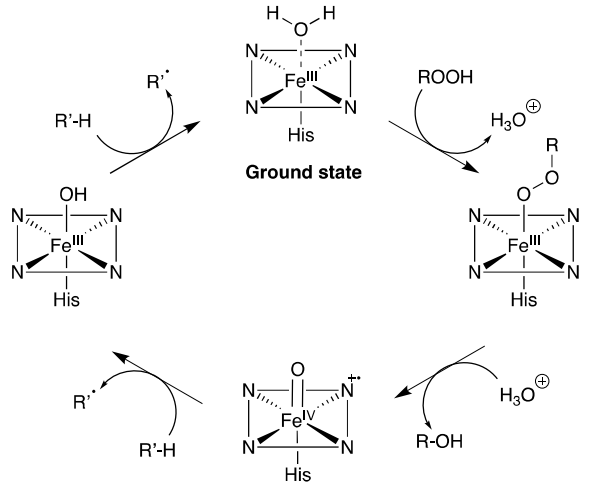
Figure 2.5 Initial attempt to reconstitute of AerJ flavin-dependent halogenase activity *in vitro* monitored by UHPLC-ESI-Q-TOF-MS analysis. Spectra show: (**A**) *apo*-AerA(PCP); (**B**) 4-HPLA-*S*-AerA(PCP) following loading of **51**; (**C**) Cl_2-4-HPLA-*S*-AerA(PCP) following incubation of reaction (**B**) with flavin-dependent halogenase AerJ and cofactors; (**D**) Product of control reaction, in which the reaction in (**C**) was repeated in the absence of halogenase AerJ. Red asterisks (*) indicate a repeating series of peaks with a +16 or +17 Da mass shift which follow the peak corresponding to Cl_2-4-HPLA-*S*-AerA(PCP).

To reconstitute halogenase activity of AerJ *in vitro*, 4-HPLA-*S*-AerA(PCP) (50 µM) was generated *in situ* and then incubated for 1 h with Fre (10 µM), AerJ (10 µM) and an excess of FAD and NADH. The charge state deconvoluted mass spectrum from intact protein UHPLC-ESI-Q-TOF-MS analysis showed an approximately +68 Da shift in mass of 4-HPLA-*S*-AerA(PCP) following introduction of the halogenase, consistent with substitution of two hydrogen atoms with chlorine atoms (figure **2.5**). Additionally, in a control reaction performed in the absence of the halogenase, the +68 Da mass shift was not observed. This provided preliminary evidence that AerJ could dichlorinate the substrate tethered to the AerA PCP domain.

Noticeably, a sequence of +16 or +17 Da peaks following the expected peak for Cl_2-4-HPLA-*S*-AerA(PCP) species were also detected in the deconvoluted

mass spectrum suggesting addition of an oxygen atom or hydroxyl group. It was therefore plausible that oxidation of the protein was occurring during the reaction, potentially caused by reactive oxygen species generated during the reaction.

Initially, it was thought that the proposed oxidation could be a result of reactive oxygen species generated by the halogenase AerJ. As discussed in section 1.4.3, the consensus flavin-dependent halogenase mechanism is thought to proceed *via* a peroxyflavin species which is proposed to oxidise chloride ions to hypochlorous acid. The structurally and mechanistically related flavin-dependent monooxygenase family of enzymes are also thought to utilise a peroxyflavin intermediate and catalyse oxidative reactions such as hydroxylations, epoxidations and Baeyer-Villiger oxidations.[172] However, the same pattern of oxidation peaks is present in the control lacking the halogenase (figure **2.5 D**), suggesting that the halogenase is not responsible for the oxidation observed. Oxidation is therefore likely derived from the introduction of FAD, NADH and flavin reductase Fre. It is known that reduced flavins can spontaneously react with molecular oxygen to generate hydrogen peroxide and superoxide,[173,174] and flavin reductases have been shown to be responsible for both *in vitro* and *in vivo* generation of hydrogen peroxide.[175] The site of oxidation of 4-HPLA-*S*-AerA(PCP) remains cryptic. In figure **2.5 C-D**, residual *apo*-AerA(PCP) also has this pattern of oxidation peaks, which suggests oxidation of the protein is occurring on the protein rather than on the 4-HPLA-*S*-phosphopantheine arm. Typical candidate residues for oxidation of proteins would be surface cysteines and methionines, owing to the variable oxygen state of their sidechain sulfur atoms.[176,177] In the AerA PCP domain amino acid sequence there is only one methionine residue and no cysteine residues, but oxidation appears to occur up to 4 times based on the charge-state deconvoluted mass spectra (figure **2.5 C**). To probe this oxidation further, the halogenation reaction could be carried out and followed by peptide mass fingerprinting[178] to gain residue-level understanding of the oxidation reactions taking place.



Scheme 2.4 Mechanism for scavenging of peroxide species (ROOH) by peroxidases. The heme cofactor is represented as an abbreviated iron complex. A sacrificial substrate (R'-H) is oxidised and provides a source of protons to allow restoration to the ground state.

It was thought that addition of a peroxidase to scavenge any potential reactive oxygen species from solution may be able to supress oxidation. Peroxidases utilise a heme group to transfer electrons to peroxide species, typically hydrogen peroxide (scheme **2.4**).[179] Horseradish peroxidase type I was purchased from Sigma-Aldrich and the halogenation assays described previously in figure **2.5**. were repeated. Introduction of peroxidase into the reaction at a concentration of 10 μM supressed the proposed oxidation peaks, with only small amounts of a singly-oxidised product peak being observed (figure **2.6**) whilst activity of the halogenase was preserved. Consequently, peroxidase was used in all future halogenation reactions to minimise formation of oxidised products, which could potentially obscure or confuse results.

In vitro reconstitution of the halogenase activity with peroxidase corroborated the results of those without. Interestingly, it was clear from these results that only dichlorinated product could be observed after 1 hour of incubation. This contrasts with the aeruginosin congeners produced by *M. aeruginosa* NIES-98, which produce non- and mono- chlorinated products alongside the di-chlorinated species. It is likely the concentrations of the proteins utilised in assay conditions are much higher than in an *in vivo* context, contributing to the complete turnover to the dichlorinated product. It would be expected that the second chlorination would require a greater activation energy that the first, as the addition of an electron-withdrawing chlorine atom would deactivate the

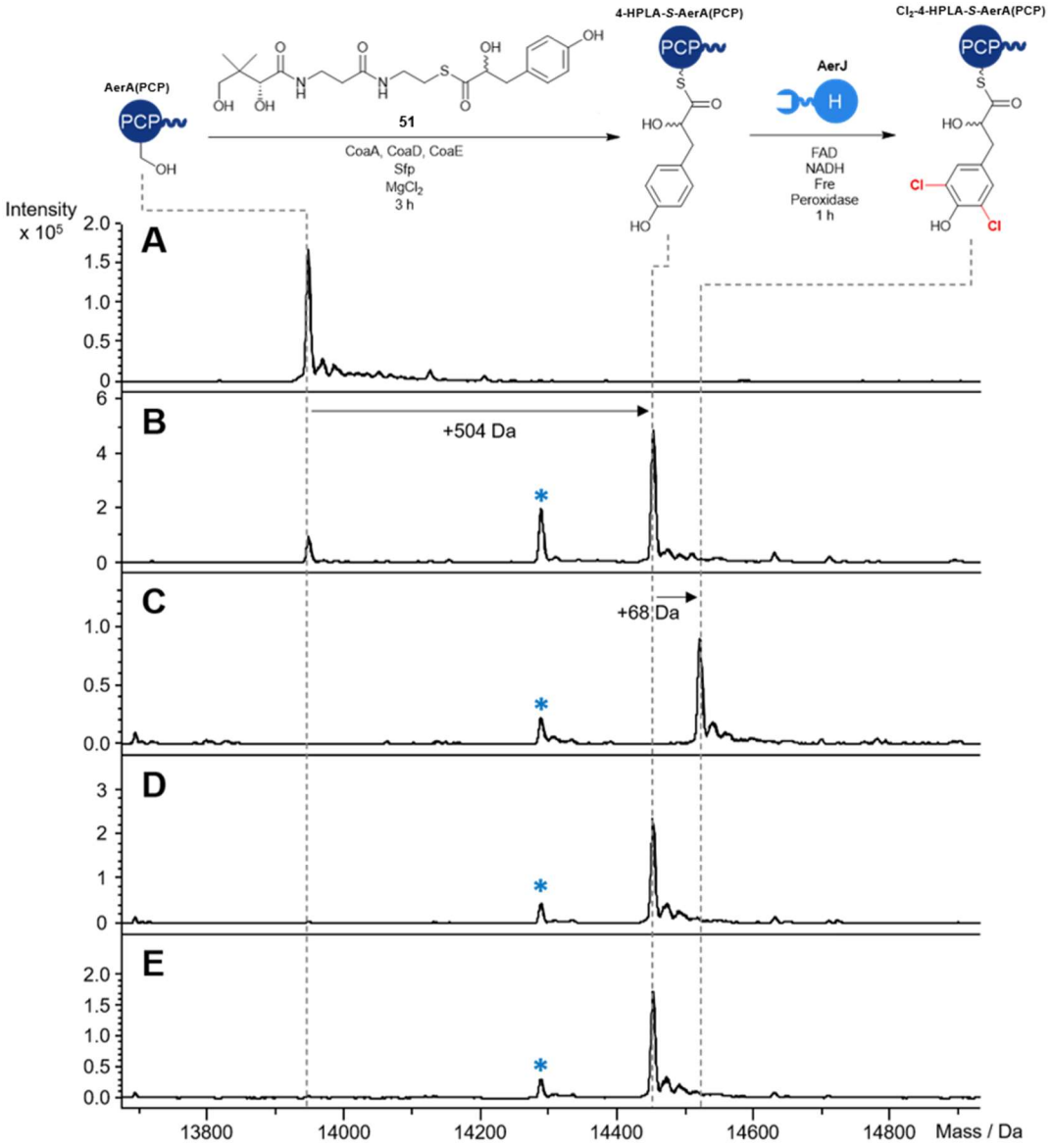


Figure 2.6 Reconstitution of AerJ flavin-dependent halogenase activity *in vitro* with horseradish peroxidase type I monitored by UHPLC-ESI-Q-TOF-MS analysis. Spectra show: (**A**) *apo*-AerA(PCP); (**B**) 4-HPLA-*S*-AerA(PCP) following loading of **51**; (**C**) Cl_2-4-HPLA-*S*-AerA(PCP) following incubation of reaction (**B**) with flavin-dependent halogenase AerJ, cofactors and peroxidase; (**D**) Product of control reaction in which the reaction in (**C**) was repeated in the absence of halogenase AerJ; (**E**) Product of control reaction in which the reaction in (**C**) was repeated with AerJ replaced by active site point mutant K156A. Blue asterisks (*) indicate *holo*-AerA(PCP) resulting from thioester hydrolysis.

phenol ring towards electrophilic aromatic substitution. However, rigorous quantitative kinetics of chlorination were beyond the scope of this study and were not investigated further. Furthermore, complete turnover to di-chlorinated 4-HPLA-*S*-AerA(PCP) demonstrates that both stereochemical configurations of the α-hydroxyl group of 4-HPLA were tolerated by the halogenase, as the synthesised pantetheine thioester derivative **51** contained a mixture of

epimers. This gives initial indications that AerJ may tolerate non-native AerA PCP domain-tethered substrates

To further confirm the halogenase is catalysing the chlorination, a point mutant of AerJ where the active site lysine was mutated to an inert alanine residue was generated. The active site lysine was identified by sequence alignment of AerJ with flavin-dependent halogenases that act upon phenolic substrates (figure **2.7 A**). Lysine 156 in the AerJ construct (corresponding to K139 in the native protein with no affinity tag) aligned with K87 in PltM and K74 in RadH, which have been experimentally validated by mutagenesis studies to be essential for halogenation,[90,180] as well as K74 in Rdc2 and K72 in CmlS.[92,181] Additionally, PltM and CmlS both have reported X-ray crystal structures (PDB accession: 6BZA and 3I3L respectively), where the aligned lysine is directed towards the centre of the active site (figure **2.7 B-C**), and in the case of PltM towards the co-crystalised substrate, phloroglucinol (figure **2.7 B**). These observations provide a high degree of certainty of the aligned lysine residues being the desired active site residue.

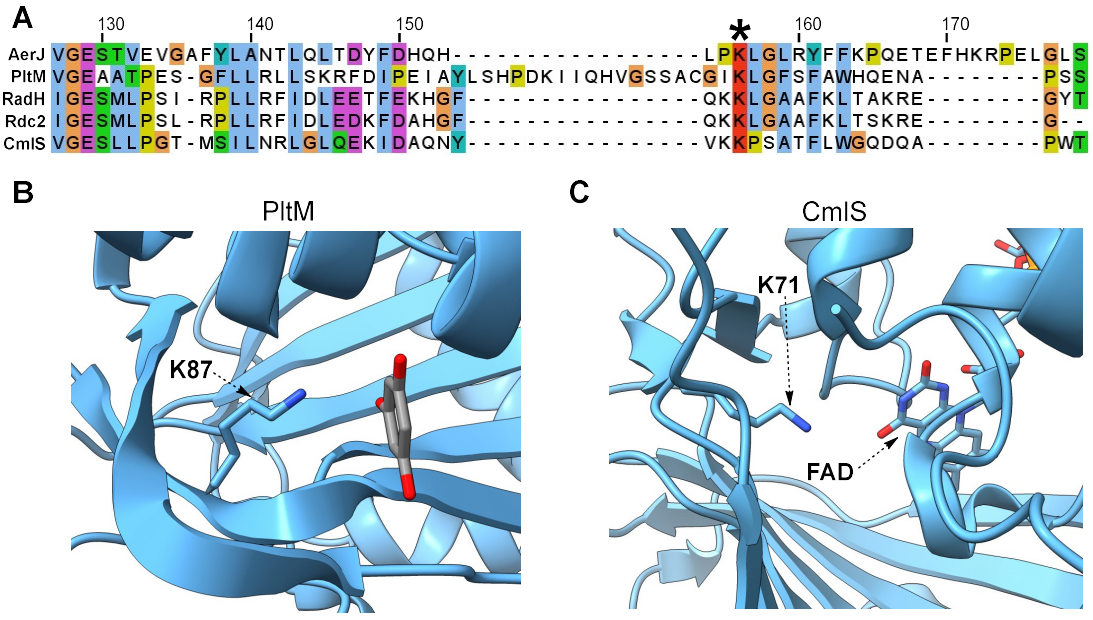


Figure 2.7 Identification of the active site lysine of flavin-dependent halogenase AerJ. (**A**) Sequence alignment of AerJ with known flavin-dependent halogenases that also act on phenolic substrates. An asterisk (*) denotes the active site lysine, which is experimentally verified for PltM and RadH. Residue numbering provided is relative to the AerJ sequence. (**B**) The X-ray crystal structure of the PltM active site (PDB accession: 6BZA). Active site lysine K87 is directed towards the co-crystalised substrate, phloroglucinol (grey). (**C**) The X-ray crystal structure of the CmlS active site (PDB accession: 3I3L). Lysine K71 is directed towards the same channel as the FAD isoalloxazine ring.

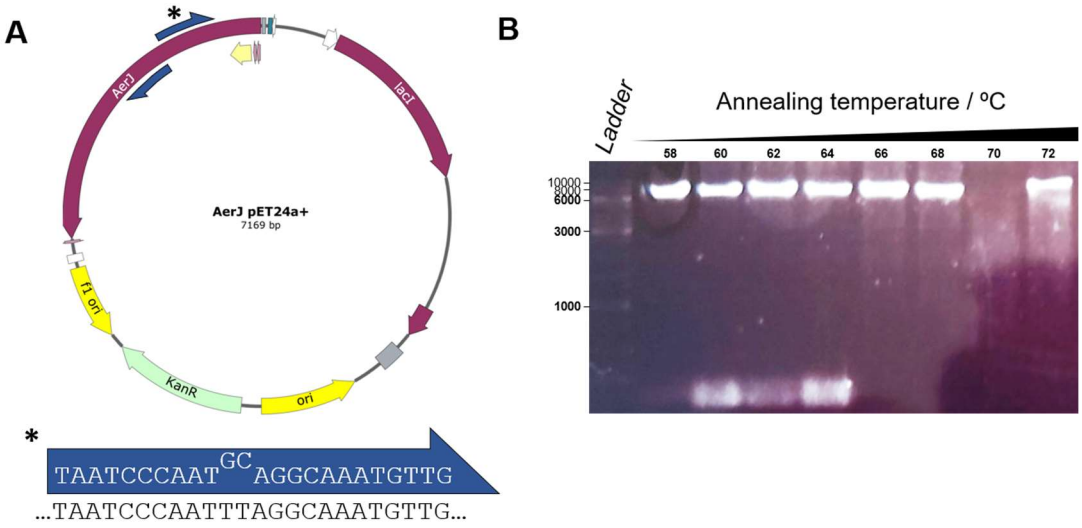


Figure 2.8. Mutagenesis to generate point mutant pET24a-(+)-AerJ(K156A). (**A**) Back-to-back design of primers for substitution of two bases. (**B**) Agarose gel electrophoresis displaying PCR products resulting from mutagenesis. Expected size = 7.2 kb.

Mutagenesis of K156 to alanine was performed using a Q5 site-directed mutagenesis kit (New England Biolabs). Back-to-back primers were designed to amplify around the pET24a-(+)-AerJ plasmid with replacement of the lysine codon with an alanine-encoding codon. PCR was carried out according to the manufacturer's instructions, with a temperature gradient during the annealing step of ± 6 °C of the calculated annealing temperature (figure **2.8**). The resulting PCR product was blunt-end ligated using a commercial KLD enzyme mixture (New England Biolabs). KLD mix contains a kinase and ligase for rapid phosphorylation and circularisation of the PCR product, and DpnI, a restriction endonuclease which only acts on DNA with methylated adenine bases, which can be found in template plasmids which have been passed through *E. coli* but not PCR-generated DNA.[182] The final plasmid was amplified in TOP10 *E. coli* and successful mutagenesis was confirmed by Sanger sequencing (Eurofins Scientific). Following overproduction and purification of AerJ K156A by Ni-affinity chromatography, the mutation was found to abolish turnover of the halogenase, with neither mono- or di-chlorinated 4-HPLA-*S*-AerA(PCP) being observed (figure **2.6 E**).

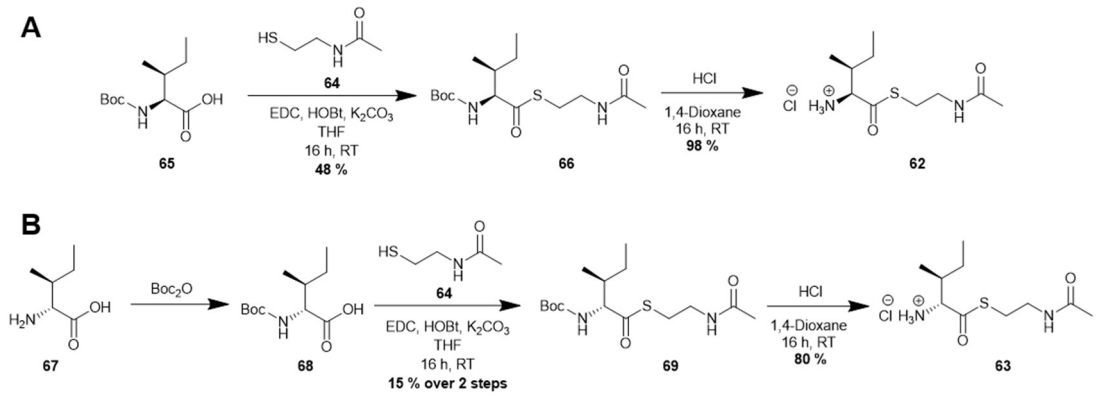
2.2. *In vitro* reconstitution of AerB condensation domain activity

Following successful demonstration of AerJ halogenase activity, it was desirable to also demonstrate the AerB condensation domain activity with a view to understanding competition between the two βHDD-bearing proteins. Condensation domains have been extensively studied and their reconstitution *in vitro* has been accomplished numerous times.[183] To achieve this, a condensation acceptor and donor are required which can either be carrier protein-tethered or a small molecule carrier protein thioester mimic. As a procedure to generate 4-HPLA-*S*-AerA(PCP), which can act as the condensation donor, had been established in section 2.1.3, a condensation acceptor was required.

2.2.1. Attempted reconstitution of AerB C domain activity using *N*-acetyl cysteamine (NAC) thioesters

Initially, it was thought a facile method for observing condensation domain activity would be to use *N*-acetyl cysteamine (NAC) thioesters as diffusible probes. These molecules can be easily synthesised with a variety of thioester-appended moieties and mimic the terminus of a substrate-tethered pantetheine arm.[184] NAC thioesters have been shown to be important substrates for investigation of the activity of a variety of NRPS and PKS catalytic domains.[184] Of relevance to reconstitution of the AerB C domain activity, aminoacyl-*S*-NACs have been shown to be readily accepted as both condensation donors and acceptors for *in vitro* assays of C domain activity.[185]

Prediction of the substrate specificity of the AerB A domain using the online PKS/NRPS predictor tool suggested L-isoleucine would be the amino acid adenylated and loaded onto the AerB PCP domain.[186] However, the amino acid incorporated into aeruginosin is D-allo-isoleucine, with the E domain of AerB catalysing the epimerisation. Current biochemical evidence suggests that epimerisation at the α-carbon occurs following condensation.[187] Nevertheless, both the L-isoleucyl-*S*-NAC (**62**) and D-allo-isoleucyl-*S*-NAC (**63**) were targeted as synthetic substrates that could act as condensation donor to investigate C domain activity.



Scheme 2.5 Synthesis of: (**A**) L-Ile-*S*-NAC hydrochloride **62**; (**B**) D-*allo*-Ile-*S*-NAC hydrochloride **63**.

62 and **63** were synthesised as their hydrochloride salts using the routes shown in scheme **2.5**. *N*-acetyl cysteamine (**64**) and Boc-L-isoleucine (**65**) were purchased from commercial sources whilst Boc-D-*allo*-isoleucine (**68**) was synthesised by reacting D-*allo*-isoleucine with Boc anhydride under basic conditions. **64** was coupled to **65** and **68** using amide coupling reagent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and 1-hydroxybenzotriazole as a rate-enhancing agent yielding the corresponding Boc-protected isoleucyl-*S*-NACs **66** and **69**. Trifluoroacetic acid (TFA)-catalysed deprotection of the Boc protecting groups was unsuccessful, presumably as the strong acidic conditions catalysed rapid hydrolysis of the labile thioester bond. Using milder deprotection conditions of 0.5 M HCl in 1,4-dioxane yielded the appropriate isoleucyl-*S*-NAC hydrochloride salts **62** and **63** in good yield. Analysis of the ^{1}H NMR spectrum showed only a single diastereomer was observed in each case, suggesting minimal racemisation of the α-carbon had occurred during the synthesis.

The AerB C domain was overproduced using a pET24a-(+) construct originally designed by Dr. Matt Jenner and Dr. Sławomir Potocki and synthesised and cloned by Epoch Life Sciences Inc. The construct contained the same custom His_8 tag as AerJ and AerA(PCP) constructs described in section 2.1.2. The nucleotide sequence was taken from the deposited aeruginosin biosynthetic gene cluster (GenBank accession: FJ609416.1), and the C-terminal C domain boundary was determined using both conserved domain and protein family (Pfam) database searches using their respective webservers.[160,188] The

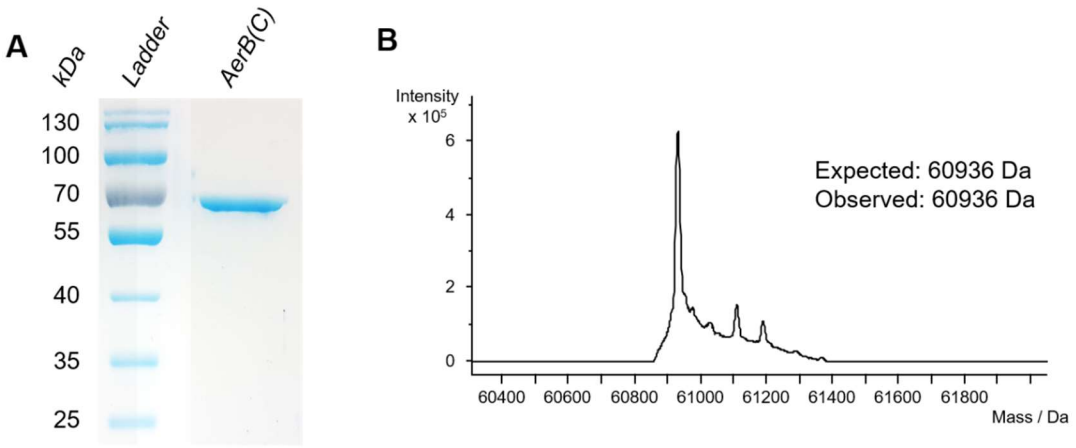


Figure 2.9 Analysis of recombinant AerB C domain following nickel affinity chromatography purification. (**A**) SDS-PAGE analysis of the AerB C domain. Expected mass = 60.9 kDa. (**B**) Charge-state deconvoluted mass spectrum from UHPLC-ESI-Q-TOF analysis of the AerB C domain confirming identity of the purified protein.

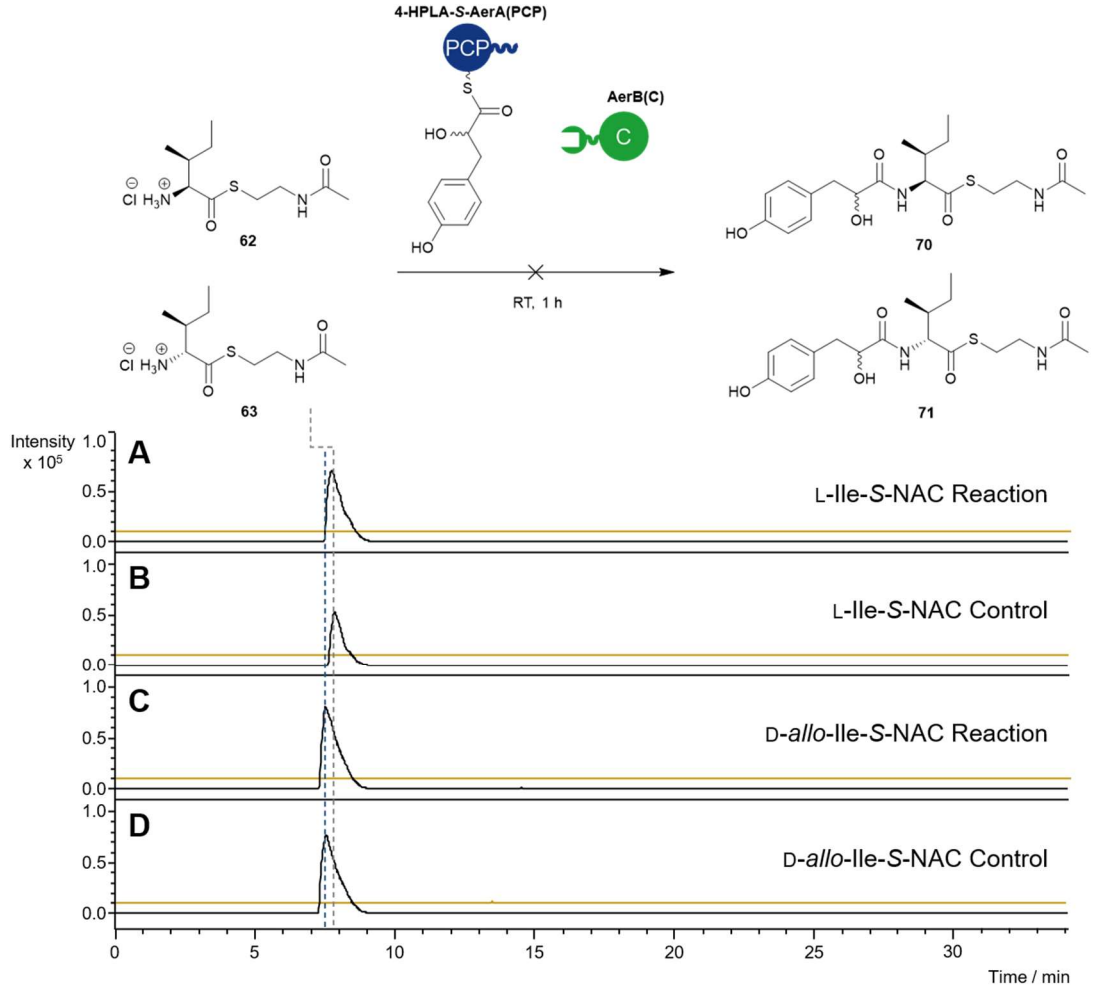


Figure 2.10 Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analysis of attempts to reconstitute AerB(C) activity with acyl-*S*-NAC substrates **62** and **63**. In all chromatograms: black EIC = 255.1143 ± 0.05 Da (**62** or **63** $[M+H]^+$); yellow EIC = 397.1917; 419.1617 ± 0.05 Da (**70** or **71** $[M+H]^+$ and $[M+Na]^+$). (**A**) The reaction of L-Ile-*S*-NAC (**62**) with 4-HPLA-*S*-AerA(PCP) and the AerB C domain. (**B**) Control reaction, where the AerB C domain in reaction (**A**) was replaced by buffer. (**C**) The reaction of D-*allo*-Ile-*S*-NAC (**63**) with 4-HPLA-*S*-AerA(PCP) and the AerB C domain. (**D**) Control reaction, where the AerB C domain in reaction (**C**) was replaced by buffer. Blue dashed line = **62**, grey dashed line = **63**.

protein was overproduced and purified by Ni-affinity chromatography (figure 2.9).

62 and 63 were incubated for 1 h with the AerB C domain and 4-HPLA-*S*-AerA(PCP) prepared as in section 2.1.3. Then, methanol was added to precipitate proteins which were then pelleted by centrifugation. UHPLC-ESI-Q-TOF-MS analysis of the supernatant detected only the starting material NAC thioesters 62 and 63, with neither elongated product (70 or 71) observed (figure 2.10). These results suggested that the AerB C domain does not readily accept SNAC substrates, and indicates that interactions between the AerB C domain and PCP domain are critical for condensation to occur.

2.2.2. Reconstitution of AerB C domain activity using the full AerB subunit

To overcome this issue, an assay was developed that used isoleucyl-loaded AerB as the condensation acceptor, so the reaction could be monitored *via* intact protein mass spectrometry. Synthesis and loading of amino acyl pantetheine thioesters has traditionally been challenging, with only one currently reported example in the literature.[154] Instead, the full length AerB subunit was overproduced, and the native AerB A domain used to load L-isoleucine.

The full-length *aerB* gene from the aeruginosin gene cluster (GenBank accession: FJ609416.1) was designed with the same N-terminal His_8-tag as in the AerJ, AerA(PCP) and AerB(C) constructs, with the TEV cleavage site replaced with a thrombin cleavage site. This was synthesised and cloned BamHI to NdeI into expression vector pET24a-(+) by Epoch Life Sciences Inc. Initial trials of overexpression and nickel ion affinity chromatography showed soluble overproduction of a protein with the desired molecular weight, 183 kDa (figure 2.11 A). However, this protein co-purified with proteins of a smaller molecular weight (approximately 60-100 kDa). These smaller proteins likely correspond to C-terminal truncations of the AerB subunit, with an intact His_8 tag, resulting either from truncated expression of the gene or degradation of

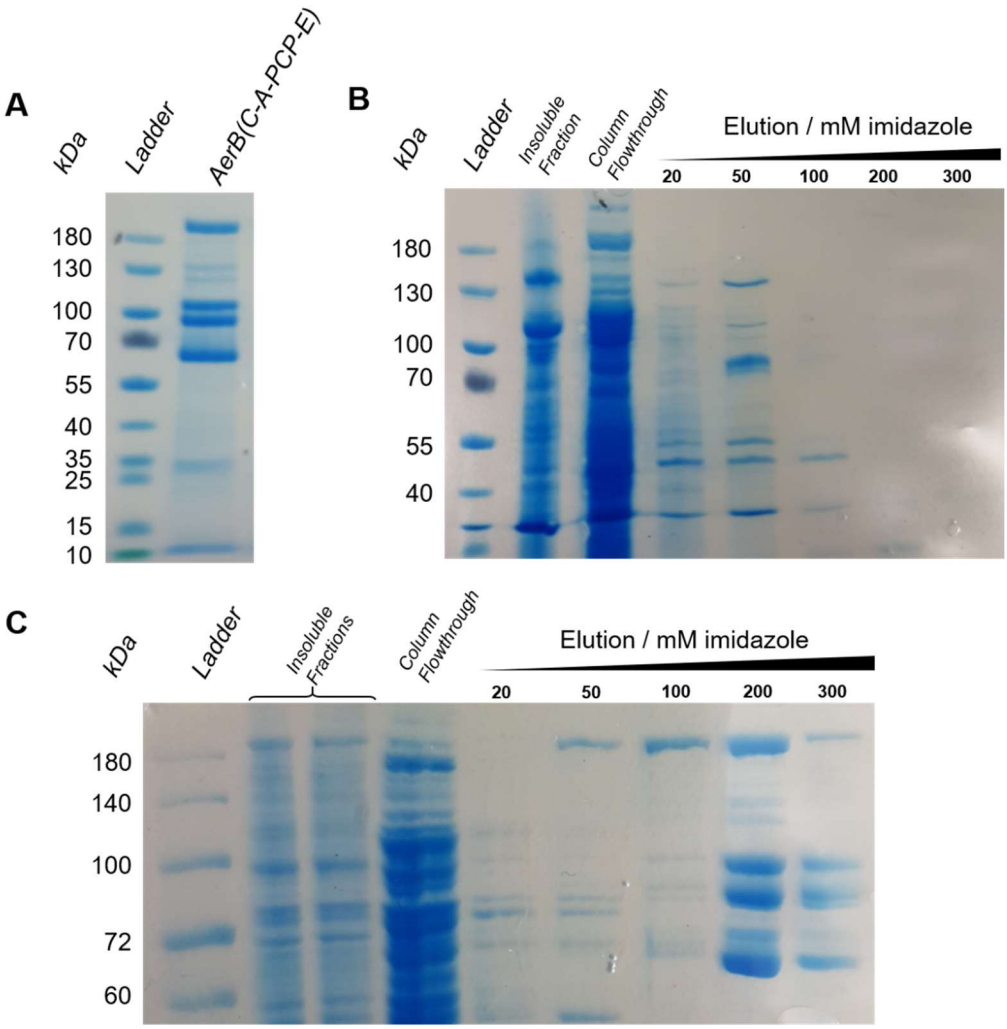


Figure 2.11 SDS PAGE gels demonstrating expression and purification of AerB(C-A-PCP-E). (**A**) AerB(C-A-PCP-E) following overexpression in *E. coli* BL21(DE3) and purification by Ni-affinity chromatography. (**B**) Elution profile from Ni-affinity chromatography of AerB(C-A-PCP-E) overexpressed in *E. coli*. C43(DE3). (**C**) Elution profile from Ni-affinity chromatography of AerB(C-A-PCP-E) overexpressed in *E. coli* BL21(DE3) with all buffers supplemented with cOmplete™ protease inhibitor cocktail.

the protein following translation. The former could be due to the high metabolic load associated with ribosomal synthesis of such a large recombinant protein, whilst the latter could result from proteases either produced by the *E. coli* or found in exogenous sources (such as contaminated buffer or other reagents) causing proteolysis.

To try to eliminate the production of the smaller proteins, expression was trialled in *E. coli* C43(DE3) cells. This *E. coli*. strain is often used in place of the more common BL21(DE3) cell lines for expression of recombinant proteins that may be toxic to *E. coli* and display expression profiles distinct to

BL21(DE3) *E. coli*.[189] However, following overproduction of the protein in *E. coli* C43(DE3), the AerB subunit was no longer observed in SDS-PAGE analysis of the elution fractions from Ni-affinity chromatography (figure **2.11 B**). Similarly, the protein was re-purified from overproduction in *E coli* BL21(DE3) using buffers supplemented with cOmplete™ protease inhibitor cocktail, but this did not lead to reduction in the intensity of the smaller protein bands in SDS-PAGE analysis (figure **2.11 C**).

Instead, the AerB subunit from Ni-affinity chromatography was further purified to near-homogeneity *via* size-exclusion chromatography prior to use in assays (figure **2.12 A-B**). Due to the low initial yield of protein and subsequent loss of

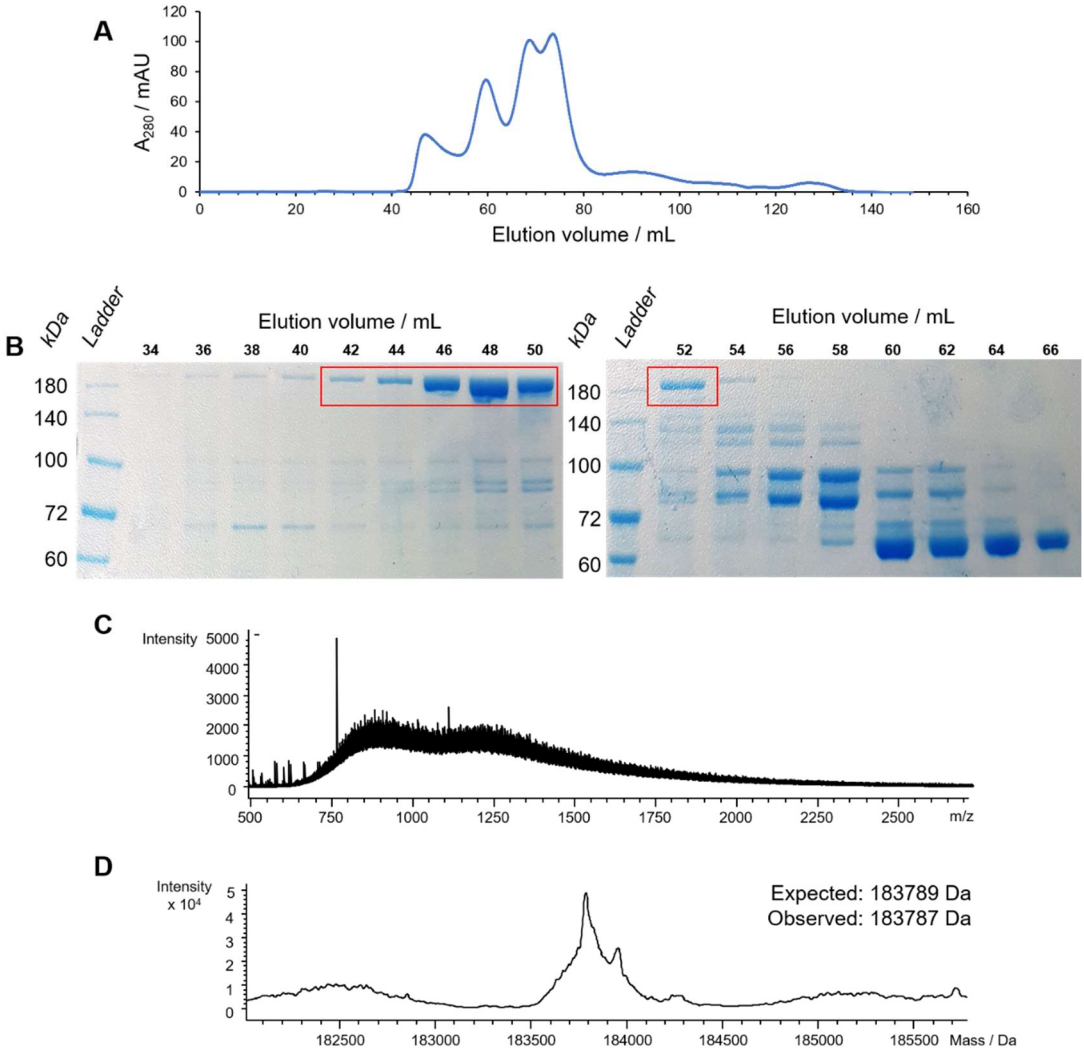


Figure 2.12 Purification by size-exclusion chromatography and confirmation of mass of AerB(C-A-PCP-E). (**A**) Size-exclusion chromatogram of the Ni-affinity purified AerB subunit. (**B**) SDS-PAGE analysis of fraction collected from size-exclusion chromatography. The bands corresponding to AerB(C-A-PCP-E) are highlighted in red, and those highlighted were collected and further concentrated for *in vitro* experiments. (**C**) Mass spectrum of AerB(C-A-PCP-E) from UHPLC-ESI-Q-TOF-MS analysis. (**D**) Charge-state deconvoluted mass spectra of AerB(C-A-PCP-E) from UHPLC-ESI-Q-TOF-MS.

samples during several round of purification, purification from 4 L of LB culture of *E. coli* BL21(DE3) cells was necessary to recover enough recombinant protein for use in *in vitro* reconstitution experiments. UHPLC-ESI-Q-TOF-MS analysis of the protein gave a charge-deconvoluted mass spectra displaying a peak correlating with the expected mass of AerB(C-A-PCP-E) (figure **2.12 C-D**). Moreover, resolution of this peak suggested detecting the mass changes for the condensation reaction would be plausible using intact mass spectrometry of this protein.

AerB(C-A-PCP-E) was first converted to its *holo*-form by incubation with coenzyme A, phosphopantetheinyl transferase Sfp and $MgCl_2$ (figure **2.13 A-B**). After confirmation of conversion by UHPLC-ESI-Q-TOF-MS, excess ATP and L-isoleucine were added for activation by the adenylation domain. 3 h at ambient temperature gave complete conversion to the isoleucine-loaded protein, as determined by the +113 Da shift in the deconvoluted intact protein mass spectrum (figure **2.13 C**). Finally, 4-HPLA-*S*-AerA(PCP), prepared as in section 2.1.3, was introduced in a 1:1 stoichiometry with L-Ile-S-AerB(C-A-PCP-E) with concentrations determined by A_{280} measurements using a Nanodrop spectrophotometer. The deconvoluted intact protein MS demonstrated a mass shift of 164 Da, consistent with condensation of the 4-HPLA moiety (figure **2.13 D**). Approximately 50 % conversion was achieved in 1 hour. A control reaction was conducted in which all the required substrates, cofactors and enzymes for generation of 4-HPLA-*S*-AerA(PCP) excluding the AerA PCP domain were preincubated then introduced to L-Ile-*S*-AerB(C-A-PCP-E). No condensation was observed, demonstrating the carrier protein is critical for condensation and the AerB C domain cannot accept residual 4-HPLA-*S*-pantetheine (**51**) or 4-HPLA-*S*-CoA (**61**) in solution (figure **2.13 E**). This could be because protein-protein interactions between the two subunits induce a conformational change in the condensation domain, allowing access to the acceptor channel by the AerA PCP domain-tethered substrate. A similar hypothesis was concluded from both structural studies and molecular dynamics simulations of the βHDD-bearing condensation domain Bamb_5915 from enacyloxin IIa (**35**) biosynthesis.[123] Structural studies of the AerB C

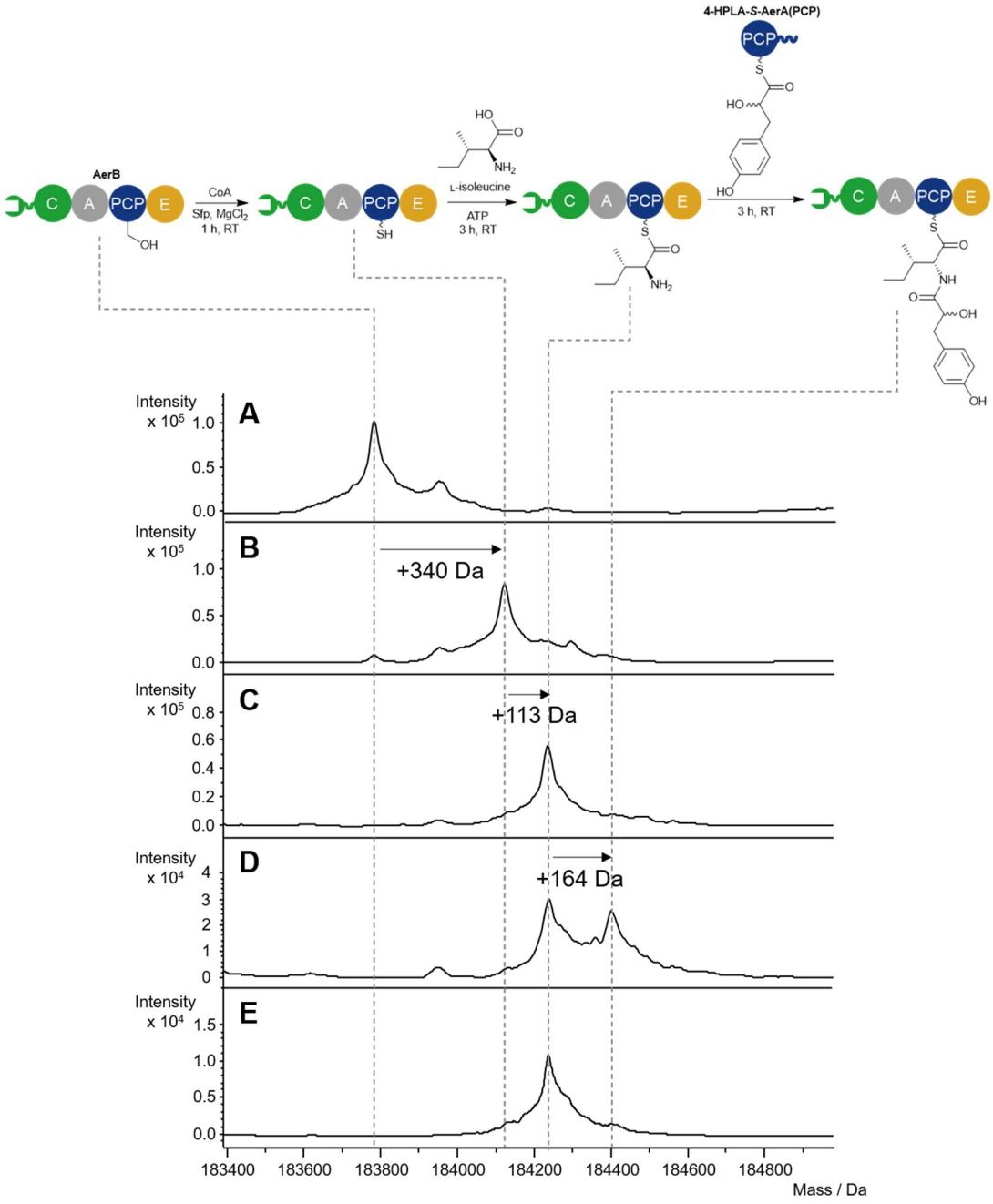


Figure 2.13 Reconstitution of AerB condensation activity *in vitro* monitored by UHPLC-ESI-Q-TOF-MS analysis. (**A**) *apo*-AerB(C-A-PCP-E). (**B**) *holo*-AerB(C-A-PCP-E) following incubation with CoA, Sfp and $MgCl_2$. (**C**) L-Ile-S-AerB(C-A-PCP-E) following incubation of reaction (**B**) with L-isoleucine and ATP. (**D**) Condensation product following incubation of (**C**) with 4-HPLA-*S*-AerA(PCP). (**E**) Product of the reaction of (**C**) where the loading reaction used to generate 4-HPLA-*S*-AerA(PCP) was conducted in the absence of AerA(PCP). This control reaction shows that the AerB C domain does not accept residual 4-HPLA-*S*-pantetheine (**51**) or 4-HPLA-*S*-CoA (**61**) as a condensation acceptor.

domain and its interaction with the AerA PCP domain are necessary to confirm this hypothesis.

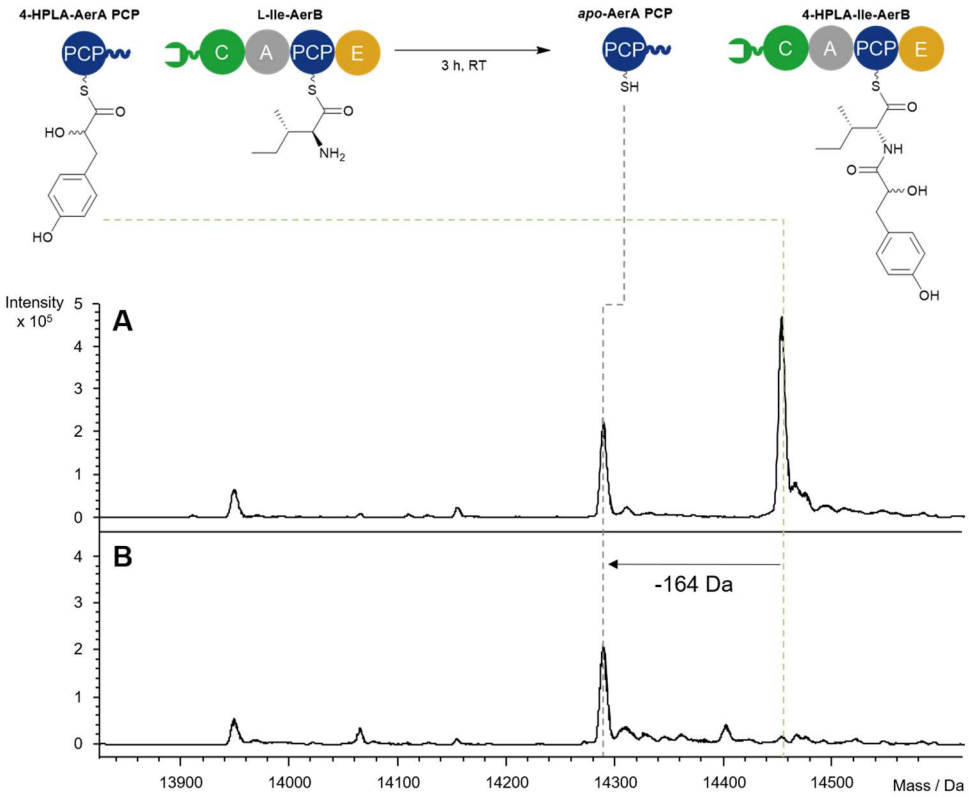


Figure 2.14 UHPLC-ESI-Q-TOF-MS analysis of the AerA(PCP) species before and after condensation with L-Ile-S-AerB(C-A-PCP-E). (**A**) 4-HPLA-*S*-AerA(PCP) from pantetheine loading reaction prior to incubation. (**B**) *holo*-AerA(PCP) following incubation, displaying a loss of 164 Da compared to 4-HPLA-*S*-AerA(PCP).

Full conversion to the condensed product could not be attained during the *in vitro* condensation. However, UHPLC-ESI-Q-TOF-MS analysis of the AerA PCP domain before and after introduction to L-Ile-*S*-AerB(C-A-PCP-E) showed that the carrier protein was fully converted from 4-HPLA-*S*- to *holo*- during the reaction (figure **2.14**). This suggests that L-Ile-*S*-AerB(C-A-PCP-E) was in excess of 4-HPLA-*S*-AerA(PCP) species in the reaction. This could be rationalised by inaccuracies in the calculation of protein concentration from A_{280} measurements or discrepancies in pantetheine thioester stability leading to loss of the 1:1 stoichiometry of the condensation acceptor and donor. To eliminate the possibility of the former, the concentrations of each protein were also determined by a Bradford assay.[190] Comparison of values obtained from both methods were broadly congruent, suggesting the 1:1 stoichiometry used was reasonable. Therefore, the inability to observe full conversion is likely attributable to instability of the pantetheine thioester reducing the effective amount of 4-HPLA-*S*-AerA(PCP) available to act as a condensation donor.

Indeed, in UHPLC-ESI-Q-TOF-MS analysis of 4-HPLA-*S*-AerA(PCP) prior to incubation, approximately 37 % conversion to *holo*- was already observed (figure **2.14 A**) To attempt to observe complete turnover, the ratio of 4-HPLA-*S*-AerA(PCP) relative to L-Ile-*S*-AerB(C-A-PCP-E) could be increased. However, increasing the concentration of 4-HPLA-*S*-AerA(PCP) in the reaction mixture was not attempted to attempt to preserve the MS detector from potential saturation.

2.3. Investigating competition of AerJ and AerB

2.3.1. Attempted *in vitro* reconstitution of a 3-component system

Having established that both AerJ and the AerB C-domain could interact productively with 4-HPLA-*S*-AerA(PCP), the origins of the mixture of non-, mono- and di-chlorinated aeruginosins was investigated *in vitro*. To do this, the reactions of AerJ and the AerB C domain with 4-HPLA-*S*-AerA(PCP) were reconstituted with both proteins present in the same reaction mixture.

Initially, the condensation reaction outlined in section 2.2.2 was conducted in the presence of AerJ and the required cofactors for halogenation, with AerJ in 1:1 stoichiometry with the AerB subunit (figure **2.15 D**). A mass shift of 198 Da was detected in the deconvoluted intact protein MS, consistent with condensation of chlorinated 4-HPLA. However, poor resolution of the deconvoluted spectrum made this result ambiguous. In an attempt to improve signal intensity in the deconvoluted spectrum, 4-HPLA-*S*-AerA(PCP) was incubated with AerJ prior to introduction of L-Ile-*S*-AerB(C-A-PCP-E) to facilitate full conversion to the di-chlorinated species. Upon analysis of the resulting product spectra, peaks with mass shifts of 198 and 232 Da were observed, corresponding to the mono-chlorinated and di-chlorinated products (figure **2.15 E**). However, increase in resolution compared to figure **2.15 D** was minimal, again providing ambiguity as to the degree of chlorination of the pantetheine-tethered substrate. Despite promising initial indications that AerJ

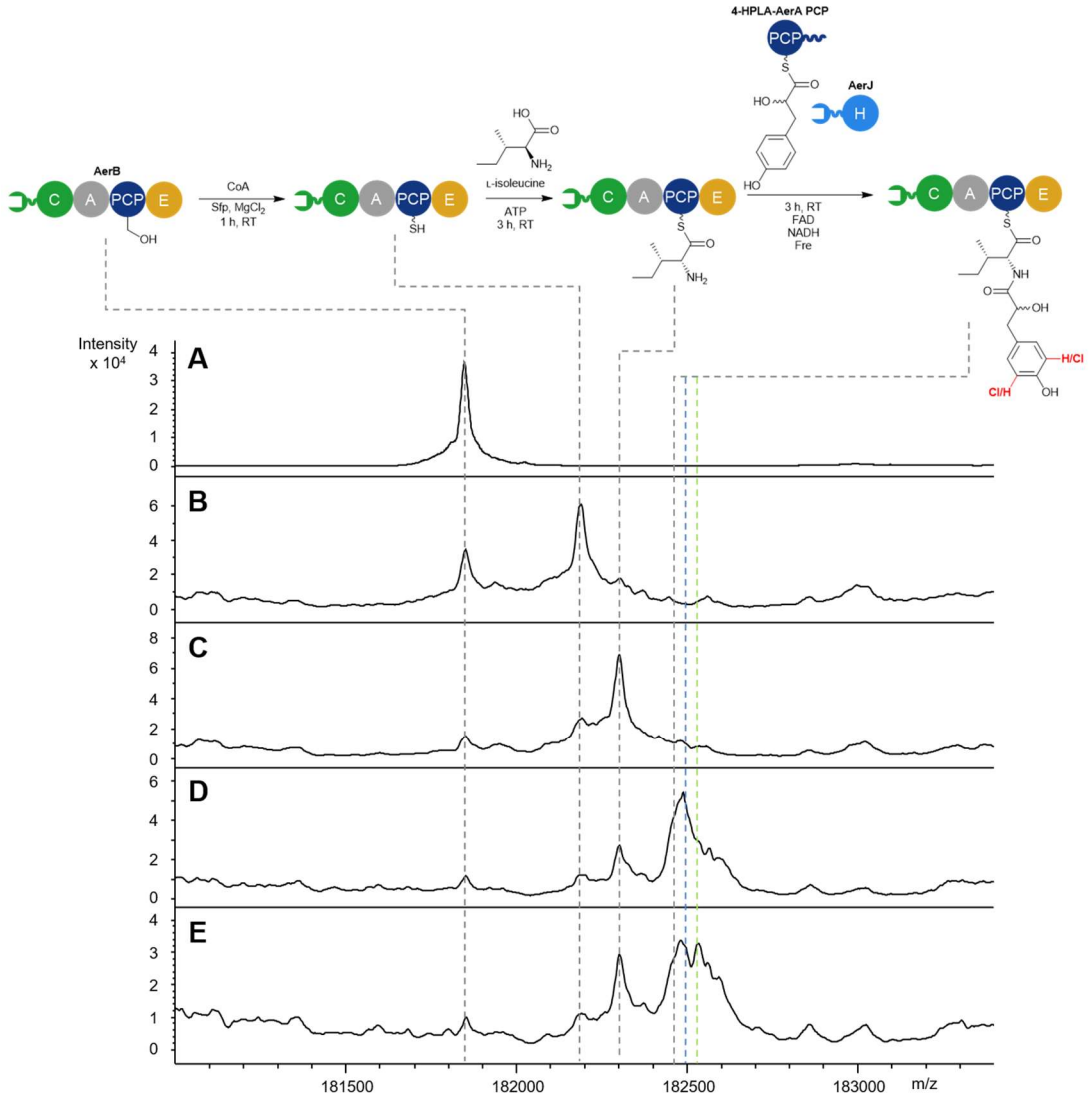


Figure 2.15 Charge-state deconvoluted intact protein mass spectra from UHPLC-ESI-Q-TOF-MS analysis demonstrating attempts to reconstitute simultaneous halogenase and C domain activity. (**A**) *apo*-AerB(C-A-PCP-E). (**B**) *holo*-AerB(C-A-PCP-E) following incubation with CoA, Sfp and $MgCl_2$. (**C**) L-Ile-S-AerB(C-A-PCP-E) following incubation of reaction (**B**) with L-isoleucine and ATP. (**D**) Condensation product following incubation of (**C**) with 4-HPLA-*S*-AerA(PCP) and AerJ and relevant cofactors simultaneously. (**E**) Condensation product following incubation of (**C**) with 4-HPLA-*S*-AerA(PCP) which had been pre-incubated with AerJ and relevant cofactors prior to condensation. Blue dashed line = expected mass of the mono-chlorinated product, green dashed line = expected mass of the di-chlorinated product.

and the AerB C domain could compete for AerA PCP domain-tethered substrates, it was clear that detection of small mass differences such as chlorination for such a large protein were not accessible at the resolutions provided by UHPLC-ESI-Q-TOF-MS analysis of the full length AerB subunit.

2.3.2. *In vitro* reconstitution of AerJ and AerB competition using a split AerB subunit

To circumvent the need to monitor the reaction on the 183 kDa AerB protein, an alternative approach, in which the subunit was split into its constituent domains was devised. Monitoring the reaction on a smaller AerB PCP domain-containing construct would allow greater resolution in UHPLC-ESI-Q-TOF-MS mass spectra and allow resolving of the different chlorination patterns of the product.

Domain boundaries between each of the constitutive domains required careful selection to both facilitate soluble expression and to preserve the ability for the domains to interact intermolecularly. The AerB C-A domain boundary was determined by alignment of the sequence with the C-A didomain of SrfA-C, the crystal structure of which Bode and co-workers have previously used to identify an ideal region of the flexible linker between the C and A domains for productive excisions.[191] This was a different boundary compared to the one chosen for the standalone AerB C domain construct in section 2.2.1, preserving more of the interdomain linker at the C-terminus of the C domain (figure **2.16 A**). The A-PCP linker from SrfA-C contains a conserved LPXL motif which interacts with a C-terminal subdomain of the A domain and is implicated in mediating conformational changes.[192] This motif was identified in the AerB subunit by sequence alignment with other A-PCP linker regions and was included at the C-terminus of the A domain-containing constructs (figure **2.16 B**). PCP-E domain boundaries are less extensively studied, with no exemplary excision sites described in the literature. For this, secondary structure predictions by the Psipred 4.0 webserver[193,194] were generated (figure **2.16 C**) to help ensure the excision site was not disrupting any predicted secondary structure elements, improving the chances of maintaining catalytic activity and domain-domain interactions of the excised proteins.

AerB split module constructs were generated by mutagenesis (figure **2.17**). Initially, a pET24a-(+)-AerB(C-A-PCP) construct was made by deletion mutagenesis using a Q5 mutagenesis kit using pET24a-(+)-AerB(C-A-PCP-E) as a template. Back-to-back primers were designed to delete the E domain in-

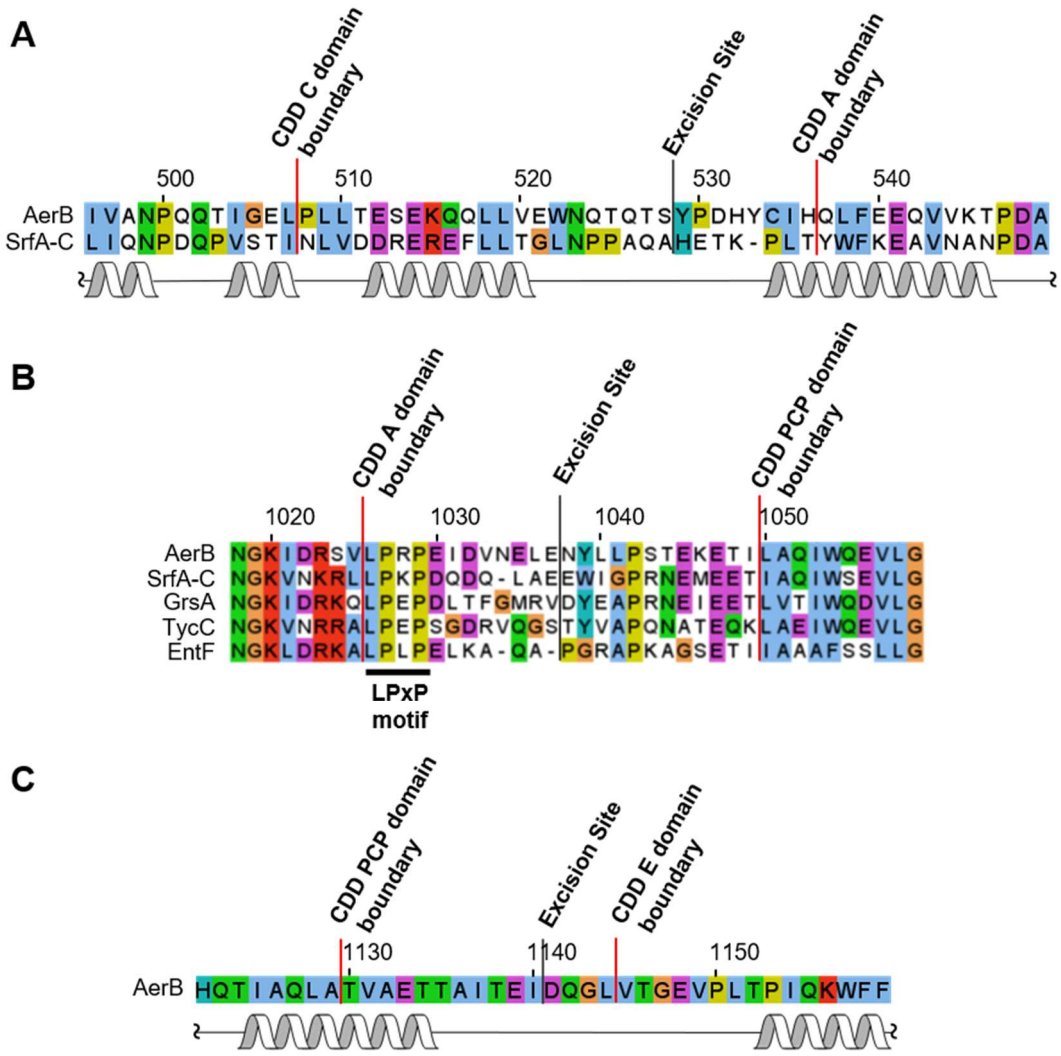


Figure 2.16 Sequence alignments used to determine domain boundaries for excision of domains from the AerB subunit. (**A**) Sequence alignment of AerB with SrfA-C at the C-A domain boundary. Below the sequence alignment, the known secondary structural elements from the crystal structure of SrfA-C (PDB accession: 2VSQ) are provided. (**B**) Alignment of A-PCP domain boundaries from several NRPS subunits. A conserved LPxP motif, which is important for A domain function, is highlighted. (**C**) The sequence of the PCP-E domain boundary. A secondary structure prediction by PsiPred 4.0 is provided below the sequence. In each schematic, the positions of excision sites and boundaries of the domains predicted by conserved domain searches is given above the sequence. Amino acid numbering is provided with reference to the AerB subunit. Domain boundaries were decided using a Conserved Domain Database (CDD) search.

frame whilst preserving the stop codon (figure **2.17 A**). PCR and ligation was performed as described in section 2.1.4. Once this construct had been generated and fidelity of the plasmid confirmed by sequencing, it was used as a template to generate pET24a-(+)-AerB(C)' (an AerB C domain construct with a different C-terminal boundary to that in section 2.2.1), pET24a-(+)-AerB(C-A) and pET24a-(+)-AerB(A-PCP) constructs using the same deletion mutagenesis strategy. Subsequently, constructs for the AerB(A) and

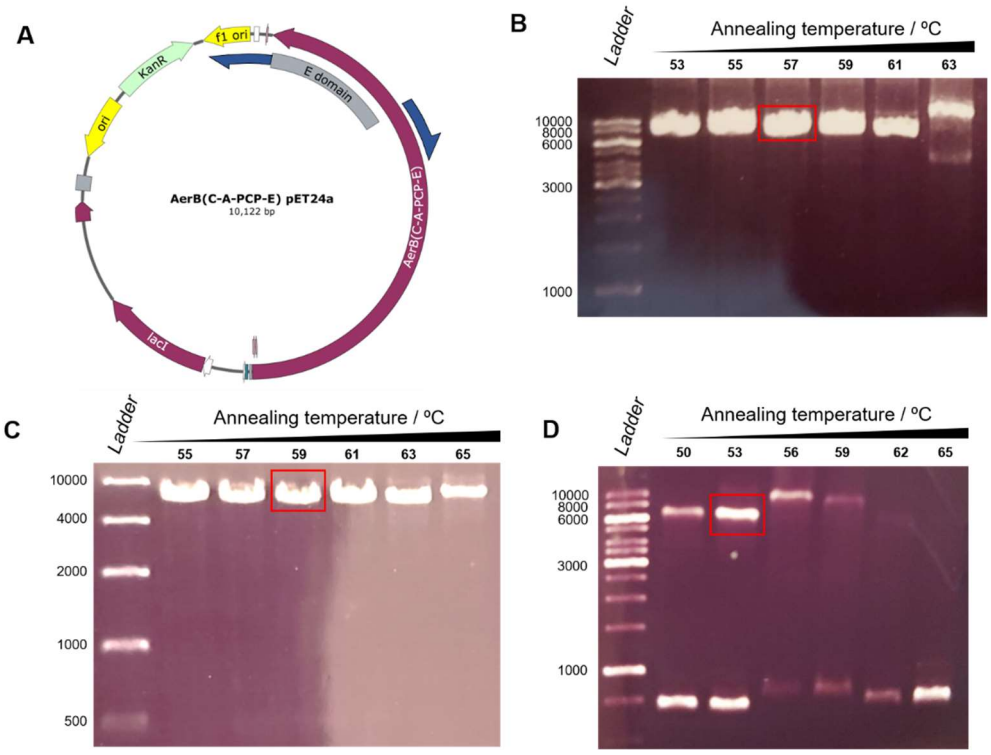


Figure 2.17 Mutagenesis to generate AerB split module expression constructs in pET24a-(+). (**A**) Schematic displaying general deletion mutagenesis strategy. Blue arrows indicate position of primers. (**B**) Agarose gel electrophoresis displaying PCR products resulting from mutagenesis to generate pET24a-(+)-AerB(C-A-PCP). Expected size = 8.7 kb. (**C**) Agarose gel electrophoresis displaying PCR products resulting from mutagenesis to generate pET24a-(+)-AerB(C-A). Expected size = 7.2 kb. (**D**) Agarose gel electrophoresis displaying PCR products resulting from mutagenesis to generate pET24a-(+)-AerB(PCP). Expected size = 5.6 kb. Bands highlighted in red were used for KLD ligation to generate the desired plasmid.

AerB(PCP) domains were generated using the pET24a-(+)-AerB(A-PCP) construct as a template.

Each of the constructs was then overproduced in *E. coli* BL21(DE3) cell lines and purified by Ni-affinity chromatography to identify the most suitable combination of proteins for reconstituting the activity of the module (figure **2.18**). AerB(C-A-PCP) was purified as a soluble protein, but copurified as a mixture with the same pattern of truncated products as with the full-length protein (figure **2.18 A**). Because the mass difference between the protein of interest and these by-products was now smaller, the protein could not be separated by preparative size-exclusion chromatography. Dilute bands corresponding to AerB(A-PCP) appeared to elute in the 100 mM and 200 mM imidazole fractions during purification by SDS PAGE analysis (figure **2.18 B**),

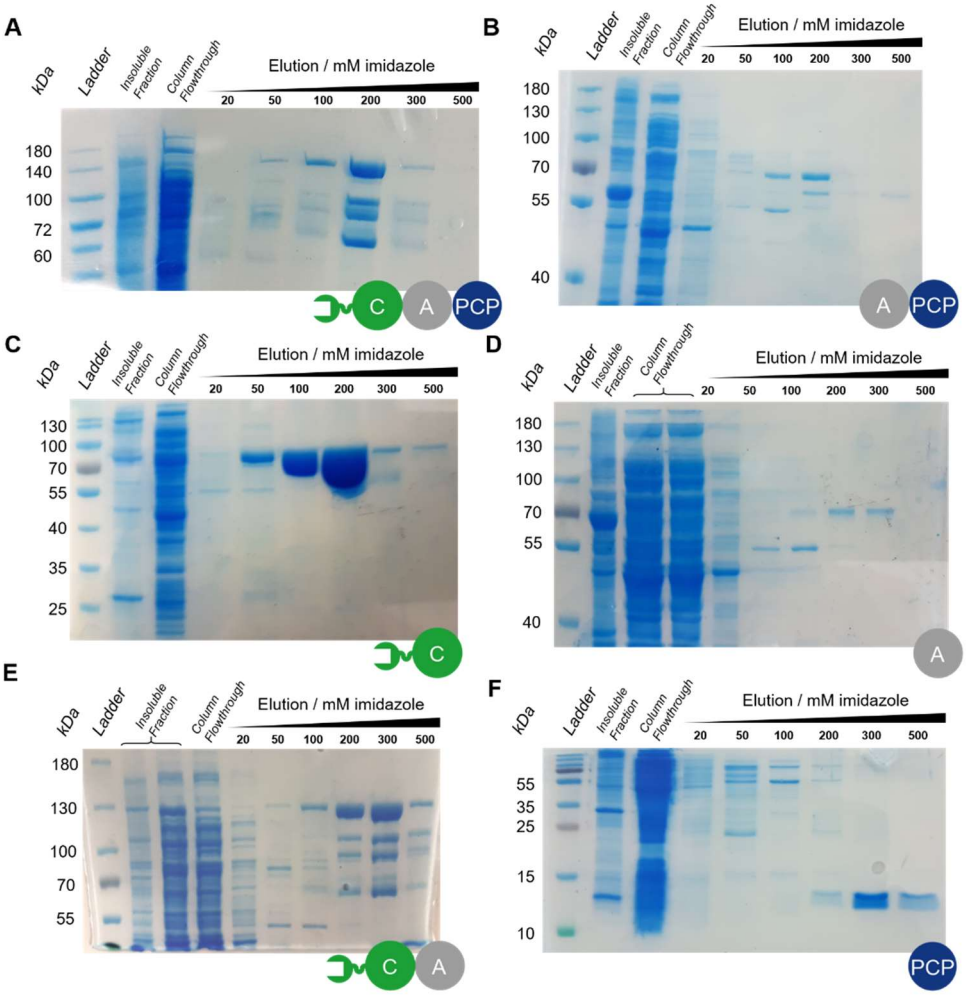


Figure 2.18 SDS PAGE analysis of elution profiles from Ni-affinity chromatography of: (**A**) AerB(C-A-PCP), expected mass = 130.2 kDa; (**B**) AerB(A-PCP), expected mass = 70.0 kDa; (**C**) AerB(C)', expected mass = 60.2 kDa; (**D**) AerB(A), expected mass = 59.9 kDa, (**E**) AerB(C-A), expected mass = 120.4 kDa; (**F**) AerB(PCP), expected mass = 12.8 kDa.

but UHPLC-ESI-Q-TOF-MS analysis of the resulting protein did not detect any protein with the correct mass in these fractions. The AerB(C)' domain expressed and purified well (figure **2.18 C**), suggesting the change of C-terminal boundary as compared to that in section 2.2.1. did not substantially affect protein overproduction or solubility. AerB(A) appeared to be produced as an insoluble protein, with a large band observed in the insoluble fraction on the SDS PAGE gel between 55-70 kDa (figure **2.18 D**). Insolubility of recombinant A domains is a common problem, owing to the inherent conformational flexibility required for their mechanism.[195] MbtH-like proteins have been implicated in A domain activity, expression and solubility.[196-199] Bioinformatics analysis of the aeruginosin gene cluster, performed using the

BLAST-P webserver, found no MbtH-like protein. However, it remains possible that an MbtH-like protein encoded elsewhere in the genome of *M. aeruginosa* NIES-98 may aid soluble expression of the A domain-containing constructs, and identification and co-expression of such a protein could also be utilised in future studies.

Both the AerB C-A didomain and AerB PCP domain were overproduced as soluble proteins (figure **2.18 E-F**) and the identity of each was confirmed by mass spectrometry (figure **2.19**). Interestingly, the AerB PCP domain purified approximately 40% in the *holo*-form. This likely results from phosphopantetheinylation of the conserved serine *in vivo* by an *E. coli* 4'-phosphopantetheinyltransferase when overproduced. As all other A domain-containing constructs were poorly overproduced or insoluble, these two constructs were chosen to generate the split module. The advantage of this would be that reactions could be monitored on the small AerB(PCP) construct which would allow simple resolution of chlorinated species. The AerB C-A didomain co-purified with small amounts of truncated protein, but because

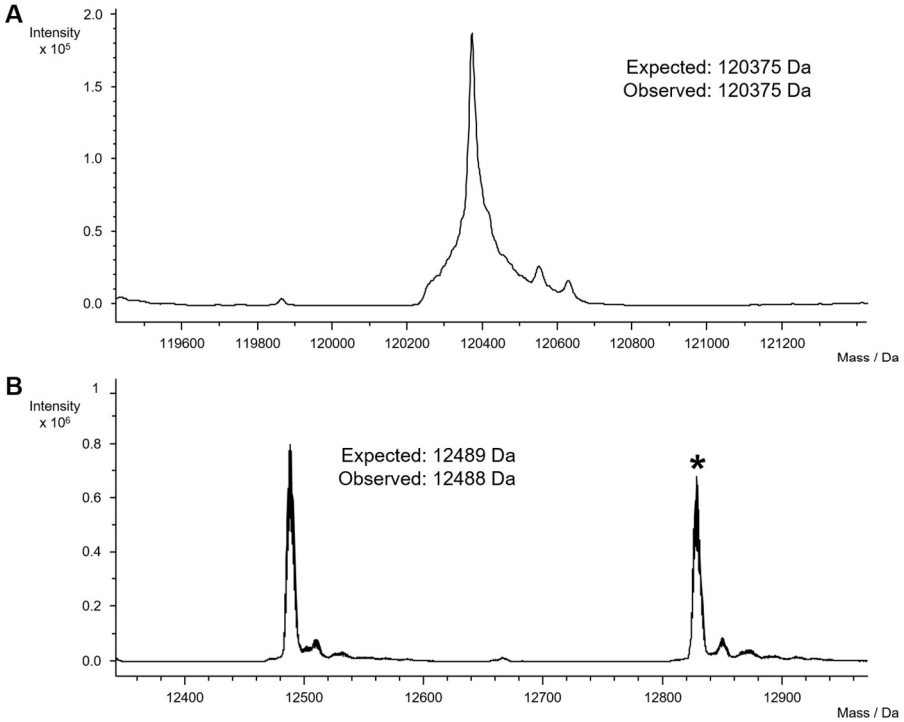


Figure 2.19 UHPLC-ESI-Q-TOF-MS analysis confirming the identity of Ni-NTA purified: (**A**) AerB(C-A); (**B**) AerB(PCP). In (**B**), the asterisk (*) indicates a peak corresponding to *holo*-AerB(PCP) which co-purified with the *apo*-protein from BL21(DE3) cells.

monitoring of the reaction by intact protein MS could now take place on the isolated PCP domain, this required no further purification.

To determine if this combination of proteins was viable for *in vitro* reconstitution of the condensation activity, the assay described in section 2.2.2. was repeated using the split module constructs. AerB(PCP) was converted from *apo*- to *holo*-form by incubation with CoA, $MgCl_2$ and Sfp (figure **2.20 A-B**). L-isoleucine was then loaded *via* intermolecular reaction with a stoichiometric amount of AerB(C-A) (figure **2.20 C**). Loading of the amino acid was not as efficient as in the full-length protein, likely due to the intermolecular reaction being slower than intramolecular reaction, so hydrolysis of the pantetheine thioester is favoured kinetically. Finally, condensation was reconstituted by addition of 4-HPLA-*S*-AerA(PCP) and co-incubation for 1 h (figure **2.20 D**). A +164 Da shift was again observed, showing that condensation could be reconstituted using the split module AerB. Again, the reactions were conducted without 4-HPLA-*S*-AerA(PCP) (figure **2.20 E**) and no turnover was exhibited, proving that the condensation domain was not accepting residual 4-HPLA-*S*-pantetheine (**51**) or 4-HPLA-*S*-CoA (**61**) in solution.

To prove that condensation activity was being catalysed by the C domain, and not a result of non-catalysed displacement of the upstream PCP thioester, a C domain active site mutant was generated. His_{222} was determined to be the active site residue by identification of the conserved C domain HHxxxDG motif by sequence alignment to characterised C domains (figure **2.21 A**).[200,201] A H222A mutant (AerB(C-A)H222A) was constructed by site-directed mutagenesis using methods discussed for AerJ(K156A) in section 2.1.4., and purified as a soluble protein (figure **2.21 B-C**). Repeating the assay demonstrated that AerB(C-A)H222A could still load L-isoleucine onto *holo*-AerB(PCP), but could no longer catalyse condensation (figure **2.20 F**). This negative control also eliminates the possibility that the impurities that co-purified with AerB(C-A) were catalysing condensation.

Having established that splitting AerB into constituent domains does not prevent condensation, the halogenase AerJ and relevant cofactors were

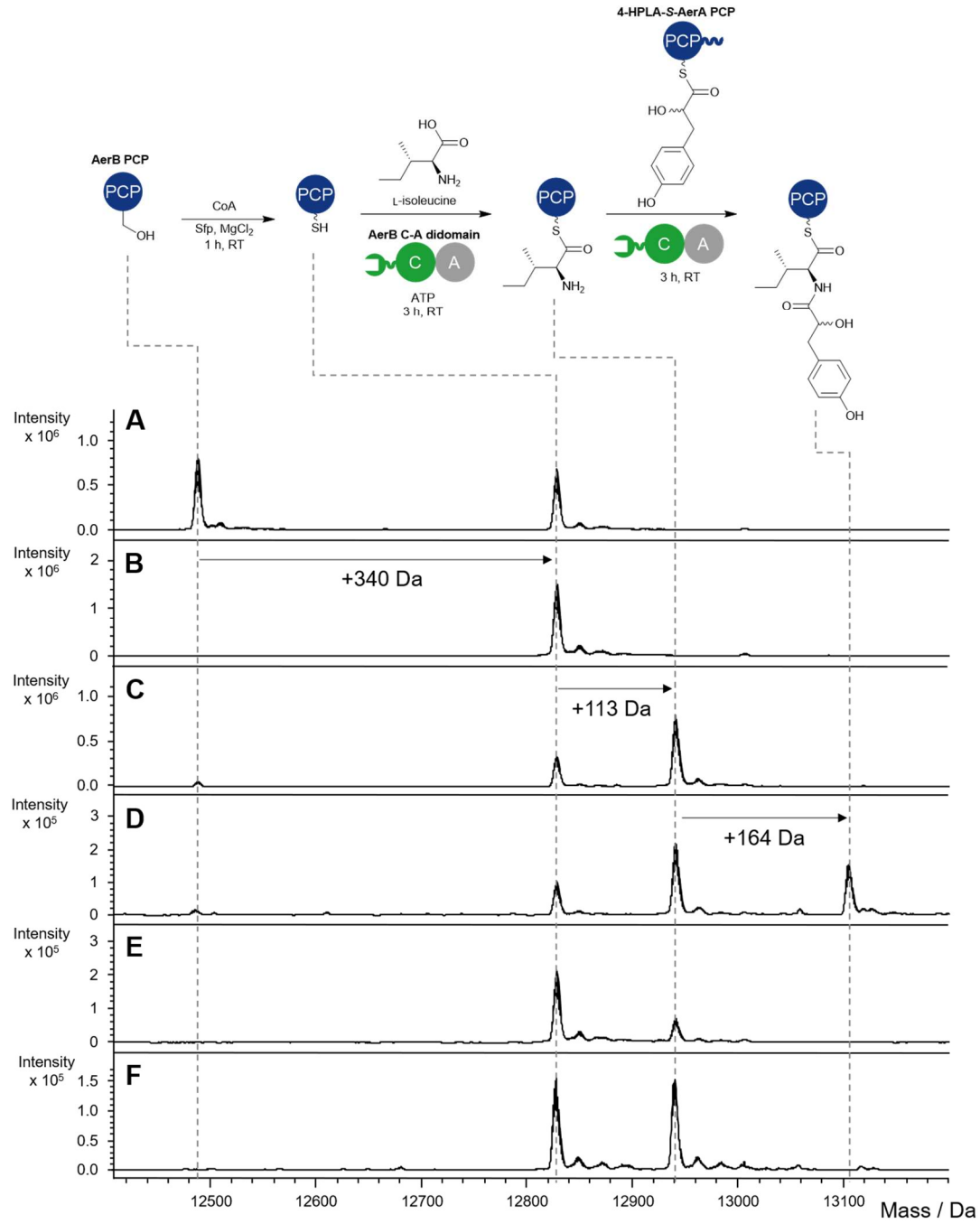


Figure 2.20 UHPLC-ESI-Q-TOF-MS analysis of the reconstitution of AerB C domain activity using AerB(C-A) didomain and AerB(PCP) constructs, with monitoring of the AerB(PCP) species. (**A**) *apo*-AerB(PCP). (**B**) *holo*-AerB(C-A-PCP-E) following incubation with CoA, Sfp and $MgCl_2$. (**C**) L-Ile-*S*-AerB(PCP) following incubation of reaction (**B**) with AerB(C-A), L-isoleucine and ATP. (**D**) Condensation product following incubation of (**C**) with 4-HPLA-*S*-AerA(PCP). (**E**) Product of reaction of (**C**) where the loading reaction used to generate 4-HPLA-*S*-AerA(PCP) was conducted in the absence of AerA(PCP). (**F**) Product of reaction (**B**)-(**D**) repeated with AerB(C-A) replaced with AerB(C-A)H222A

added to the reaction (figure **2.22**). Incubation of L-Ile-*S*-AerB(PCP) and AerB(C-A) with 4-HPLA-*S*-AerA(PCP) and AerJ and relevant cofactors in a 1:1:1:1 stoichiometry yielded a mixture of non-, mono- and di-chlorinated condensed products in an approximate 2:3:4 ratio (figure **2.22 C**). This

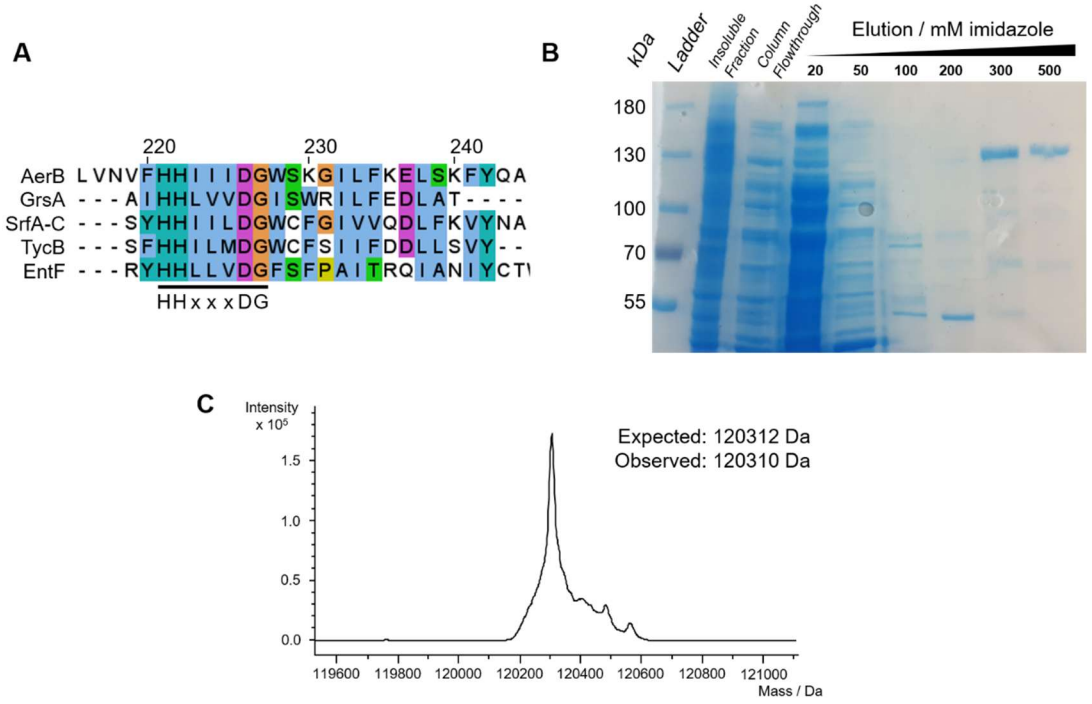


Figure 2.21 Generation of a AerB(C-A)H222A point mutant. (**A**) Identification of the conserved HHxxxDG motif in the active site of the AerB C domain by sequence alignment with characterised C domains. (**B**) SDS PAGE analysis of the elution profile from Ni-affinity chromatography of AerB(C-A)H222A. Expected mass = 120.3 kDa. (**C**) Charge-state deconvoluted mass spectrum from UHPLC-ESI-Q-TOF analysis of AerB(C-A)H222A confirming identity of the purified protein.

confirmed the predicted simultaneous action of the halogenase and condensation domain at this subunit junction in aeruginosin biosynthesis and provides a biochemical basis for the mixture of aeruginosin congeners isolated from *M. aeruginosa* NIES-98.

Reducing the amount of AerJ present in solution compared to AerB(C-A) didomain decreased the conversion to mono-chlorinated and di-chlorinated products (figure **2.22 D-F**), as would be expected. Presumably increasing the amount of AerJ in solution would increase the conversion to chlorinated products. However, it was not possible to test this because as AerB(C-A) was required for adenylation and loading of isoleucine, its concentration could not be varied and increasing the concentration of AerJ was not possible due to the potential of saturating the MS detector. Nevertheless, for potential future biosynthetic engineering applications, careful control of relative protein concentration could be critical for accessing different chlorinated analogues.

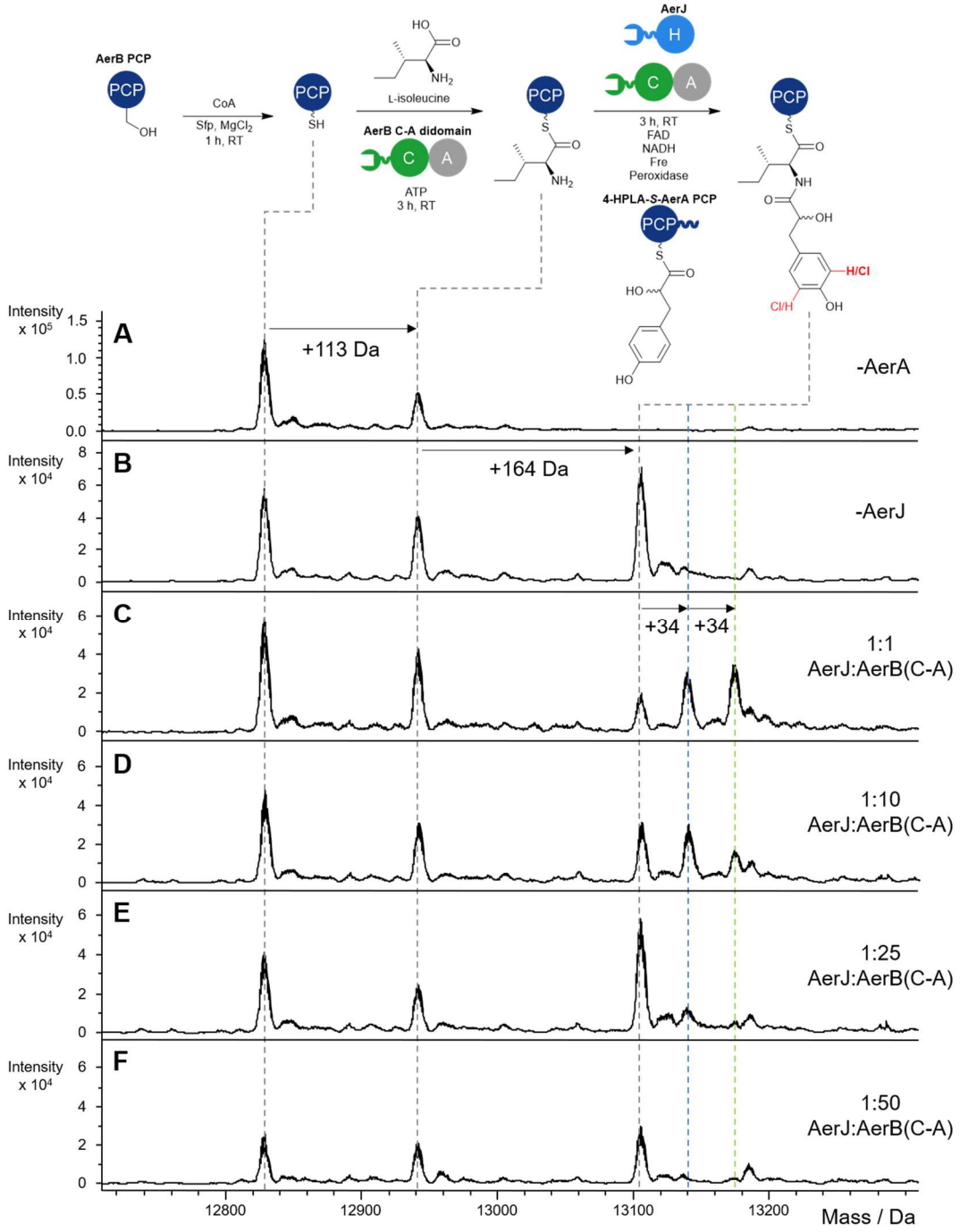


Figure 2.22 UHPLC-ESI-Q-TOF-MS analysis of reconstitution of halogenase and condensation domain activity using AerJ, AerB(C-A) didomain and AerB(PCP) constructs, with monitoring of the AerB(PCP) species. (**A**) Control reaction where condensation had been conducted in the absence of 4-HPLA-*S*-AerA(PCP). (**B**) Control reaction where condensation had been conducted in the absence of AerJ. (**C**) Condensation reaction with 1:1 ratio of AerJ:AerB(C-A), peroxidase and the relevant cofactors for $FADH_2$ generation. (**D**) As in (**C**) but with 1:10 ratio of AerJ:AerB(C-A). (**E**) As in (**C**) but with 1:25 ratio of AerJ:AerB(C-A). (**F**) As in (**C**) but with 1:50 ratio of AerJ:AerB(C-A).

2.3.3. Confirming the order of condensation and halogenation

Whilst the results in section 2.3.2. provided initial indications that both AerJ and the AerB C domain could act upon the 4-HPLA-*S*-AerA(PCP) intermediate, another potential origin for the halogenated products observed provided in figure **2.22 C** could be that AerJ can halogenate the 4-HPLA-L-Ile-*S*-AerB(PCP) condensation product (figure **2.23**). From the data provided, it was unclear as to whether this could occur and was significantly contributing to the chlorinated product formation.

To determine whether AerJ can directly chlorinate 4-HPLA-L-Ile-*S*-AerB(PCP), it needed to be generated as a substrate. Since full turnover to 4-HPLA-L-Ile-*S*-AerB(PCP) using the split module system was never observed, it was instead chemoenzymatically synthesised using the 4'-phosphopantetheinylation methodology described in section 2.1.3., appending

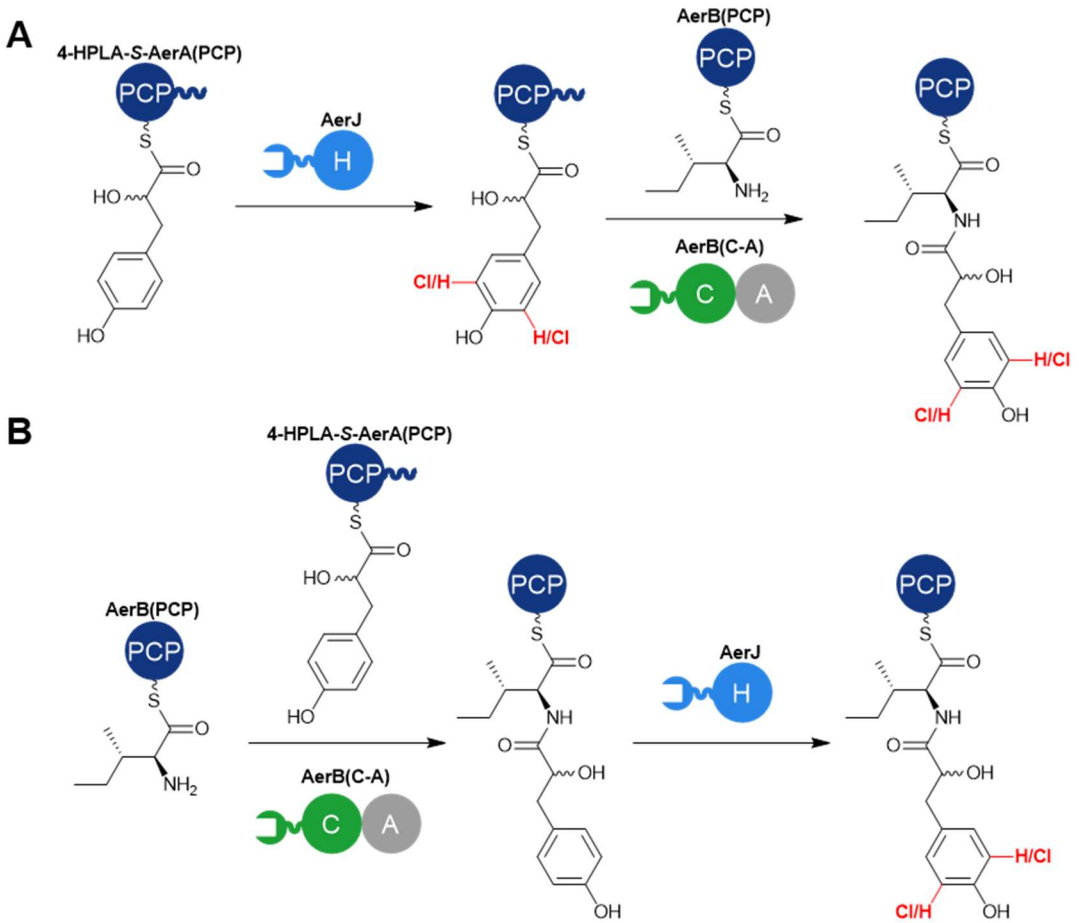
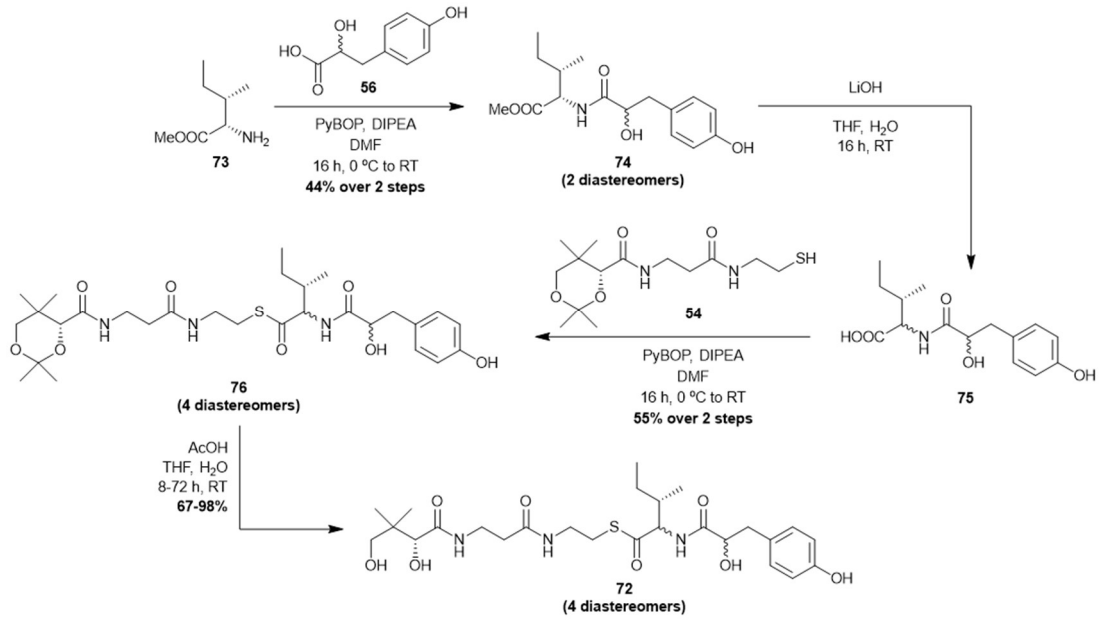


Figure 2.23 Possible scenarios leading to chlorinated 4-HPLA-L-Ile-*S*-AerB(PCP) products observed. (**A**) Halogenation of 4-HPLA-*S*-AerA(PCP) by AerJ followed by condensation with L-Ile-S-AerB(PCP) by AerB(C-A). (**B**) Condensation of 4-HPLA-*S*-AerA(PCP) with L-Ile-*S*-AerB(PCP) by AerB(C-A) followed by halogenation of 4-HPLA-L-Ile-*S*-AerB(PCP) by AerJ.



Scheme 2.6 Synthesis of 4-HPLA-L-Ile-*S*-pantetheine (**72**).

chemically synthesised 4-HPLA-Ile-*S*-pantetheine (**72**) onto the conserved serine residue of the AerB PCP domain.

4-HPLA-Ile-*S*-pantetheine was chemically synthesised *via* a 4-step procedure (scheme **2.6**). L-isoleucine methyl ester (**73**) was coupled to racemic 4-HPLA (**56**) via a PyBOP coupling similar to that discussed in section 2.1.1. This generated two diastereomers which were separable by silica gel chromatography. Deprotection of the methyl ester of each diastereomer afforded acids (**75**), the crude products of which were coupled to acetonide-protected pantetheine (**54**) to give the corresponding pantetheine thioester derivatives (**76**). Each diastereomer of **74** yielded two diastereomers of **76**, likely due to base-catalysed racemisation of the α-carbon of the isoleucine group. This likely occurred in the ester deprotection step, where the strong base lithium hydroxide could facilitate racemisation by deprotonation of the acidic α-proton and re-protonation of the opposing face. The β position of the isoleucine remains fixed in the (*S*)-configuration as the β-proton is not acidic enough to be efficiently deprotonated. Each pair of diastereomers were separated by semi-preparative reverse-phase HPLC. Each diastereomeric acetonide-protected pantetheine thioester was then deprotected to give the desired product **72** as four distinct diastereomers.

From the ^{1}H and ^{13}C NMR spectra of the four diastereomers generated, the stereochemistry of the isoleucine moiety could be predicted. In a previous study by Anderson *et al.*, several C- and N-substituted L-isoleucine and D-*allo*-isoleucine analogues were generated.[202] It was found that general trends in the NMR spectra allowed the L- and D-*allo*- epimers to be distinguished

Table 2.1 ^{1}H and ^{13}C NMR spectroscopic data relating to the isoleucyl α-proton and α-carbon of the four 4-HPLA-Ile-*S*-pantetheine (**72**) diastereomers. Values in bold represent the larger values for each measurement. Spectroscopic data was obtained on a Bruker Avance III HD 500 spectrometer in d_4-methanol.

Diastereomer	^{1}H		^{13}C	Isoleucine moiety stereochemistry
	Chemical Shift (ppm)	$^{3}J_{CH-CH}$ (Hz)	Chemical Shift (ppm)	
1	4.43	**6.0**	**63.3**	L
2	**4.57**	4.0	61.4	D-*allo*
3	**4.62**	4.5	61.6	D-*allo*
4	4.40	**6.0**	**63.1**	L

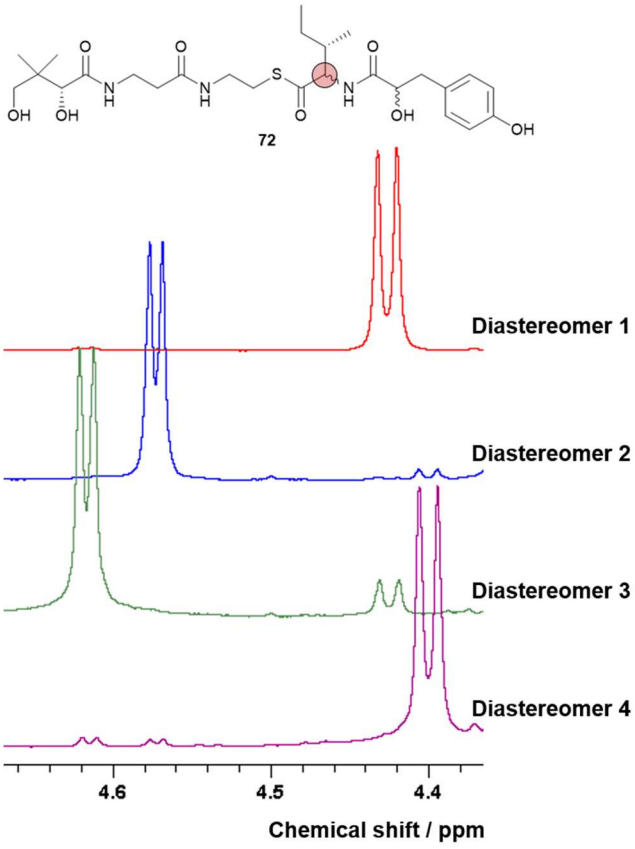


Figure 2.24 ^{1}H NMR spectra displaying doublets corresponding to the L-Ile α-protons of diastereomers of **72**.

regardless of the moiety attached to the amino group. Specifically, in all cases: the ^{1}H chemical shift value of the α-proton is lower for the L- epimer than the D-*allo*- epimer; the $^{3}J_{CH\text{-}CH}$ coupling constant between the α- and β-protons is larger for the L- epimer than the D-*allo*- epimer; and the ^{13}C chemical shift value of the α-carbon is higher for the L- epimer than the D-*allo*- epimer.[202] Each of these values for the diastereomers of **72** is given in table **2.1**, with all NMRs obtained in d_4-methanol and referenced to the solvent peak (figure **2.24**). Relating these values to the trends reported by Anderson *et al.*, the stereochemical configuration about the α-carbon of the isoleucine moiety of each diastereomer of **72** was assigned (table **2.1**).[202]

Unlike the α-carbon of the isoleucine moiety, there is no literature precedent for the stereochemical assignment of the 4-HPLA moiety by NMR. However, the ^{1}H spectral data for the 4-HPLA α-proton for diastereomers 1 and 3 are similar, with chemical shifts of 4.28 and 4.29 ppm respectively. and $^{3}J_{CH\text{-}CH}$ coupling constants of 8.0 and 4.5 Hz for both (as measured to the nearest 0.5 Hz). Similarly, diastereomers 2 and 4 have similar spectral data for the α-proton, with chemical shifts of 4.35 and 4.32 ppm respectively and identical $^{3}J_{CH\text{-}CH}$ coupling constants of 6.5 and 4.0 Hz. This indicates the presence of both stereochemical configurations at the 4-HPLA α-carbon, as would be expected from using racemic 4-HPLA starting material. As diastereomers 1 and 3 of **72** both derive from one diastereomer of **74**, diastereomers 2 and 4 of **72** both derive from the other diastereomer of **74**, it is almost certain that each pair has the same configuration at the 4-HPLA α-carbon. However, it is unclear without further experimentation how the ^{1}H NMR spectroscopic data relates to the absolute stereochemical configuration in this case.

Evidence from the NMR spectroscopic data presented here strongly indicates that all four diastereomers at the α-positions of the isoleucine and 4-HPLA residues had been produced and purified during this synthesis. Nevertheless, in future work the absolute configuration of the synthetic pantetheine thioesters could potentially be determined by chemical derivatisation and NMR spectroscopy, for example by Mosher's ester analysis.[203] This involves the α-hydroxyl group being derivatised as an ester with (*R*)- and (*S*)-α-methoxy-α-

trifluoromethylphenylacetic acid, with comparative analysis of the ^{1}H spectra of the esters allowing distinction between the two configurations. Alternatively, the absolute configuration could be determined by X-ray crystallography.[204]

Following synthesis, each diastereomer of **72** was loaded onto the AerB PCP domain using the enzymatic reactions described in section 2.1.1 prior to incubation with AerJ, Fre, FAD, NADH and horseradish peroxidase. This assay was repeated in triplicate for each diastereomer. No halogenation was observed for any of the diastereomers by intact protein MS (figure **2.25**). This validates the hypothesis that AerJ must act upon the AerA carrier protein-tethered 4-HPLA in aeruginosin biosynthesis prior to condensation. It is unclear whether selectivity for the AerA PCP domain-tethered substrate is primarily controlled by preference of the halogenase active site for the shorter-chain substrate or whether selective protein-protein interactions with the AerA PCP domain are essential in controlling the order of actions. Further investigations into the origins of this selectivity will be discussed in Chapter 3. However, if selective protein-protein interactions are the major factor contributing to selectivity at this interface, this could give credence to the necessity of the SLiM-βHDD interface in mediating halogenation by AerJ.

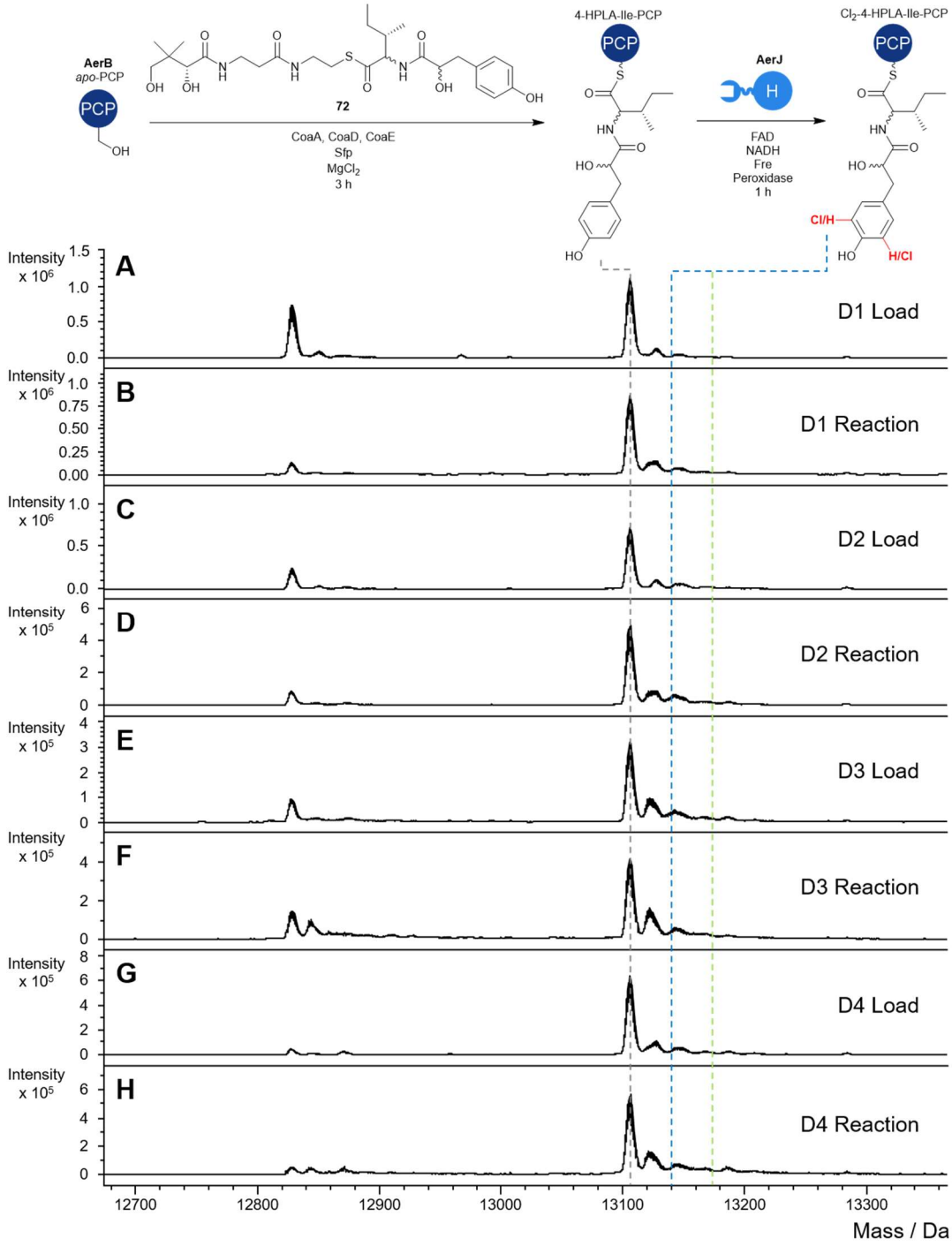


Figure 2.25 Charge-state deconvoluted mass spectra from UHPLC-ESI-Q-TOF-MS analysis demonstrating the products of: (**A**) Loading of 4-HPLA-L-Ile-*S*-pantetheine diastereomer 1 onto AerB(PCP); (**B**) Incubation of 4-HPLA-L-Ile-*S*-AerB(PCP) diastereomer 1 with AerJ and relevant cofactors for 1 h; (**C**) as (**A**) but with diastereomer 2; (**D**) as (**B**) with diastereomer 2; (**E**) as (**A**) but with diastereomer 3; (**F**) as (**B**) with diastereomer 3; (**E**) as (**A**) but with diastereomer 4; (**F**) as (**B**) with diastereomer 4. Blue dashed line = expected mass of the mono-chlorinated product, green dashed line = expected mass of the di-chlorinated product.

2.4. Conclusions

2.4.1. Summary

In this chapter, the initial steps of the biosynthesis of aeruginosin in *Microcystis aeruginosa* NIES-98 have been reconstituted *in vitro*. Synthesis of 4-HPLA-*S*-pantetheine (**51**) has enabled chemoenzymatic synthesis of 4-HPLA-*S*-AerA(PCP). This has been used as a substrate to understand the reactions catalysed by the two βHDD-bearing domains, AerJ and the AerB C domain *via* intact protein mass spectrometry.

As hypothesised, the flavin-dependent halogenase AerJ is responsible for mono- and di-chlorination of 4-HPLA tethered to the AerA PCP domain prior to condensation catalysed by the downstream C domain of AerB. The data presented evidences competition between AerJ and the AerB C domain for 4-HPLA-*S*-AerA(PCP) being responsible for the mixture of chlorination patterns observed in aeruginosin isolates of *M. aeruginosa* NIES-98. Chlorination cannot occur following condensation, as AerJ does not accept 4-HPLA-L-Ile-*S*-AerB(PCP) as a substrate.

2.4.2. Chlorination patterns in related biosynthetic systems

Here, the origins of the chlorination patterns in aeruginosins from *M. aeruginosa* NIES-98 is described. However, other *Microcystis* strains containing homologous *aerJ* genes are known to produce aeruginosins with different degrees of chlorination (figure **2.26**). For example, isolates from *Microcystis aeruginosa* PCC 7806 and PCC 9812 are reported to exclusively consist of a mono-chlorinated aeruginosin congener, aeruginosins 686 (**77**) and 608 (**78**) respectively.[158] Meanwhile, strains PCC 7813 and 7820 produce a mixture of non- and mono-chlorinated aeruginosins (**79** and **80**), whilst strains PCC 9701, PCC 9804, PCC 9805 and NIES-101 biosynthesise only di-chlorinated aeruginosin analogues (**81** and **44**).[158] This is despite the sequences of AerJ and AerA homologues possessing >91% identity in these strains. This includes identical sequences of the AerA SLiM and β2 strand of

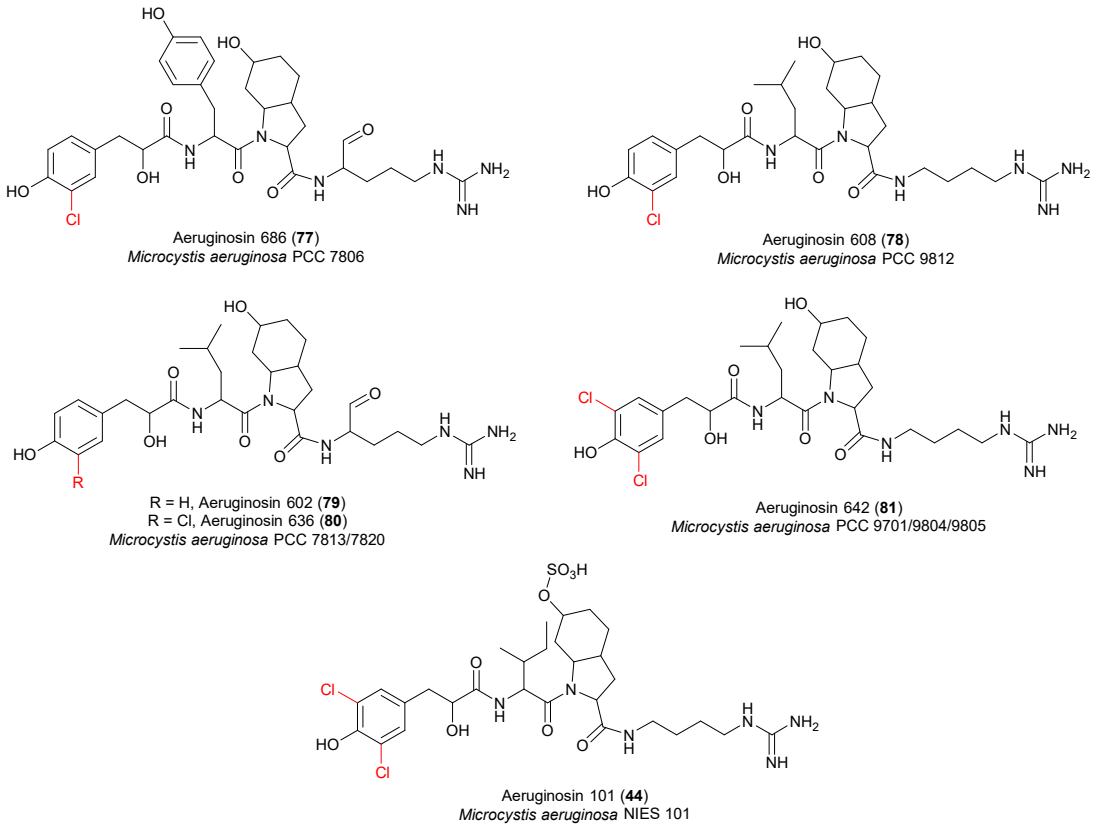


Figure 2.26 Chlorinated aeruginosin congeners from *Microcystis aeruginosa* strains. Stereochemical assignments for these derivatives are not currently available.

the AerJ and AerB βHDDs (figure **2.27**), which are known to form critical electrostatic interactions in association of SLiM-βHDD pairs.[101,126]

It is interesting that such highly related interfaces, both within *Microcystis* strains and in other cyanobacteria, can produce differing chlorination patterns. It is plausible that these differences are merely artefacts of the cyanobacterial strain culturing and natural product isolation process. For instance, insufficient amounts of chloride ions in laboratorial culture conditions may prevent halogenase activity. It may also be the case that relative levels of expression of each of the genes plays a role in the generation of a mixture of chlorination patterns, although *aerA*, *aerJ* and *aerB* genes would be expected to be under the control of the same promoter. Nevertheless, it would be of interest in future experiments to demonstrate whether these differences can be attributed to structural or residue-level divergence of the biosynthetic proteins, and, if so, underpin the key structural elements responsible for control of degree of

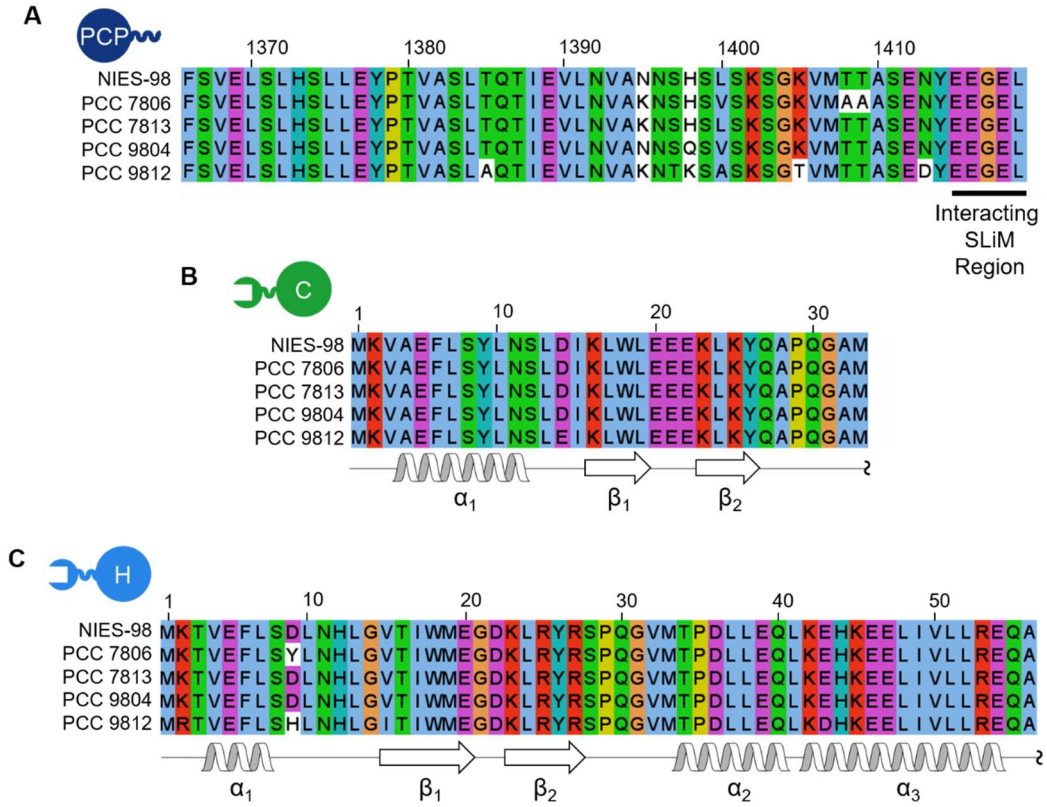


Figure 2.27 Sequence alignments of proteins encoded by *aerA*, *aerJ* and *aerB* genes in various *Microcystis* species. (**A**) Alignment of the C-terminus of AerA. (**B**) Alignment of the N-terminus of AerB. Note that sequence coverage of the entire βHDD is not available for PCC strains. (**C**) Sequence alignment of the N-terminus of AerJ. In all alignments, sequence numbering is provided relative to the protein encoded by the protein from *Microcystis aeruginosa* NIES-98 strain. For (**B**) and (**C**), expected secondary structural elements from PsiPred 4.0 webserver predictions of the *Microcystis aeruginosa* NIES-98 protein are provided below the alignment. GenBank Accessions: NIES-98 = AM773660.1; PCC 7806 = AM773662.1; PCC 7813 = AM773661.1; PCC 9804 =AM773659.1; PCC 9812 = AM773664.1.

halogenation. This would give insights into the relationship between structure, protein-protein interactions and chemistry in natural product biosynthesis.

Moreover, a related SLiM-βHDD interface is found in the biosynthesis of two related compounds of the cyclic depsipeptide class; cyanopeptolins (**82** and **83**) in *Planktothrix*, *Microcystis*, *Sphaerocavum* and *Nostoc* species and anabaenopeptolides (**84** and **85**) in *Anabaena* species (figure **2.28**).[158,205-208] In both cases, a mixture of non- and mono-chlorinated natural products have been isolated from strains containing a putative flavin-dependent halogenase gene, with chlorination occurring at the L-tyrosine moiety incorporated by module 6 of their respective NRPSs. This module appears at the C-terminus of the McnC and ApdB subunits, and both the downstream C domain, at the

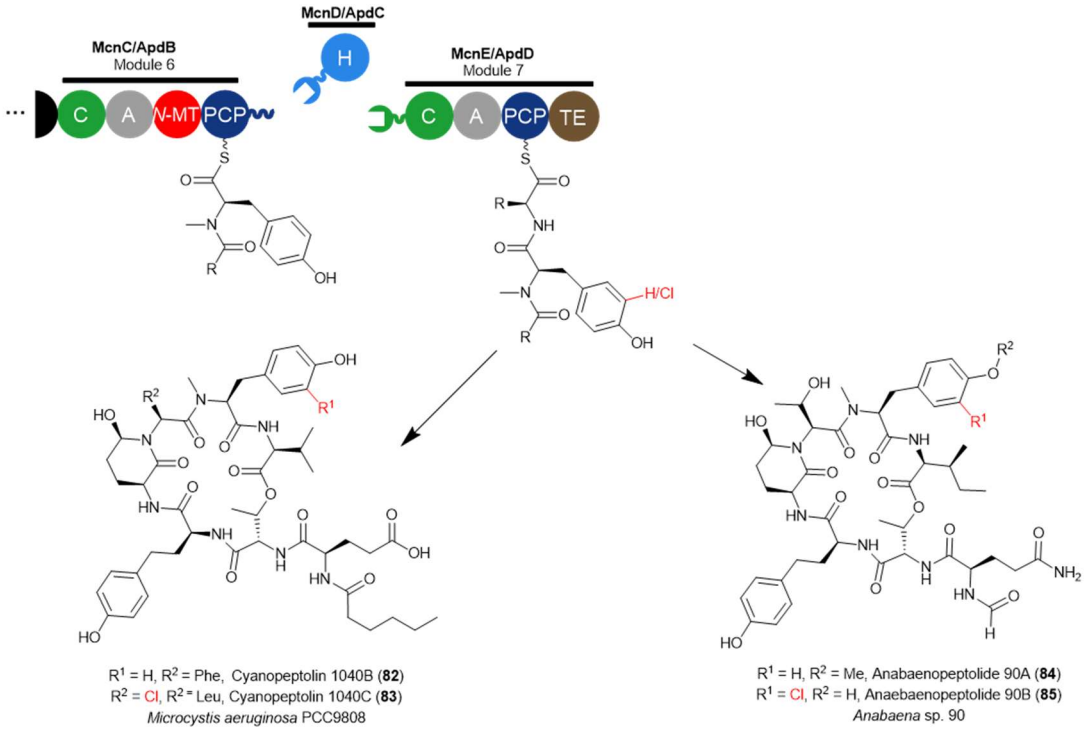


Figure 2.28 The proposed biosynthesis of cyanopeptolin (**82** and **83**) and anabaenopeptolide (**84** and **85**) congeners in *Microcystis aeruginosa* PCC 9806 and *Anabaena* sp. 90, respectively. A SLiM-βHDD docking interface is proposed to be involved in recruiting the halogenase to the PCP of module 6.

N-terminus of McnE and ApdD, and standalone halogenases McnD and ApdC, are predicted from sequence alignments and secondary structure predictions to harbour N-terminal βHDDs (figure **2.29 A**). Similarly, the C-terminal PCP domains of module 6 appear have glutamate-rich sequences, aligning well with other known SLiMs (figure **2.29 B**). It is therefore likely that the halogenases act upon the module 6 PCP domain-tethered substrates, directed by a SLiM-βHDD interaction similar to the aeruginosin system described herein.

It is unclear in each case whether the halogenase acts on the tyrosinyl-PCP domain, hexapeptide-PCP domain or indeed both. In either case, the origins of non- or mono-chlorination are very interesting. It is likely that the mixtures of these halogenation patterns can be explained by βHDD-mediated C domain and halogenase competition observed in this study. The apparent inability to di-chlorinate the tyrosine moiety could also be attributed to this competition or instead be a consequence of substrate selectivity of the halogenase or C

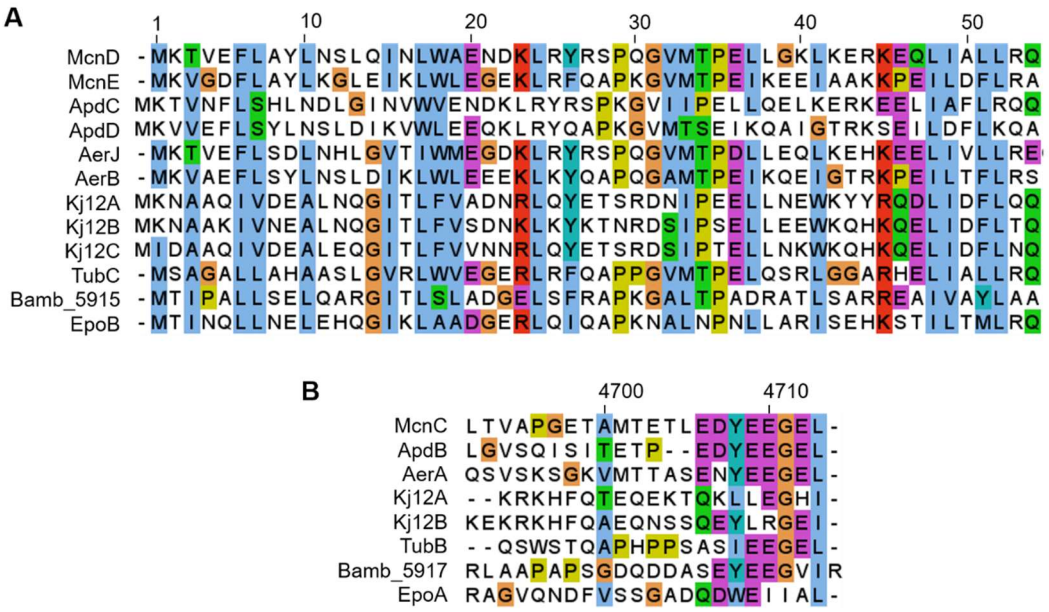


Figure 2.29 Sequence alignments of (**A**) βHDDs and (**B**) SLiMs from the biosynthesis of cyanopeptolin and anabaenopeptolide with characterised examples.

domain active sites. Similar experiments to those used in this study would be of great importance to further understanding docking domain-mediated recruitment of halogenases and underpin the ubiquity of this mechanism in modulating halogenation during non-ribosomal peptide assembly.

Chapter 3: Investigating protein-protein interactions in aeruginosin biosynthesis

3.1. Biochemical studies of protein-protein interactions in aeruginosin biosynthesis

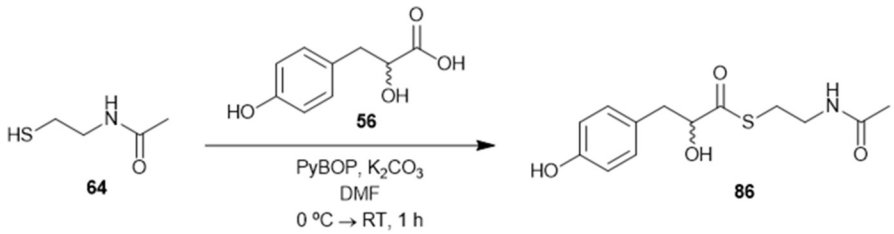
Having reconstituted the tandem halogenation and condensation of 4-HPLA-*S*-AerA(PCP) by βHDD-bearing proteins AerJ and AerB respectively, the next aim was to understand the role of the SLiM-βHDD interface compared to other interacting regions of the proteins in mediating recruitment of the two proteins to the carrier protein domain. This would improve understanding of the biosynthetic origins of chlorinated congeners of aeruginosin, and provide information to progress development of these proteins and other SLiM-βHDD interfaces into useful biosynthetic engineering tools.

Furthering understanding of the AerA PCP domain-AerJ interface was of particular interest. Whilst examples of flavin-dependent halogenases that act upon carrier protein-tethered substrates are known,[209-212] no biochemically-verified examples appear to have appended docking domains, and little is understood regarding the protein-protein interactions governing halogenation. Furthermore, engineering of protein-protein interaction interfaces to enable recruitment of tailoring enzymes such as halogenases would be highly valuable for facile introduction of chemical diversity into engineered natural products.

3.1.1. Attempted halogenation of small molecules demonstrates the necessity of protein-protein interactions for halogenation

To understand whether protein-protein interaction between AerJ and the AerA PCP domain is required for halogenation, the ability of AerJ to halogenate small molecules was investigated.

Three substrates were trialled initially; racemic 4-HPLA and its *N*-acetyl cysteamine and pantetheine derivatives. Unfortunately, the commercially available 4-HPLA could not be observed in positive ion electrospray ionisation LC-MS. Due to the ionisability of the labile carboxylic acid O-H bond, negative mode electrospray of the carboxylic acid may have been preferable for monitoring of the reaction, but this was not explored. Racemic 4-HPLA-*S*-NAC



Scheme 3.1 The synthesis of racemic 4-HPLA-*S*-NAC (**86**).

(**86**) was synthesised by coupling 4-hydroxyphenyllactic acid (**56**) to commercially available *N*-acetyl cysteamine (NAC, **64**) (scheme **3.1**). 4-HPLA-*S*-pantetheine (**51**) was used as previously synthesised in section 2.1.1.

Attempted halogenation of the small molecules was carried out as described for intact protein mass spectrometry, with 4-HPLA-*S*-AerA(PCP) replaced by the appropriate molecule. In both cases, only small amounts of turnover to the mono-chlorinated species, **87** and **88**, was observed (approximately 10% and 16% conversion of **86** and **51**, respectively) in 1 h following quenching by precipitation by precipitation of the protein with MeOH and UHPLC-ESI-Q-TOF-MS analysis (figure **3.1 B & G**).

Monitoring of halogenation of small molecules in this manner is not directly comparable to the results for intact mass spectrometry, as 500 µM of the small molecule was used to generate enough product for detection, compared to the 50 µM of 4-HPLA-*S*-AerA(PCP) used for intact protein mass spectrometry. Nevertheless, the low turnover with these small molecule analogues implies that protein-protein interaction is important for the activity of AerJ. It would be of interest in future work to carry out the same reaction with 500 µM 4-HPLA-*S*-AerA(PCP) to allow direct comparison of the degree of halogenation with and without tethering to the carrier protein. However, the AerA PCP domain has been typically not stable at the higher concentrations that would be required for this.

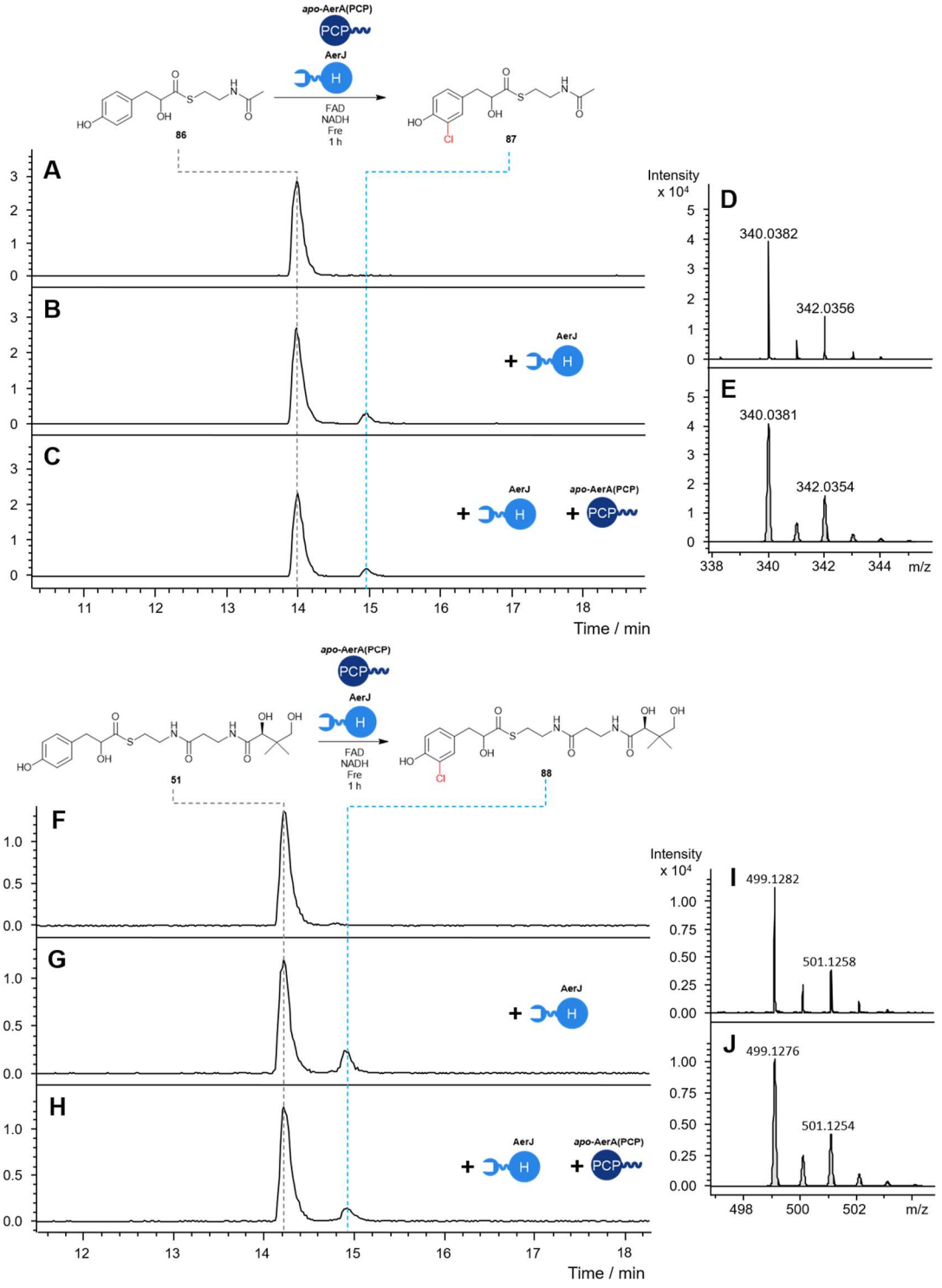


Figure 3.1 UHPLC-ESI-Q-TOF-MS analyses of halogenation of small molecule 4-HPLA-*S*-AerA(PCP) mimics by AerJ. (**A**) Extracted ion chromatogram of control reaction containing **86** in the absence of AerJ. (**B**) as (**A**) with AerJ added. (**C**) as (**B**) with *apo*-AerA(PCP) added. EICs in (**A**)-(**C**) are the sum of 284.0957, 306.0776, 318.0567, 340.0386, 352.0177 and 373.9997 ± 0.01 Da. (**D**) Mass spectrum resulting from the mono-chlorinated product **87** highlighted with the blue dashed line in (**B**) and (**C**). (**E**) Simulated mass spectrum for the $[M+Na]^+$ ion derived from **87** with peak height 4000 A.U. and width 0.1 Da. (**F**) Extracted ion chromatogram of control reaction containing **51** in the absence of AerJ. (**G**) as (**F**) with AerJ added. (**H**) as (**G**) with *apo*-AerA(PCP) added. EICs in (**F**)-(**H**) are the sum of 443.1852, 465.1671, 477.1462, 499.1282, 511.1073 and 533.0892 ± 0.01 Da. (**D**) Mass spectrum of the $[M+Na]^+$ ion resulting from **88** highlighted with the blue dashed line in (**G**) and (**H**). (**J**) Simulated mass spectrum for the $[M+Na]^+$ ion derived from **88** with peak height 1000 A.U. and width 0.1 Da.

A potential explanation for the lower turnover observed with the small molecule carrier protein-thioester mimics could be that binding of the AerA PCP domain induces a conformational change in AerJ, which promotes halogenation. As a simple test to determine whether a similar conformational change could be occurring with AerJ in the presence of the AerA PCP domain, the *apo*-protein was introduced to the small molecule halogenation reactions in five times excess of the halogenase. An increase in turnover compared to the case without the PCP domain could suggest that the halogenase is stabilised by binding of the PCP domain in a conformation that promotes halogenation. However, no significant increase in halogenation was observed (figure **3.1 C & H**). Indeed, a small decrease in the degree of halogenation was observed, possibly because binding of the AerA PCP domain to the halogenase blocked the active site channel preventing the NAC and pantetheine thioesters from accessing the active site.

The low degree of turnover of small molecule 4-HPLA-*S*-AerA(PCP) mimics suggests that AerJ may not be appropriate for biocatalytic application without significant engineering. Because of the capability of AerJ to halogenate carrier protein-tethered substrates, focus should instead be placed on use in biosynthetic pathway engineering applications.

3.1.2. Protein-protein interactions play a key role in directing halogenation in aeruginosin biosynthesis

In section 2.3.3., it was shown that Ile-4-HPLA-*S*-AerB(PCP) could not be significantly halogenated by AerJ, indicating that 4-HPLA-*S*-AerA(PCP) is almost certainly the substrate of the halogenase and that halogenation occurs before condensation. However, it remained unclear whether the tethered substrate or the carrier protein is the predominant factor responsible for the selectivity of the halogenase, or indeed whether both play a role.

To investigate this, the substrates tethered to the AerA and AerB PCP domain were switched, with the post-condensation substrate tethered to the AerA PCP domain and the pre-condensation substrate tethered to the AerB PCP domain, and the extent of halogenation by AerJ was compared to that for the substrates

tethered to their natural carrier protein domains. 4-HPLA-*S*-AerB(PCP) and all four diastereomers of Ile-4-HPLA-*S*-AerA(PCP) were chemoenzymatically synthesised by incubating CoaA, CoaD, CoaE and Sfp enzymes with the appropriate PCP domain and pantetheine thioester (**51** or **72**). Following loading, halogenations were carried out as previously described, with peroxidase added, and quenched after 1 hour by addition of a 10 % formic acid solution to a final concentration of 1 %. The extent of halogenation was then assessed by UHPLC-ESI-Q-TOF-MS analysis of the intact protein. To quantitate the extent of halogenation, the peak heights at the expected mass for the non-, mono- and di-chlorinated states in the charge-state deconvoluted mass spectra of each species were measured. Halogenation of 4-HPLA-*S*-AerA(PCP) with AerJ K156A was also conducted to establish background ion intensities resulting from oxidised or hydroxylated species, as discussed in section 2.1.4. Measurements were conducted in triplicate and the mean and standard deviation was determined.

4-HPLA-*S*-AerA(PCP) is 88% di-chlorinated and 10% monochlorinated, whereas no more than 40% dichlorination was observed for any of the Ile-4-HPLA-*S*-AerA(PCP) diastereomers (figure **3.2**). This demonstrates that the Ile-4-HPLA-*S*- substrates are less well tolerated by AerJ than the natural 4-HPLA-*S*- substrate, as may be expected. Interestingly, diastereomers 2 and 3 give comparable degrees of chlorination when tethered to the AerA PCP domain, whilst diastereomers 1 and 4 also give comparable chlorination levels which in turn are both lower than diastereomers 2 and 3. These similarities correlate with the stereochemistry of the isoleucyl moieties determined from analysis of ^{1}H and ^{13}C NMR spectroscopic data described in section 2.3.3, which showed that diastereomers 2 and 3 possess the D-*allo*- configuration and diastereomers 1 and 4 are the L-configured diastereomers. These results suggest that the stereochemistry of the isoleucine moiety can significantly affect the activity of AerJ. This may be a common trait when considering activity towards other dipeptide-like substrates and may need to be accounted for if attempting to use AerJ to chlorinate other non-ribosomal peptides in future bioengineering work. It also suggests that the stereochemistry of the 4-HPLA hydroxyl group does not significantly affect turnover, as the

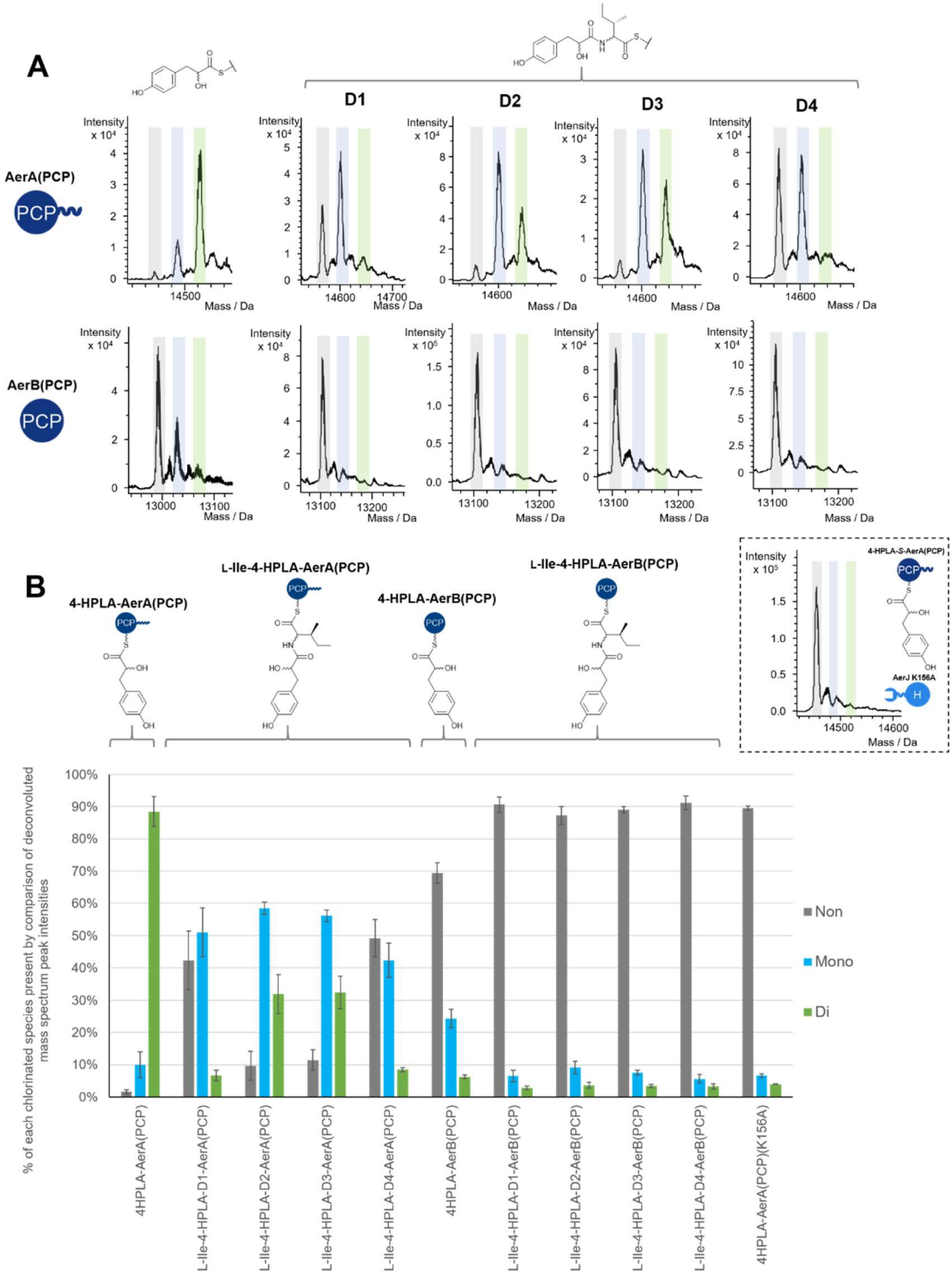


Figure 3.2 Comparison of degrees of halogenation by AerJ of AerA(PCP) and AerB(PCP) with 4-HPLA or L-Ile-4-HPLA diastereomers loaded as pantetheine-tethered substrates. (**A**) Charge state-deconvoluted mass spectra from UHPLC-ESI-Q-TOF-MS analysis of the products of incubation of substrate-tethered carrier protein domains with AerJ, Fre, NADH and FAD for 1 h following quenching with 1% formic acid. *Inset*: the reaction of 4-HPLA-S-AerA(PCP) with inactive halogenase mutant AerJ K156A. (**B**) Plot of percentage of non-, mono- and di-chlorinated species observed in the mass spectra shown in (**A**) based on the intensity count of the corresponding mass. Error bars show the mean ± standard deviation calculated from three independent experiments.

diastereomers with the same isoleucine stereochemistry but different 4-HPLA stereochemistry had similar chlorination profiles and were within a standard deviation of each other in all cases. This is consistent with AerJ being able to fully di-chlorinate racemic 4-HPLA-*S*-AerA(PCP). On the other hand, 4-HPLA-*S*-AerB(PCP) has reduced levels of chlorination compared to 4-HPLA-*S*-AerA(PCP), as only 24% mono- and 6% di-chlorinated species were produced. Furthermore, no significant halogenation of any diastereomers of Ile-4-HPLA-*S*-AerB(PCP) was observed. This suggests that AerJ preferentially acts upon the AerA PCP domain instead of the AerB PCP domain.

From these results, it is clear that the nature of the tethered substrate and the carrier protein domain are both important determinants of recognition by the halogenase. However, since low degrees of chlorination were observed for any substrate attached to the AerB PCP domain, it is clear that the nature of the carrier protein domain is a key factor in halogenase recognition and that halogenation exclusively occurs on AerA in aeruginosin biosynthesis. This corroborates with the results of investigations of a number of other βHDD-bearing catalytic domains. For instance, the standalone condensation domain Bamb_5915 from enacyloxin biosynthesis accepts several DHCCA (**37**) analogues and acyl groups as acceptor and donor substrates,[29] respectively, yet will only productively condense acyl donors tethered to its native interaction partner, the Bamb_5917 PCP domain. Acyl donors tethered to the Bamb_5919 ACP domain are not accepted, necessitating a specialised KS^0 domain to transfer the acyl donor between the two to enable the Bamb_5915 to catalyse condensation with DHCCA (**37**).[29,123] Similarly, the standalone C domain Kj12C from rhabdopeptide biosynthesis will condense peptide chains of a number of lengths and amino acid compositions,[126,132] yet almost exclusively accepts substrates tethered to the Kj12B PCP domain but not the Kj12A PCP domain, since the SLiM of the former possesses greater affinity for the Kj12C βHDD.[126,132] It may be that relaxed specificity towards carrier protein-tethered substrates is a common trait of βHDD-bearing catalytic domains, with selective protein-protein recognition of complementary docking domains directing chain assembly. Characterisation of the substrate specificity of additional examples

of βHDD-bearing enzymes will be necessary to examine this hypothesis further.

These results give the first indication that AerJ has relaxed specificity towards non-native PCP-tethered substrates. The ability of AerJ to accept non-native substrates gives a good indication that it could be used to engineer pathways that incorporate a variety of phenol-containing precursors, particularly in other NRPS assembly lines. A more comprehensive study of the substrate scope of the halogenase is currently underway with particular focus on common phenolic moieties found in natural products. This will give a clearer understanding of the types of chlorinated compounds that could be generated by pathway engineering using AerJ and help identify target pathways to apply these approaches to.

3.1.3. Halogenation of 4-HPLA tethered to heterologous carrier proteins by AerJ

Work in this section was conducted in collaboration with Dr. Shanshan Zhou, Dr. Christopher Fage and Munro Passmore who provided aliquots of Ni-NTA purified Sven_0512, BlmVIII/TycA and PcdC carrier protein domains respectively.

Following demonstration that the nature of the carrier protein is a key factor controlling halogenase specificity in aerguinosin biosynthesis, halogenation of 4-HPLA tethered to several different SLiM-bearing carrier proteins was examined. This was undertaken to determine whether differences in amino acid sequence can be linked to differences in catalytic activity and probe whether the halogenase could interact productively with heterologous carrier proteins, with a view to future biosynthetic engineering applications.

Several SLiM-bearing carrier proteins were used. The Bamb_5917 PCP domain from the enacyloxin assembly line was overproduced in BL21(DE3) *E. coli* using a pET151-Bamb_5917(PCP) expression vector previously constructed in the Challis group.[29,123] The Sven_0512 PCP domain from the watasemycin NRPS in *Streptomyces venezuelae* was kindly provided as a Ni-

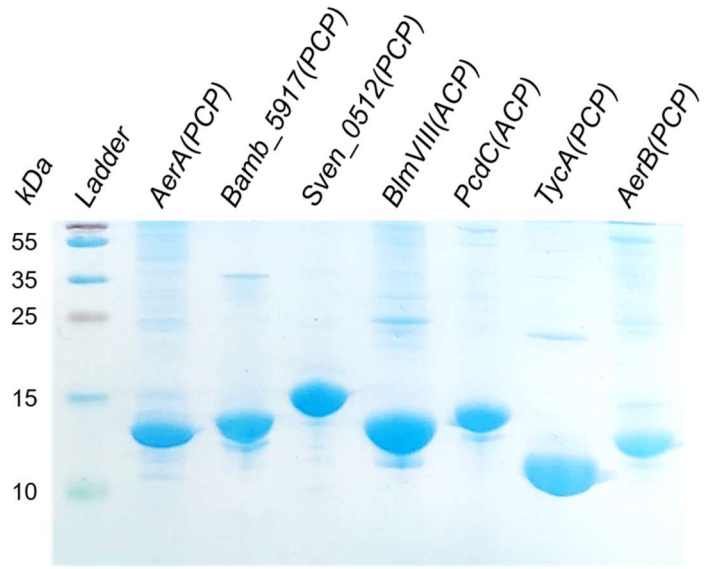


Figure 3.3 SDS-PAGE analysis of the heterologous carrier protein domains tested for halogenation in this section. Expected masses from left to right: 14.0 kDa, 14.5 kDa, 15.5 kDa, 11.7 kDa, 15.3 kDa, 11.2 kDa and 12.5 kDa.

NTA-purified protein by Dr. Shanshan Zhou. Similarly, the BlmVIII ACP domain involved in bleomycin biosynthesis in *Streptomyces verticillus* was kindly provided as a Ni-NTA-purified protein by Dr. Christopher Fage. These three carrier protein domains had been previously identified as bearing SLiMs by the hidden Markov model that was used to identify the βHDD-SLiM interfaces in the aeruginosin NRPS.[123] The PcdC ACP domain, a recently identified SLiM-bearing ACP domain from an NRPS involved in FR901375 biosynthesis in *Pseudomonas chlororaphis* was also tested. To validate the results obtained for the AerB PCP domain, the TycA PCP domain was also obtained from Dr. Christopher Fage. Similarly to the AerB PCP domain, this carrier protein is not located at the C-terminus of a subunit and does not appear to harbour a SLiM. These recombinant proteins were analysed by SDS-PAGE gel electrophoresis prior to use (figure **3.3**).

Using the same methodology for the halogenation reaction described in section 3.1.2, each carrier protein domain was loaded with 4-HPLA-*S*-pantetheine (**51**), then incubated with AerJ and cofactors for 1 h before quenching. Reactions were conducted in triplicate and analysed as described

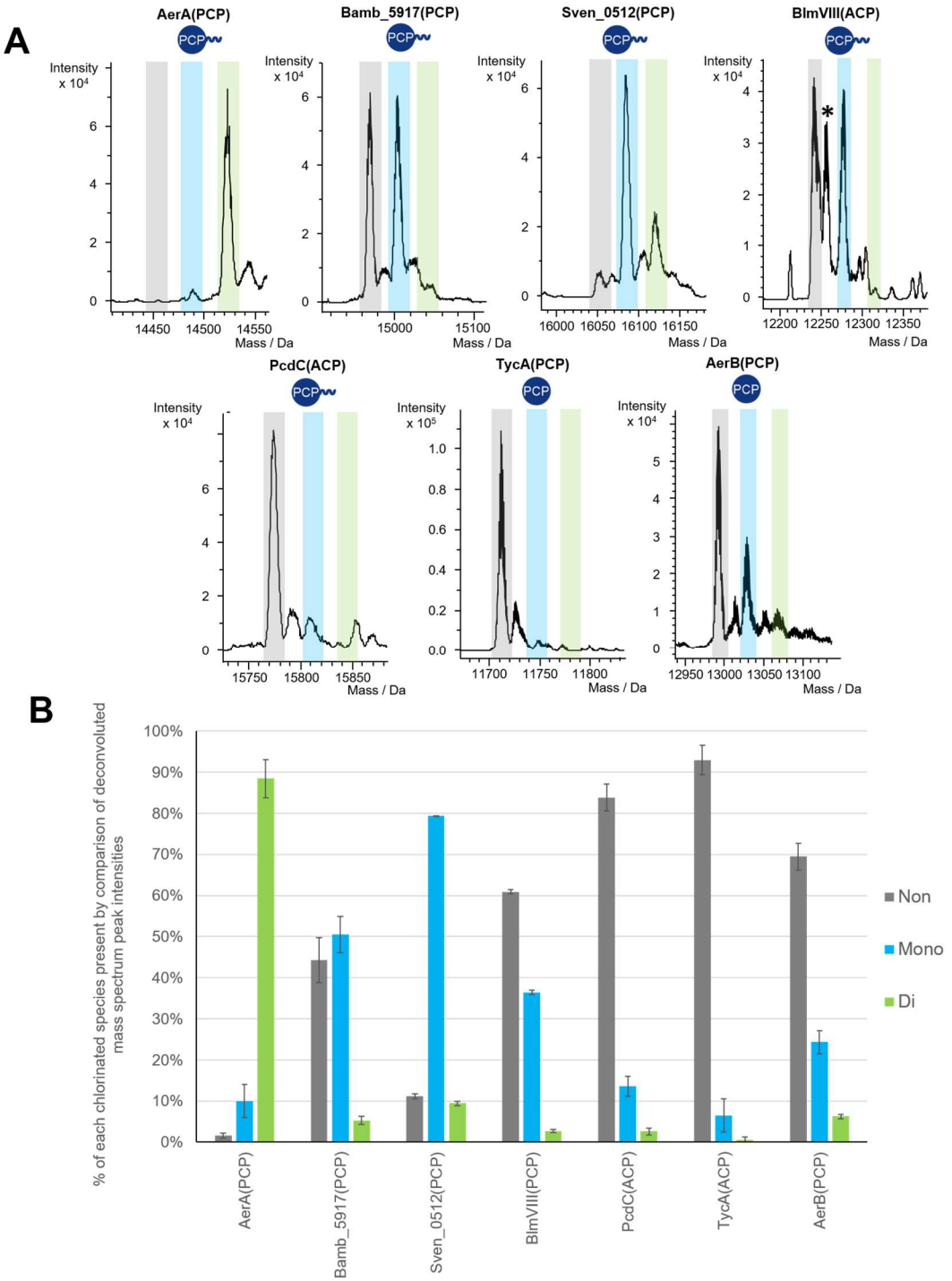


Figure 3.4 Comparison of degrees of halogenation by AerJ of 4-HPLA tethered to carrier protein domains from different biosynthetic pathways. (**A**) Charge state-deconvoluted mass spectra from UHPLC-ESI-Q-TOF-MS analysis of the products of incubation of substrate-tethered carrier protein domains with AerJ, Fre, NADH and FAD for 1 h following quenching with 1% formic acid. (**B**) Plot of percentage of non-, mono- and di-chlorinated species observed in the mass spectra shown in (**A**) based on the intensity count of the corresponding mass. Error bars show the mean ± standard deviation calculated from three independent experiments.

previously. The results are shown in figure **3.4** Notably, all of the heterologous carrier protein domains were less efficiently chlorinated than the native AerA PCP domain in the halogenation reaction. 4-HPLA-*S*-Sven_0512(PCP), 4-

HPLA-*S*-Bamb_5917(PCP), and 4-HPLA-*S*-BlmVIII(ACP) were all predominantly mono-chlorinated in descending order of efficiency. A small degree of di-chlorinated product was detected for the 4-HPLA-*S*-Sven_0512(PCP) species, suggesting that dichlorination is likely not prevented by a mechanistic effect and that differences in extent of halogenation is most likely to be due to the difference in protein-protein interaction efficiency. Interestingly, 4-HPLA-*S*-PcdC(ACP), although it bears a

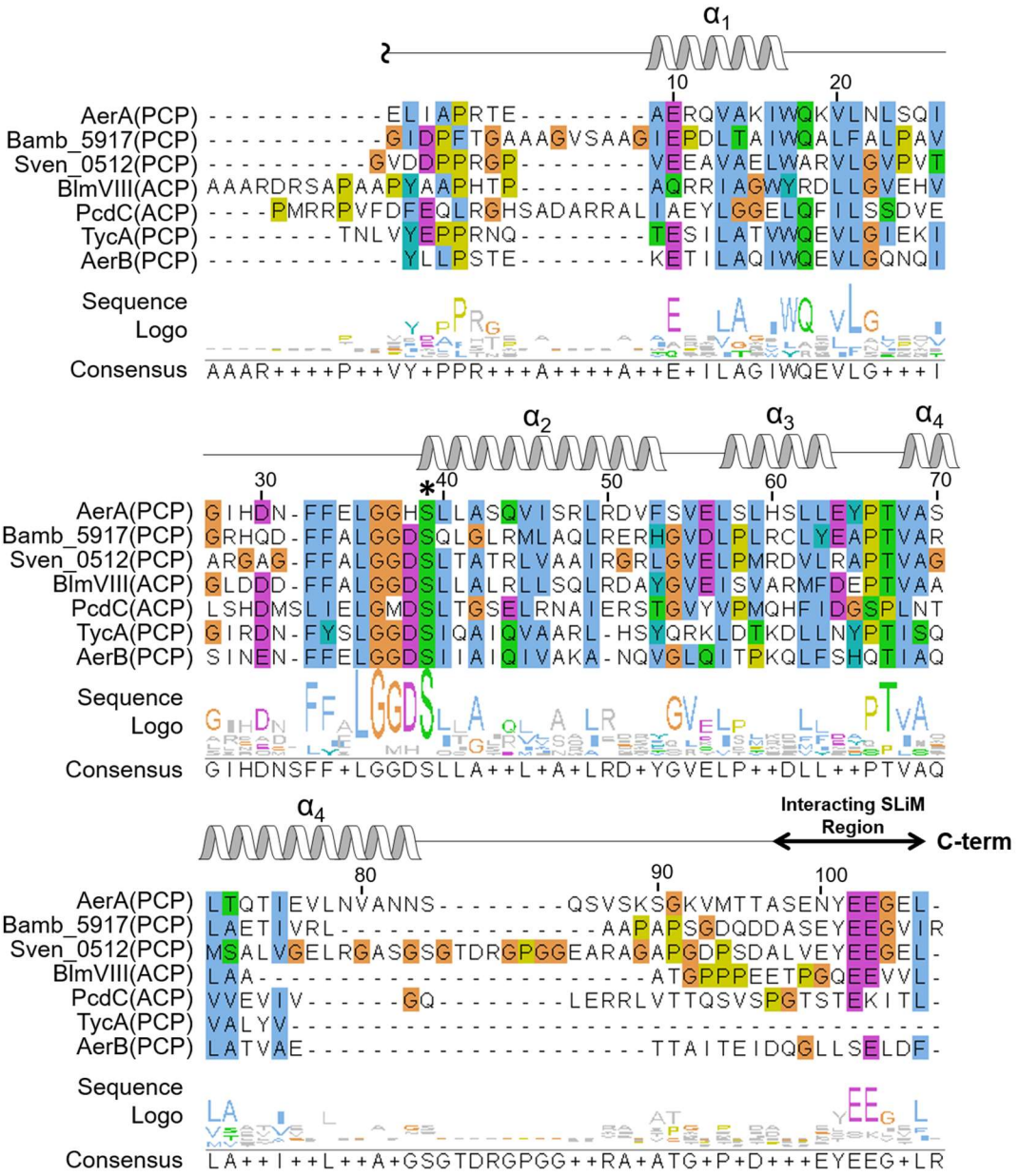


Figure 3.5 Sequence alignment of the heterologous carrier protein domains used in this study. Above the alignment, the secondary structure predicted for the AerA PCP domain from PsiPred 4.0 predictions is displayed. A sequence logo and consensus sequence are displayed below the alignment. The conserved serine to which the 4'-phosphopantetheine arm is tethered is highlighted with an asterisk (*). Residue numbering is relative to that from the AerA PCP domain.

putative SLiM was not significantly chlorinated, suggesting poor compatibility with AerJ. Only trace amounts of chlorination of 4-HPLA-*S*-TycA(ACP) was observed, even less than that seen with the AerB PCP domain, further highlighting the poor compatibility of AerJ with PCP domains not located at SLiM-βHDD intersubunit boundaries.

Comparing the sequences of these carrier protein domains revealed potential explanations for the carrier protein domain selectivity of AerJ (figure **3.5**). Although the SLiM regions of the AerA PCP, Bamb_5917 PCP, Sven_0512 PCP and BlmVIII ACP domains have very similar sequences, the length of the disordered region linking to the SLiM is highly variable. For instance, the BlmVIII ACP domain has a linker that is approximately 16 amino acids shorter than that of the AerA PCP domain, whilst the Sven_0512 PCP domain's linker is 8 amino acids longer. Having a linker with a mismatching length may affect the ability of the pantetheine-tethered substrate to access the halogenase active site when the SLiM is bound to the βHDD, especially in the case of those with shorter linkers than the AerA PCP domain. It is interesting to note that the PCP domain with the linker closest in size to the AerA PCP domain, the Sven_0512 PCP domain, gave the largest degree of halogenation, consistent with this hypothesis. The PcdC ACP domain has both a shorter linker and possesses a positively charged lysine residue in place of a conserved glutamate residue in its SLiM, which may explain the low amount of halogenation. As the AerJ βHDD has a SLiM-interacting β2 strand containing entirely positively charged amino acid residues, a repulsive electrostatic interaction likely drastically reduces binding affinity of the PcdC SLiM compared to the AerA SLiM.

There are, however, many other amino acid substitutions elsewhere in the carrier protein domains. In particular, helix 2, which houses the conserved serine residue that the pantetheine arm is tethered to, shows poor conservation between the carrier protein domains used in this study. Further examination of the interactions between this helix and AerJ is necessary to relate halogenase activity to carrier protein structure.

3.1.4. SLiM Swapping Experiments

Work in this section was conducted in collaboration with Dr. Shanshan Zhou, who provided the Bamb_5917 PCP domain fused to the AerA SLiM as a Ni-NTA purified protein.

To further probe the necessity of SLiM-βHDD binding in recruitment of AerJ, chimeric proteins with PCP domains fused to different SLiMs were created and evaluated in halogenation assays. The AerA, Bamb_5917 and Sven_0512 PCP domains were targeted for generation of chimeras, with the SLiMs of the latter two grafted onto the AerA PCP domain and the AerA SLiM grafted onto the Bamb_5917 and Sven_0512 PCP domains.

The final 10 amino acid residues of the disordered regions appended to carrier protein domains were chosen as the excision site for generating chimeras as these residues had been previously demonstrated in the Bamb_5917 PCP domain to be involved in forming part of the interaction interface with the Bamb_5915 βHDD by Kosol *et al.* in NMR titration experiments.[123] This excision site was in agreement with secondary structure and C-terminal disorder predictions made by the PsiPred 4.0 and Dynamine webservers, respectively, in which this region was shown to be disordered (appendix 11).[213,214] The AerA PCP domain-SLiM chimeras were constructed by first

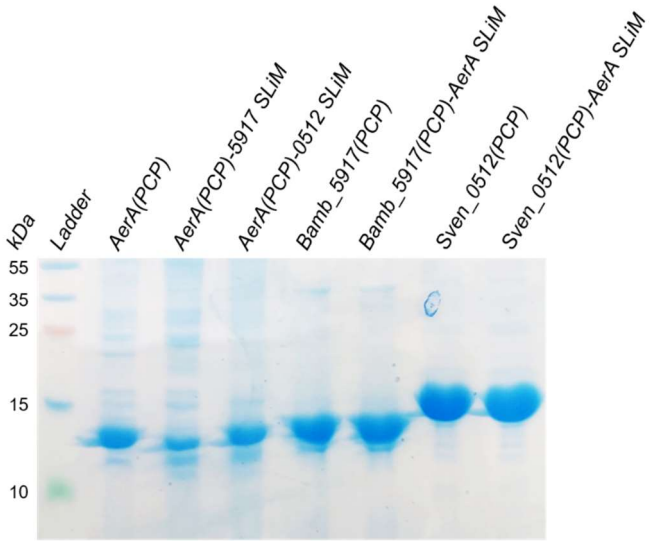


Figure 3.6 SDS-PAGE analysis of the carrier protein-SLiM chimeras tested for halogenation in this section. Expected masses from left to right: 14.0 kDa, 14.0 kDa, 14.0 kDa, 14.5 kDa, 14.5 kDa, 15.5 kDa, 15.5 kDa.

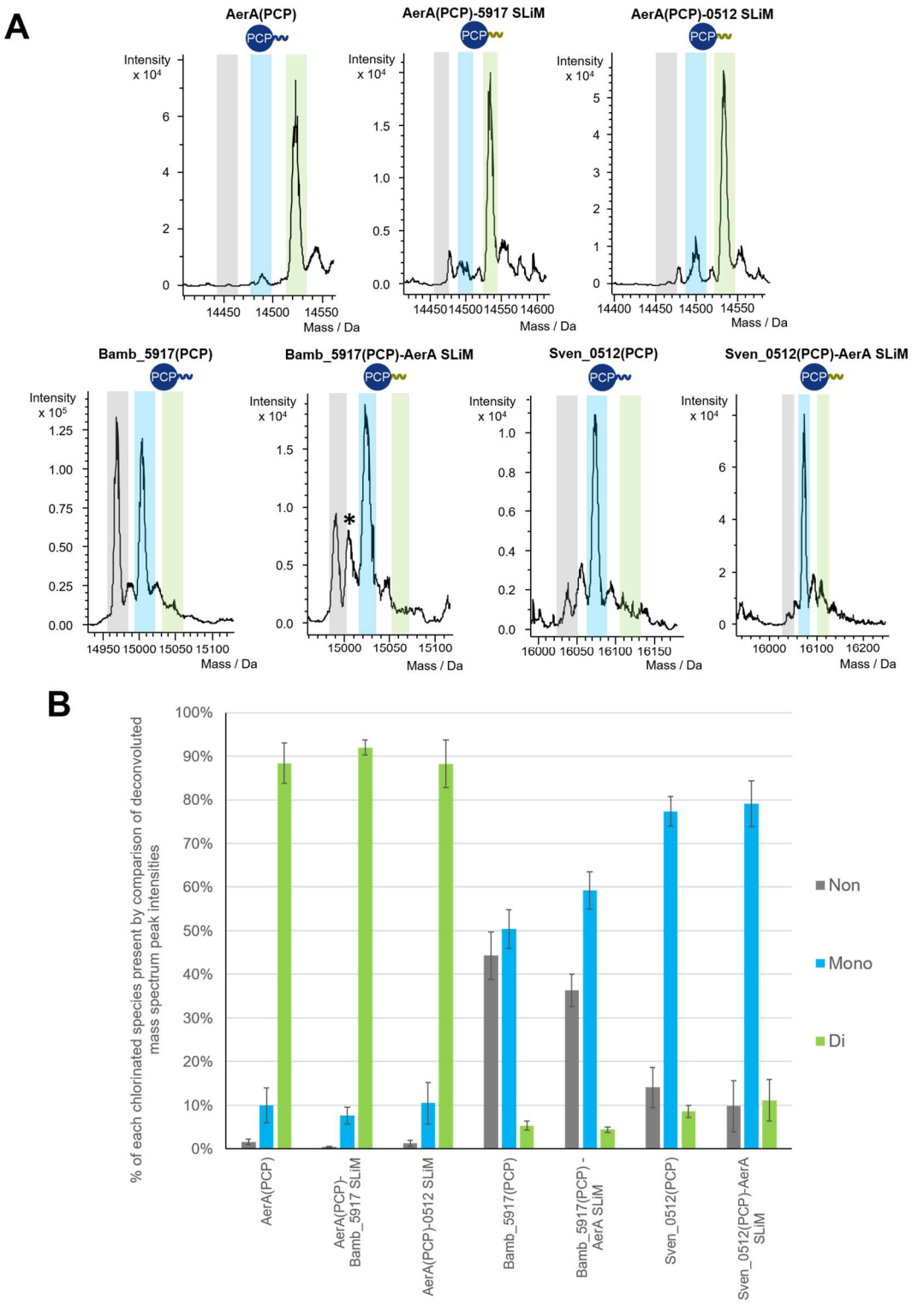


Figure 3.7 Comparison of degrees of halogenation by AerJ of 4-HPLA tethered to carrier protein domain-SLiM chimeras. (**A**) Charge state-deconvoluted mass spectra of the products of incubation with AerJ for 1 h following quenching with 1% formic acid. (**B**) Plot of percentage of non-, mono- and di-chlorinated species observed in the mass spectra shown in (**A**) based on the intensity count of the corresponding mass. Error bars show mean ± standard deviation calculated from three independent experiments.

performing deletion mutagenesis in pET24a-(+)-AerA(PCP) to remove the 10 C-terminal amino acid residues. Overproduction of the resulting

AerA(PCP)ΔSLiM mutant was attempted, but it was found to be insoluble. The constructs with the AerA PCP domain fused to the Sven_0512 and Bamb_5917 SLiMs were created from the pET24a-(+)-AerA(PCP)ΔSLiM plasmid by insertion mutagenesis using primer overhangs to insert the nucleotide sequences coding for the appropriate SLiM. The fusion of the Bamb_5917 PCP domain with the AerA SLiM was created and provided as a Ni-NTA-purified protein by Dr. Shanshan Zhou. The fusion of the Sven_0512 PCP domain with the AerA SLiM was created from the pET151-Sven_0512(PCP) construct previously produced by the Challis group.[123] PCP domain-SLiM chimeras were overproduced in *E. coli* BL21(DE3) and purified by nickel affinity chromatography (figure **3.6**).

Using the methodology for halogenation described in section 3.1.2, the PCP domain-SLiM chimeras were compared for degree of halogenation of the appended 4-HPLA thioester (figure **3.7**). Interestingly, swapping of the SLiMs did not significantly impact the degree of chlorination observed in UHPLC-ESI-Q-TOF-MS analyses. Chimeras of the AerA PCP domain with the Bamb_5917 and Sven_0512 SLiMs both produced predominantly di-chlorinated species at the same level within the limits of error. Similarly, chimeras of the Bamb_5917 and Sven_0512 PCP domains with the AerA SLiM did not support significantly higher levels of chlorination than the native carrier protein domains.

These results can be rationalised by considering the similarity of the final ten amino acid residues of AerA, Bamb_5917 and Sven_0512 (figure **3.8 A**). Excluding the C-terminal positively-charge arginine of Bamb_5917, which does not align with the other proteins, all three SLiMs possess negatively charged sequences with high identity (figure **3.8 B**). Replacing these SLiMs with more divergent SLiMs may result in greater variation in halogenation by AerJ. The carrier protein domain ArfA from the arthrofactin biosynthetic pathway possessing a positively-charged SLiM was identified in the hidden

B

Percentage Identity Matrix

	AerA			
AerA	100%	**Bamb_5917**		
Bamb_5917	44%	100%	**Sven_0512**	
Sven_0512	55%	70%	100%	**ArfA**
ArfA	33%	0%	11%	100%

Figure 3.8 (**A**) Sequence alignment of the SLiMs used to create PCP domain-SLiM chimeras. (**B**) Percentage identity matrix for the four SLiMs.

Markov model by Kosol *et al.* (figure **3.8**),[123] and a chimera of the AerA PCP domain with this divergent SLiM was targeted to investigate whether this might produce an observable decrease in halogenase efficiency. A plasmid encoding a chimera of the AerA PCP domain with the ArfA SLiM was created by insertion mutagenesis of the AerA(PCP)ΔSLiM plasmid, but the recombinant protein was insoluble.

Also, as mentioned in section 3.1.3, the linker length between α-helix 4 and the SLiM region in these three carrier protein domains is very variable. Swapping these linker regions may create larger differences in halogenase effciency and give further insights into the effect the length of this linker has on protein-protein interaction interfaces at SLiM-βHDD subunit interfaces.

3.1.5. Mutagenesis of the AerJ β-hairpin docking domain

Previous studies of SLiM-βHDD interfaces have successfully utilised point mutants in biophysical[101,126,135] and *in vivo*[130] characterisations to understand the structural basis for productive interface formation. Furthermore, constructs lacking docking domains have been demonstrated to form less efficient interactions than their docking domain-bearing counterparts.[29,118] This evidence has been used to rationalise the importance of the studied docking domain in the fidelity of their respective biosyntheses. Several mutant proteins were generated to investigate the SLiM-βHDD interface between the AerA PCP domain and AerJ in biochemical assays.

An AerJ construct lacking the βHDD (AerJΔβHDD) was produced (appendix 2). The excision site for the AerJ βHDD was determined by secondary structure predictions by PsiPred 4.0 and a homology model of the N-terminal

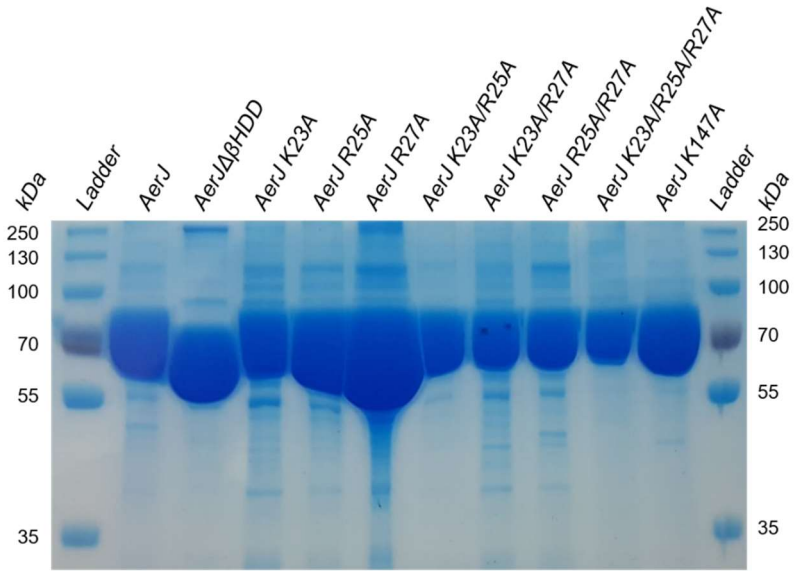


Figure 3.9 SDS-PAGE analysis of AerJ mutants. Expected masses for AerJ = 74.3 kDa, AerJΔβHDD = 67.8 kDa, AerJ K23A to AerJ R25A/R27A = 74.2 kDa, AerJ K23A/R25A/R27A = 74.1 kDa and AerJ K156A = 74.1 kDa.

region generated by Phyre2 (appendix 12).[194,215] Recognition of the conserved αββαα fold allowed identification of the putative βHDD, and an excision site was chosen close to the N-terminus of the third α-helix. A pET24a-(+)-AerJΔβHDD construct was generated by deletion mutagenesis from the pET24a-(+)-AerJ plasmid and confirmed by Sanger sequencing. To understand the importance of electrostatic interaction in mediating docking between the AerA SLiM and the AerJ βHDD, electrostatic residues present on the β2 strand of the βHDD were targeted for mutagenesis as these have been previously been demonstrated to be important for salt-bridge formation between the docking domains.[101,126] The β2 strand was also identified with the aid of the PsiPred 4.0 secondary structure predictions and Phyre2 homology model.[194,215] Mutations of K23, R25 and R27 to alanine were achieved by point mutagenesis of the pET24a-(+)-AerJ vector. Double mutants K23A/R25A, K23A/R27A, R25A/R27A and the triple mutant K23A/R25A/R27A were constructed from the single point mutants by sequential mutagenesis of the plasmids. These AerJ mutants were then overproduced in *E. coli* BL21(DE3) and purified using nickel affinity chromatography (figure **3.9**, appendix 3).

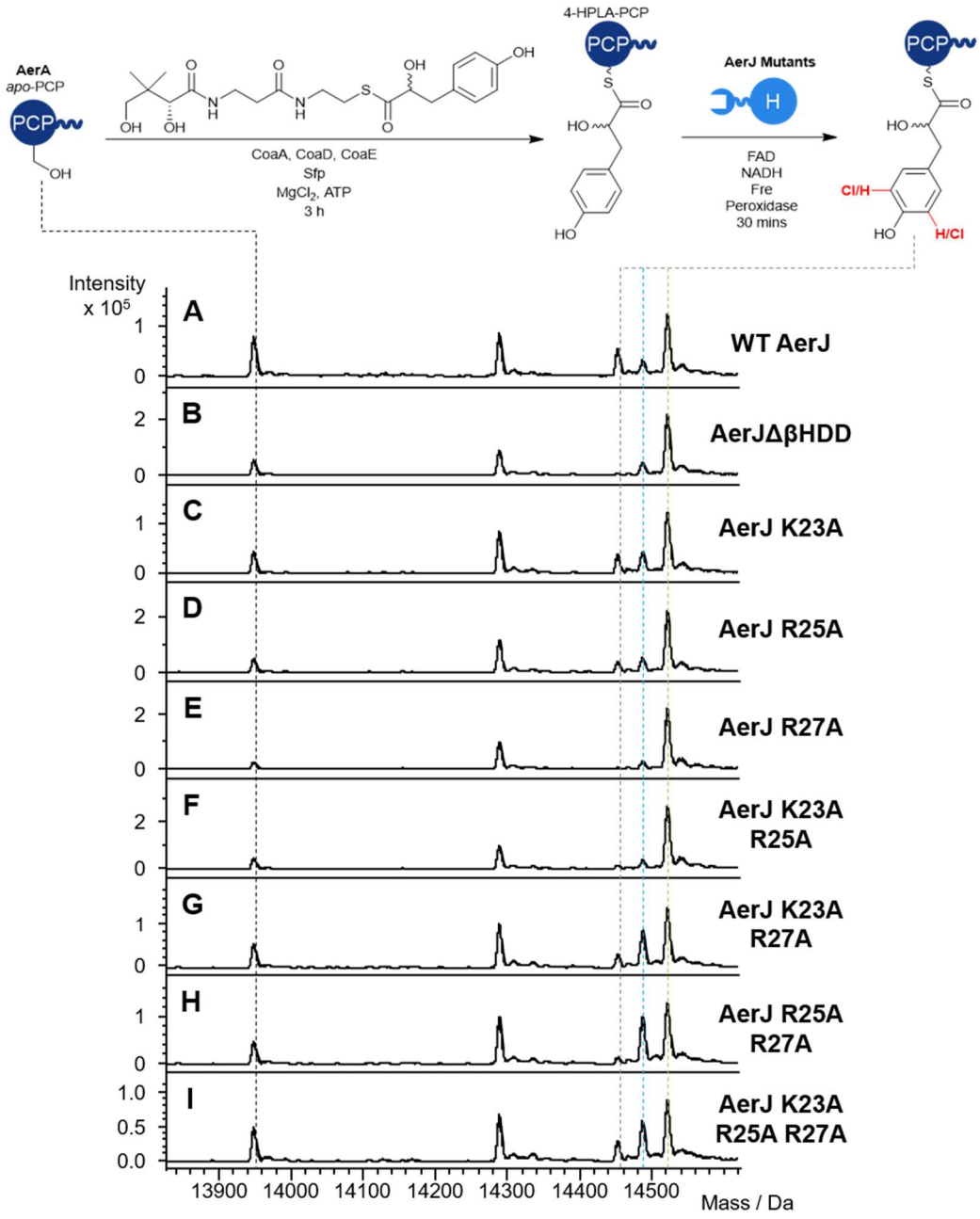


Figure 3.10 UHPLC-ESI-Q-TOF-MS analysis demonstrating halogenation of 4-HPLA-*S*-AerA(PCP) by AerJ point mutants *in vitro*. Charge-state deconvoluted spectra show halogenation by: (**A**) AerJ; (**B**) AerJΔβHDD; (**C**) AerJ K23A; (**D**) AerJ R25A; (**E**) AerJ R27A; (**F**) AerJ K23A/R25A; (**G**) AerJ K23A/R27A; (**H**) AerJ R25A/R27A; (**I**) AerJ K23A/R25A/R27A. Grey dashed line = non-chlorinated species, blue dashed line = mono-chlorinated species, green dashed line = di-chlorinated species.

To compare the efficiency of halogenation by each of the mutants, a single loading reaction of 4-HPLA-*S*-pantetheine (**51**) onto the AerA PCP domain was performed and then divided between each subsequent halogenation reaction to ensure equivalent extent of loading of the carrier protein domain. Then, the appropriate halogenase mutant was introduced with the relevant cofactors and peroxidase and incubated for 1 h at 25 °C before quenching by

addition of 1 % formic acid. The results of this experiment can be seen in figure **3.10**. All βHDD mutants displayed halogenase activity towards the carrier protein-tethered substrate. Unfortunately, comparing the mutants proved difficult, with multiple mutants appearing to provide a greater extent of halogenation than the wild type AerJ. This is inconsistent with results of previous studies of SLiM-βHDD interfaces in which mutations along the β2 strand always inhibit productivity of the appended catalytic domain.[123,126,130] Moreover, the AerJ mutant lacking the βHDD gave comparable turnover to the wild type enzyme. It was unclear whether these unusual results were an accurate comparison of the mutants or an artefact of the methodology used. In general, halogenation appeared to be most inhibited in the double and triple mutants, where a noticeable shift towards non- and mono-chlorinated 4-HPLA-*S*-AerA(PCP) can been observed. However, due to the ambiguity of results with other mutants, it is difficult to draw conclusions from this observation.

In separate experiments, AerJ and AerJΔβHDD were compared over multiple timepoints, both by quenching with 1% formic acid and UHPLC-ESI-Q-TOF-MS of the intact protein or by methylamine cleavage of the carrier protein-tethered thioester and analysis of the methylamide product by UHPLC-ESI-IT-MS (appendix 14, see sections 6.1.6.9 & 6.1.6.10 for detailed methodology). However, results proved to be inconsistent and no reproducible difference in degree of halogenation of 4-HPLA-*S*-AerA(PCP) by AerJ and AerJΔβHDD was observed (appendix 14).

Difficulties in comparing halogenation by the AerJ point mutants may result from inaccuracies in measuring protein concentration. To try to ensure that assays using mutants were robustly comparable, accurate measurements of the concentrations of each were required. Typically, protein concentrations are calculated by measuring absorbance at 280 nm using a spectrophotometer. The concentration can then be estimated from the Beer-Lambert law using predicted protein extinction coefficient based upon the number of 280 nm-absorbing tyrosine, tryptophan and cysteine residues present in the amino acid sequence. To confirm the accuracy of concentration measurements, a Bradford assay was conducted.[190] This method of concentration measurement

has the advantage that it does not rely upon estimated extinction coefficient calculations and cannot be affected by salts or other solution components. Although the Bradford assay and Nanodrop calculated concentrations did not give corresponding absolute concentrations, the ratio of concentrations between each mutant were in approximate agreement across the two methods (appendix 15).

However, SDS-PAGE analysis of each of these mutants diluted to the same concentrations as determined by these methods qualitatively suggested that the measurements were inaccurate as the intensity of each band was not equivalent (figure **3.9**). Additionally, attempted comparison of the mutants by CD spectroscopy gave spectra with different intensities attributable to differences in concentration (appendix 16). As it is crucial for such comparative experiments that all mutants are of the same concentration, in future investigations lyophilisation of the proteins and resuspension in a known amount of buffer would allow for the most accurate protein concentration measurements possible.

In NMR titrations conducted by Hacker *et al.*, it was determined that charge reversal of residues on the β2-strand of the βHDD more severely reduced binding affinity with the SLiM compared to mutation to alanine.[126] To try to determine whether charge reversal of β2-residues would provide a more obvious difference in halogenation efficiency, a mutant was created in which all positively-charged residues on the β2 strand were mutated to negatively-charged glutamate residues (AerJ K23E/R25E/R27E). However, this construct was insoluble. It is likely that altering the charges of the residues on the β2-strand of the βHDD could cause dramatic disruption of protein folding and promote aggregation, resulting in an insoluble protein.

3.1.6. Inhibition of the AerA-AerB SLiM-βHDD interface by addition of excess excised AerJ or AerB βHDD

Another approach to infer the importance of docking domain interactions is to inhibit the reaction *in vitro* using the excised docking domain. In the

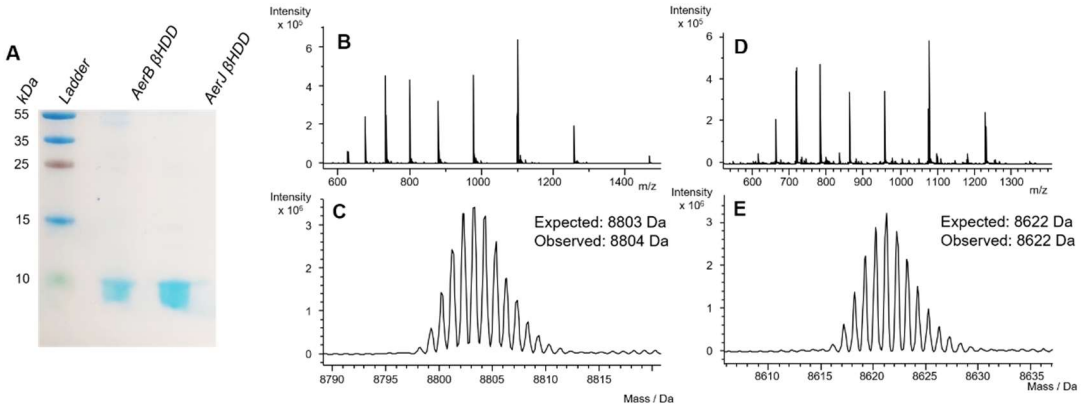
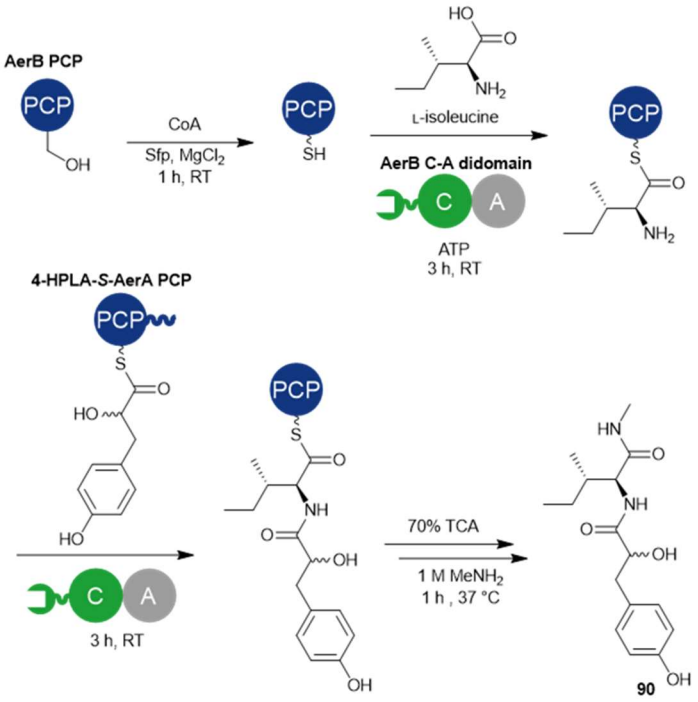


Figure 3.11 Characterisation of excised AerJ and AerB β-hairpin docking domains by SDS-PAGE and UHPLC-ESI-Q-TOF-MS. (**A**) SDS-PAGE analysis of the excised docking domains. (**B**) The mass spectrum of the AerJ βHDD. (**C**) Charge state-deconvoluted mass spectrum of the AerJ βHDD. (**D**) The mass spectrum of the AerB βHDD. (**C**) Charge state-deconvoluted mass spectrum of the AerB βHDD.

characterisation of the Bamb_5917-Bamb_5915 βHDD-SLiM interface by Masschelein *et al.*, addition of increasing concentrations either the excised SLiM or βHDD caused inhibition of product formation,[123] highlighting the importance of SLiM-βHDD binding in mediating carrier protein domain recruitment. A similar method could provide supplementary evidence for the necessity of the SLiM-βHDD interfaces in the aeruginosin biosynthetic system.

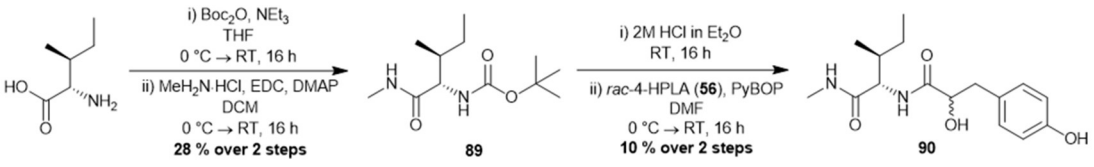
To do this, the excised β-hairpin docking domains of AerJ and AerB were produced. Expression constructs were generated *via* deletion mutagenesis of the pET24a-(+)-AerJ and pET24a-(+)-AerB(C) plasmids. The same boundary for excision for AerJ was used as described for the generation of AerJΔβHDD construct in section 3.1.5 (appendix 12), and the same rationale was applied for deciding the excision boundary for the βHDD from the AerB C domain (appendix 13). The proteins were overproduced in *E. coli* BL21(DE3) cell lines and purified by nickel-affinity chromatography (figure **3.11**).

An intact protein mass spectrometry-based experiment was not suitable for observing inhibition, as the amount of excised docking domain that could be added was fundamentally limited by the saturation limits of the ESI-Q-TOF instrument. Instead, an approach using methylamine cleavage of the pantetheine thioester was developed, adapted from Kittilä *et al.*[150] To understand the effect that the AerJ βHDD has in competing with the AerA-AerB



Scheme 3.2 Workflow for generation of methyl amide **90** following condensation of 4-HPLA-*S*-AerA(PCP) and L-Ile-S-AerB(PCP) catalysed by the AerB C-A didomain.

intersubunit interface, the *in vitro* experiment utilising the AerB C-A didomain for condensation of 4-HPLA-*S*-AerA(PCP) with L-Ile-AerB(PCP) as demonstrated in section 2.3.2 (scheme **3.2**) was used. The AerB PCP domain was converted to *holo*- form by incubating with CoA, Sfp and $MgCl_2$ and loaded with an L-isoleucyl group by incubating with L-isoleucine, the AerB C-A didomain and ATP for 3 h. Then, the appropriate excised βHDD was added, before 4-HPLA-*S*-AerA(PCP) was introduced to initiate the condensation reaction. The reaction mixtures were incubated at 25 °C for 5 minutes before quenching by addition of 10% trichloroacetic acid solution. The protein was pelleted by centrifugation, washed with 10% trichloroacetic acid and redissolved in a 1 M methylamine solution. The mixture was then lyophilised, redissolved in a known amount of methanol and the methyl amide product was analysed by UHPLC-ESI-IT-MS.



Scheme 3.3 Synthesis of an authentic standard of methyl amide **90**.

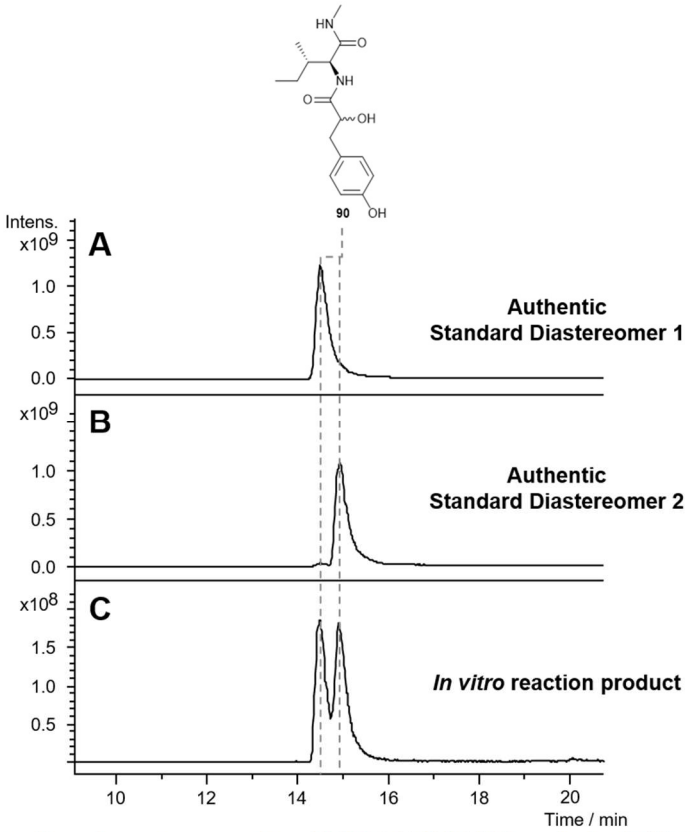


Figure 3.12 Extracted ion chromatograms from UHPLC-ESI-IT-MS analysis of: (**A**) One diastereomer of the synthetic standard **90**; (**B**) The other diastereomer of the synthetic standard **90**; (**C**) The product of methylamine cleavage of the pantetheine thioester of L-Ile-4-HPLA-*S*-AerB(PCP) following condensation of 4-HPLA-*S*-AerA(PCP) with L-Ile-4-HPLA-*S*-AerB(PCP) catalysed by the AerB C-A didomain as described in scheme **3.2**. EICs are the sum of 309.2; 331.2; 617.4; 639.4 ± 0.5 Da, referring to the expected $[M+H]^+$, $[M+Na]^+$, $[2M+H]^+$ and $[2M+Na]^+$ ions respectively.

To verify production of the expected methyl amides, authentic standards of the products were synthesised (scheme **3.3**). The amino group of L-isoleucine was protected with a *tert*-butoxycarbonyl group before coupling to methylamine to give **89**. The protecting group was then removed under acidic conditions and the resulting amine coupled to racemic 4-hydroxyphenyllactic acid (**56**) with PyBOP to give the expected methylamide product, **90**. The two diastereomers produced were then separated by semi-preparative HPLC. Comparison of the retention time of synthetic **90** to the products of the *in vitro* experiment confirmed production of the expected methyl amide compound (figure **3.12**). Both diastereomers were observed in the reaction because racemic 4-HPLA-*S*-pantetheine (**51**) was loaded onto AerA PCP domain. This result therefore also confirms the ability of the AerB C domain to accept both epimers of 4-hydroxyphenyllactic acid.

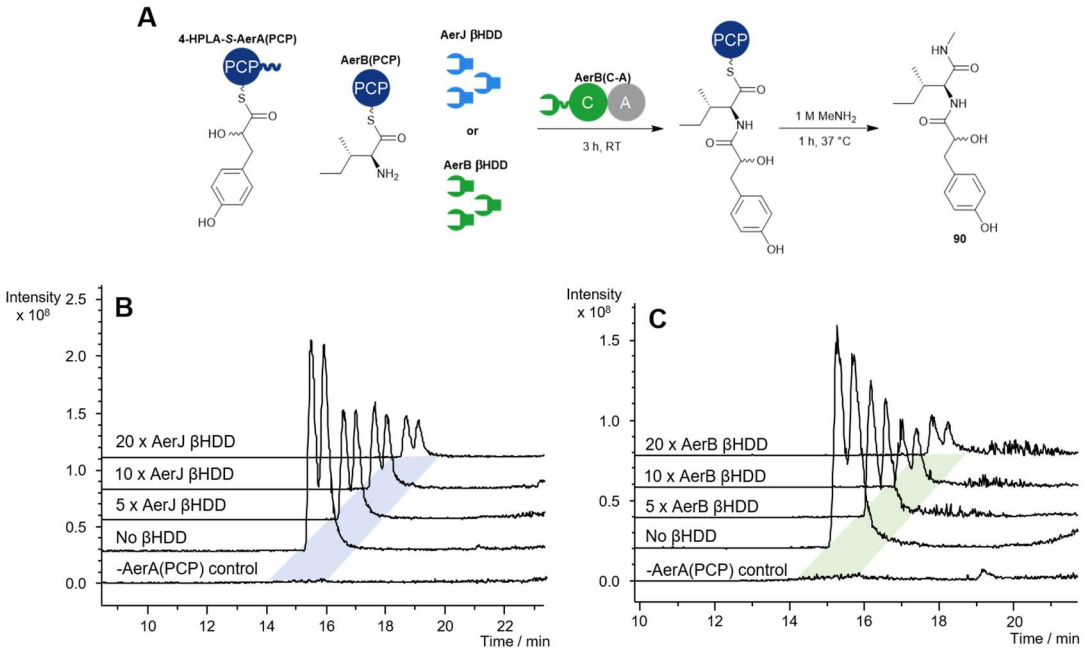


Figure 3.13 Inhibition of the AerB C-A didomain condensation activity by addition of the excised AerB or AerJ βHDDs. (**A**) Reaction scheme of the *in vitro* experiment. (**B**) Extracted ion chromatogram from UHPLC-ESI-IT-MS analysis demonstrating inhibition of condensation by introduction of AerJ βHDD. (**C**) Extracted ion chromatogram from UHPLC-ESI-IT-MS analysis demonstrating inhibition of condensation by introduction of AerB βHDD. EICs are the sum of 309.2; 331.2; 617.4; 639.4 ± 0.2 Da, referring to the expected $[M+H]^+$, $[M+Na]^+$, $[2M+H]^+$ and $[2M+Na]^+$ ions respectively.

Introduction of excised AerJ or AerB βHDDs to the reaction resulted in a significant decrease in methyl amide product formation (figure **3.13**). Increasing the amount of excised βHDD relative to the AerB C-A didomain caused further reduction in peak area of methyl amide product (figure **3.13**). This suggests that the excised docking domains interacting with the SLiM of the AerA PCP domain inhibits the condensation reaction catalysed by the AerB C domain.

To compare the difference in inhibition between the excised AerJ and AerB βHDDs, the area under the two product peaks was measured and normalised for differences in the absolute peak area by calculating percentage product formation for reactions with 5 times, 10 times and 20 times excised βHDD with respect to the AerB C-A didomain compared to a reaction where no excised βHDD was present (figure **3.14**). Reactions were performed in triplicate and the means and standard deviations were calculated.

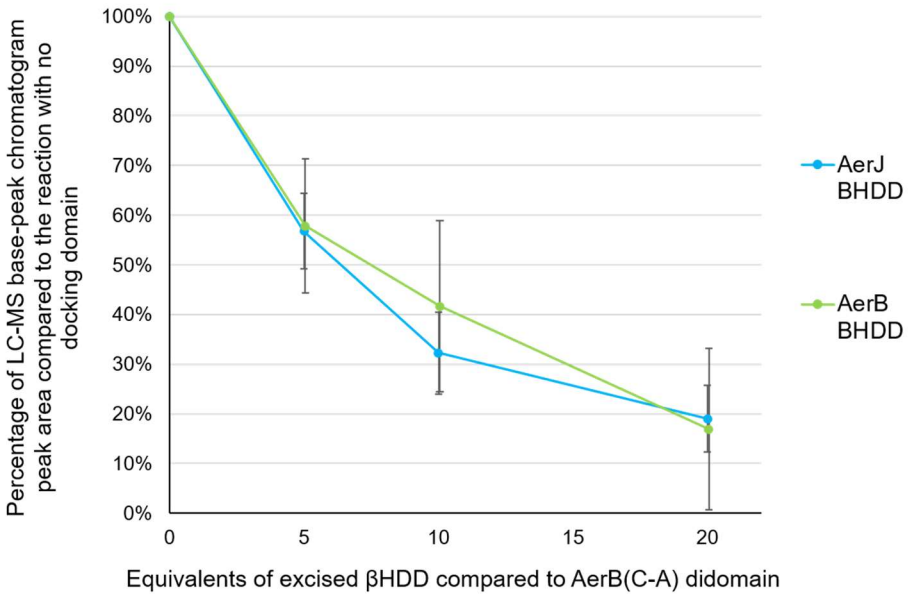


Figure 3.14 Plot of the percentage of peak area from UHPLC-ESI-IT-MS analysis of inhibition of AerB C-A didomain activity by addition of the excised AerJ or AerB βHDDs. Error bars show the mean ± standard deviation calculated from three independent experiments.

These results demonstrate that productive interaction of the AerA PCP domain and the AerB C-A didomain can be inhibited by co-incubation with either the AerJ or AerB βHDDs. This evidences the necessity of the βHDD-SLiM interaction in the productivity of the AerA PCP domain-AerB C domain interface. Furthermore, the extent of inhibition is not statistically significantly different with either βHDD, and suggests that they likely have a similar binding affinity with the AerA(PCP). This is perhaps not surprising, as both βHDDs are expected to interact with the AerA SLiM and share 58% amino acid sequence identity.

A control reaction in which the AerA PCP domain loading reaction was carried out in the absence of the carrier protein domain was performed. Following processing of the reaction and incubation with methylamine, no methyl amide product was observed (figure **3.13 B-C**), further demonstrating that the condensation domain was not accepting residual pantetheine- or CoA-thioesters in solution as the acyl donor.

In future experiments, it is of interest to observe whether excised docking domains can also inhibit halogenation of 4-HPLA-*S*-AerA(PCP) by AerJ. This

could be achieved through similar methodology to that described here, monitoring formation of mono- and di-chlorinated 4-HPLA methyl amide (**90**). It would also be useful to demonstrate whether the excised SLiM also inhibits these reactions. This could be accomplished by peptide synthesis of the AerA SLiM region and using this in similar experiments. This could give further confirmation of the necessity of SLiM-βHDD interactions for aeruginosin biosynthesis.

3.2. Biophysical studies of protein-protein interactions in aeruginosin biosynthesis

Following probing of protein-protein interactions by biochemical methods several questions remained. Inhibiting the interaction of the AerA PCP domain and AerB C-A didomain with excised β-hairpin docking domains demonstrated that SLiM-βHDD docking provides an important interaction interface between the two proteins. However, other experiments suggested mutagenesis of the AerA SLiM or AerJ βHDD does not significantly impact halogenation efficiency of AerJ towards AerA PCP domain-tethered substrates. Moreover, grafting the AerA SLiM onto heterologous carrier proteins did not improve recruitment of halogenase AerJ. Clearly, interactions between the globular parts of PCP domain and the surface of AerJ also play an important role in protein recognition. However their relative contributions to productive interface formation remains unclear.

To further probe the interactions involved in mediating the recruitment of AerJ and AerB by the AerA PCP domain, biophysical techniques were employed to begin to characterise the protein-protein interactions between these domains. This was particularly important for the interface between the AerA PCP domain and AerJ as binding between a carrier protein and a flavin-dependent halogenase has, to the best of our knowledge, not been biophysically characterised in previous studies.

3.2.1. Bio-Layer Interferometry measurements

Bio-Layer Interferometry measurements were conducted in collaboration with Dr. Simone Kosol, who performed streptavidin labelling of the AerA PCP domain and recorded data using a ForteBio Octet instrument.

Bio-Layer Interferometry (BLI) is a widely used technique to quantitate protein-protein affinity.[216] It measures the interference patterns of incident white light reflected by two surfaces at the terminus of a biosensor tip; a ForteBio proprietary internal reference surface and a biolayer consisting of immobilised ligands (figure **3.15 A**). Ligands are typically immobilised using high affinity biotin-streptavidin binding. Upon introduction to a solution containing a binding partner for the immobilised ligands, a shift in the wavelength of reflected light can be observed and measured (figure **3.15 B**). This can be used as a measure of the thickness of the biolayer, which is proportional to number of bound molecules and therefore the affinity between immobilised ligand and analyte. From this response, affinity constants can be calculated. This methodology was used to calculate affinities of the AerA PCP domain with AerJ and the AerB C domain as well as mutants lacking βHDDs to understand the contribution of docking domains to the strength of pairwise interactions.

Holo-AerA(PCP) was biotinylated by reaction with an *N*-hydroxysuccinimide (NHS) ester-containing derivative of biotin. NHS-biotin is selective for primary amines, so can react with any of the three lysine residues in the AerA PCP

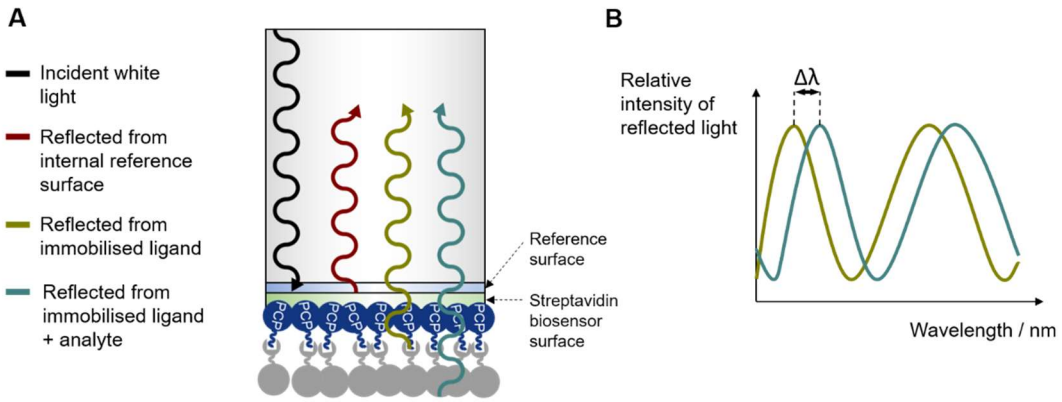


Figure 3.15 Measurement of binding of proteins by Bio-Layer Interferometry. (**A**) Representation of an BLI sensor tip with a steptavadin-tethered biolayer. (**B**) Idealised interference pattern resulting from BLI analysis. The wavelength shift (Δλ) upon binding of an analyte is recorded and used to determine specific binding.

domain amino acid sequence. Rigorous exchange into a phosphate buffer was necessary as the primary amine of trisaminomethane in the typical buffers used in this project can be biotinylated and provide undesirable auxiliary binding to the streptavidin sensor. Following desalting to remove excess NHS-biotin from solution, the biotinylated AerA PCP domain was loaded onto a streptavidin biosensor for analysis. Association and dissociation were recorded for AerJ, AerJΔβHDD, the AerB C domain and the AerB C domain ΔβHDD with the immobilised AerA PCP domain over a range of concentrations using a ForteBio Octet RED96 instrument (figure **3.16**). Measurements were conducted in triplicate (appendix 17). The AerB C domain ΔβHDD was produced by deletion the pET24a-(+)-AerB(C) plasmid (appendix 2) using the same boundary used to generate the excised AerB βHDD domain and was overproduced in *E. coli* BL21(DE3) and purified by nickel affinity chromatography.

The equilibrium response was extracted from these plots at each concentration and steady-state fitting was employed using a one-site binding model (figure **3.17**) to determine the equilibrium dissociation constant, K_D using the equation:

$$y = \frac{B_{max} \times x}{(K_D + x)}$$

where y = specific binding, B_{max} = maximum specific binding and x = concentration of the analyte (in this case, the concentration of AerJ, AerJΔβHDD, AerB C domain or AerB C domain ΔβHDD as appropriate).

Estimated K_D values are given in figure **3.17**. In the case of both AerJ and AerB C domain, removal of the β-hairpin docking domain increased the K_D, consistent with a reduced interaction efficiency. However, the K_D without the docking domain in both cases was still in the low μM range, suggesting that affinity without docking domains is still sufficient to maintain productive interaction interfaces. This is consistent with results of *in vitro* experiments, where differences between AerJ and constructs where the βHDD was mutated

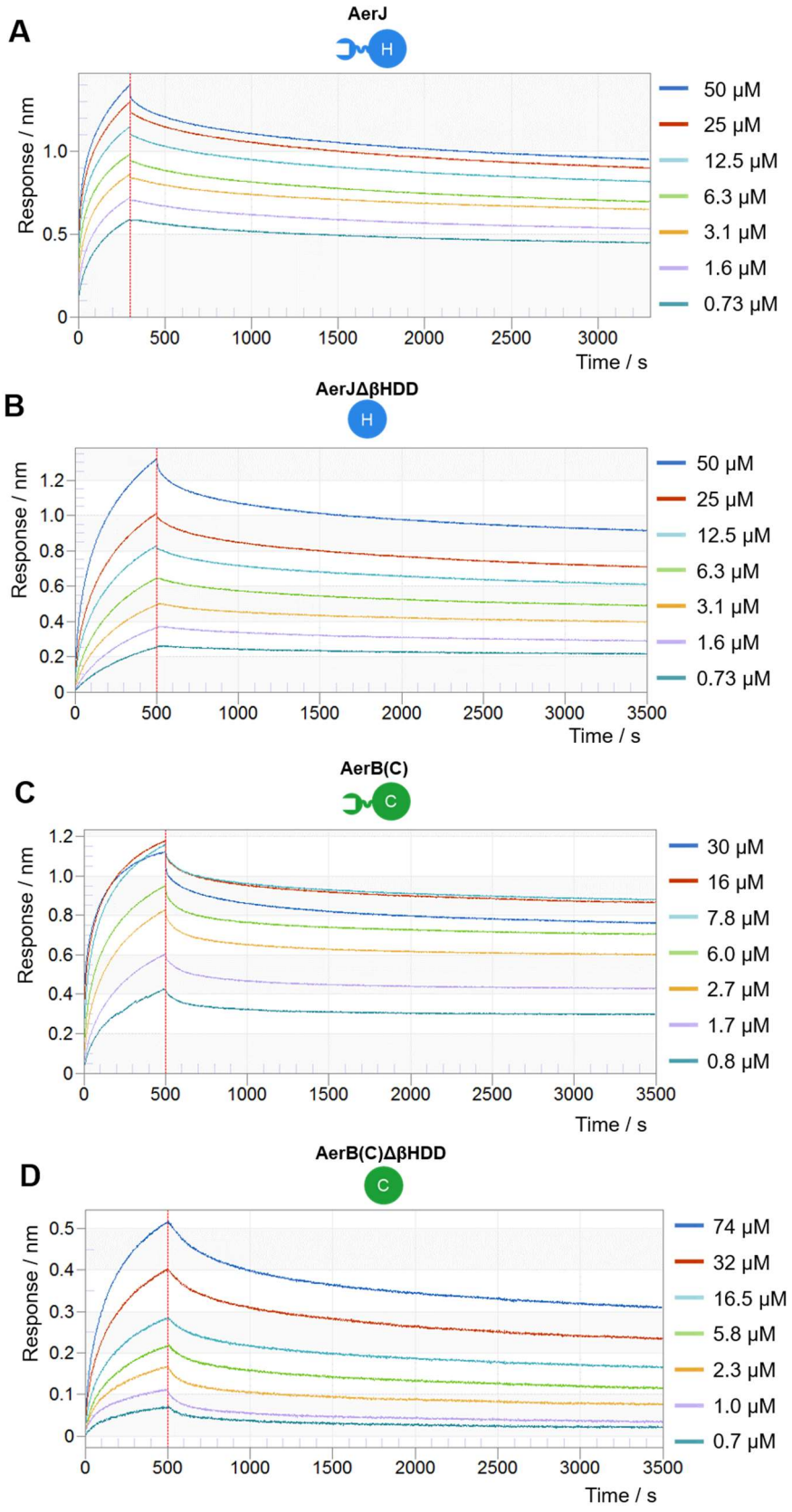


Figure 3.16 Association and dissociation curves over a range of concentration demonstrating interaction between streptavidin-immobilised AerA(PCP) and: (**A**) AerJ, (**B**) AerJΔβHDD, (**C**) the AerB C domain and (**D**) the AerB C domain ΔβHDD.

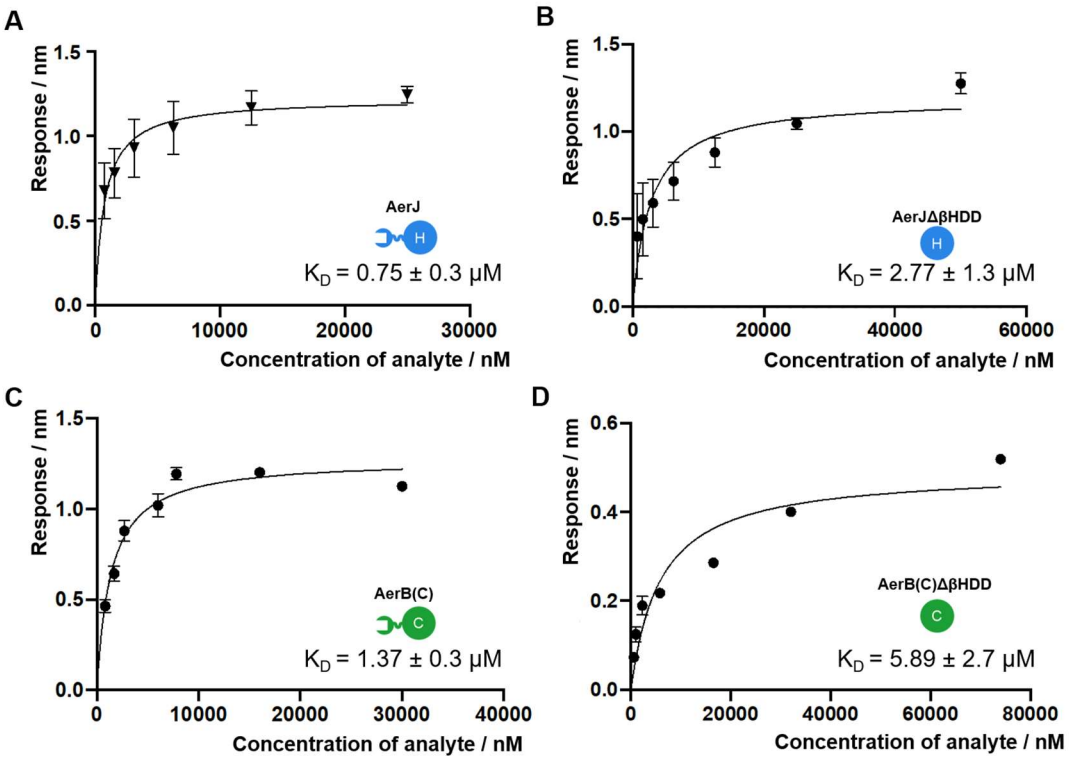


Figure 3.17 Steady-state fitting of BLI data using a one-site binding model for analytes (**A**) AerJ, (**B**) AerJΔβHDD, (**C**) the AerB C domain and (**D**) the AerB C domain ΔβHDD. Estimated K_D values calculated from the fittings are given with 95% confidence intervals. Error bars show the mean ± standard deviation from measurements in triplicate.

or removed proved difficult to observe. These results are consistent with the SLiM-βHDD interfaces being non-essential for protein-protein interaction but granting increased affinity to maintain selectivity or increase effective carrier protein concentration, which is becoming a prevailing hypothesis for the role of docking domains in natural product biosynthesis.[100,114,116,123] The K_D values for AerJ and the AerB C domain are very similar, with AerJ having only marginally stronger affinity, giving further evidence for their competition for AerA in aeruginosin biosynthesis.

In ongoing work, the binding affinities of the AerA PCP domain with the excised βHDDs from AerJ and AerB are to be determined by BLI measurements. This may give a more complete understanding of the contributions of both the βHDD and catalytic domains to the binding interfaces.

3.2.4. Carbene Footprinting experiments

Carbene footprinting experiments were conducted by Dr. Christopher Fage and Dr. Matthew Jenner.

Carbene footprinting is a recently developed technique that can localise the binding interfaces in protein-protein interactions.[217,218] The basis of carbene footprinting is labelling of solvent-exposed residues of the proteins using 4-(3-(trifluoromethyl)-3*H*-diazirin-3-yl)benzoic acid (**91**), which forms an aryl carbene (**92**) upon irradiation by a laser at a wavelength of 350 nm (figure **3.18 A**). The carbene is highly reactive and can insert into C-H, O-H, N-H or S-H bonds of the side chains of amino acid residues in proteins, allowing labelling of solvent-accessible residues. Subsequent mass spectrometry of protease digests of the protein allows detection of labelled peptides. The labelling of isolated proteins can be compared to the labelling of the proteins in complex, to elucidate their interaction interfaces (figure **3.18 B**). Suppression of labelling, or masking, in the complex suggests that residues contained within the tryptic peptide are involved in interface formation as they are less solvent-exposed. Enhanced labelling, or unmasking, can also be observed and implies

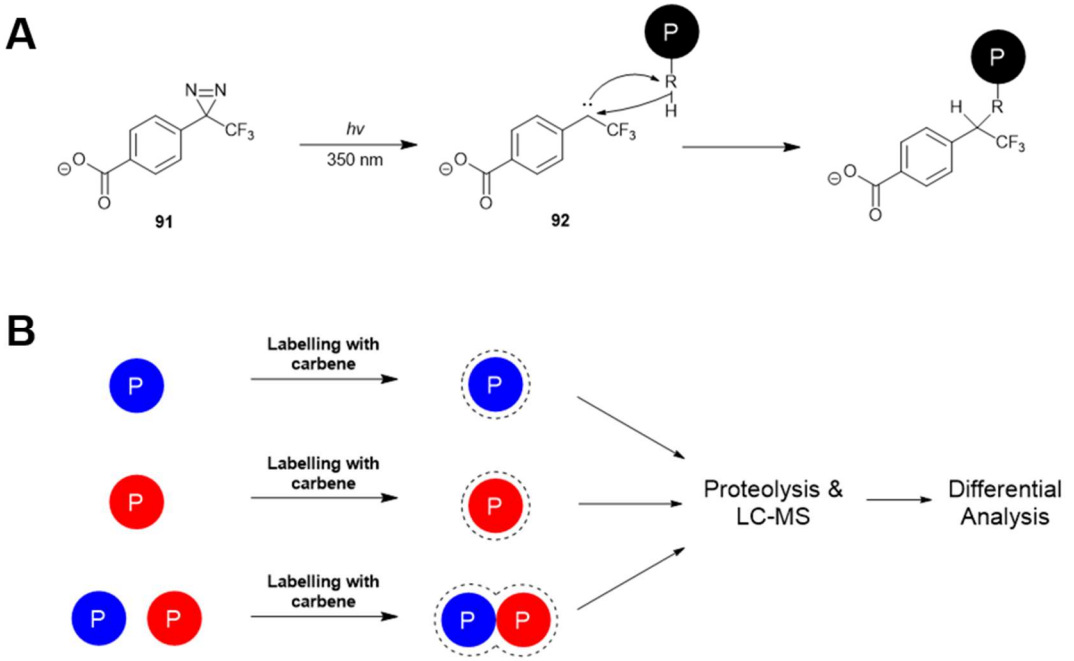


Figure 3.18 The principles of carbene footprinting to characterise protein-protein interactions. (**A**) Reaction scheme showing generation of carbene from irradiation of 4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (**91**). (**B**) Workflow for analysis of protein-protein interactions by carbene footprinting.

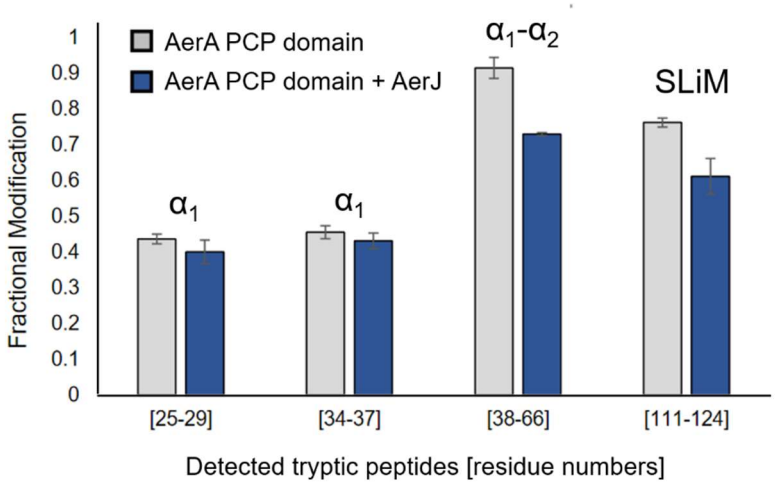


Figure 3.19 Fractional modification of tryptic peptides of AerA(PCP) by aryl diazirine **91** detected by UHPLC-ESI-Q-TOF-MS with and without AerJ in the reaction mixture. Error bars represent standard deviation from triplicate repeats. Graph provided by Dr. Matthew Jenner, University of Warwick.

that the residues are becoming more solvent-exposed, usually indicative of a conformational shift induced by binding of the partner protein.

Carbene footprinting was performed on the AerA PCP domain, AerJ and a mixture of the two by Dr. Christopher Fage and Dr. Matthew Jenner by mixing with **91**, flash-freezing and irradiating with a Nd:YLF laser. The protein mixtures were then, reduced, alkylated, digested with trypsin and analysed on a UHPLC-ESI-Q-TOF-MS instrument. Analysis of the resulting LC-MS traces was performed by Dr. Matthew Jenner, outputting fractional modification of each peptide fragment by diazirine **91**.

Four tryptic peptides of the AerA PCP domain were detected by UHPLC-ESI-Q-TOF-MS analysis. Fractional modification of the peptides in the absence and presence of AerJ was determined (figure **3.19**, appendix 18), uncovering no significant change in fractional modification of two peptides covering helix α_1, whilst statistically significant decreases in labelling were found for peptides covering the C-terminus of α_1 and part of α_2 as well as the C-terminal SLiM region. Figure **3.20** shows the masked regions of the AerA PCP domain displayed on a homology model of the AerA PCP domain generated by the Phyre2 webserver,[215] with modelling to the PCP domain from the TycC C-PCP

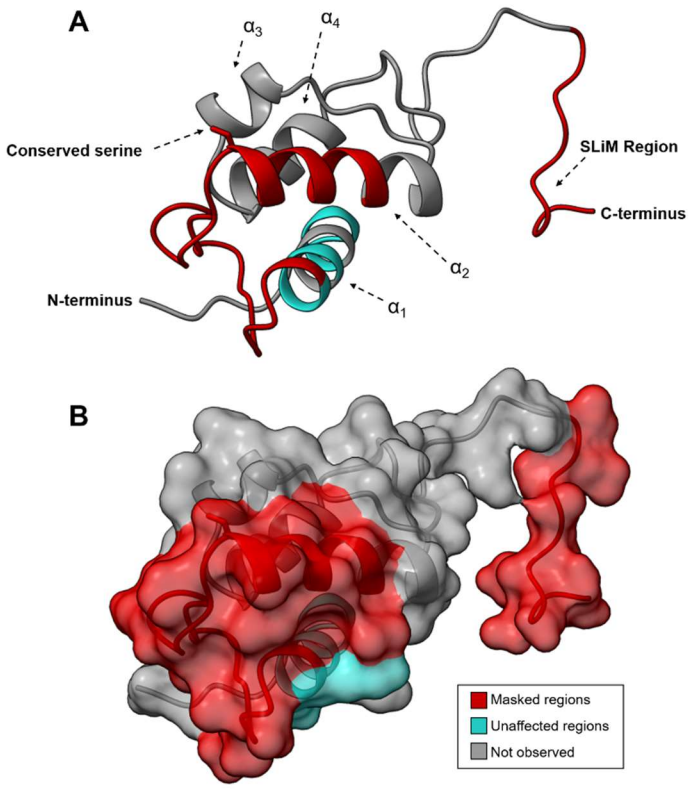


Figure 3.20 Homology model of the AerA PCP domain showing how regions of the protein that are masked during carbene footprinting when incubated with AerJ. (**A**) Cartoon depiction of the AerA PCP domain homology model. (**B**) Surface representation of the AerA PCP domain homology model. Both structures are shown in the same orientation. The homology model of the AerA PCP domain is the output of Phyre2 homology modelling with TycC PCP-C didomain crystal structure (PDB accession = 2JGP) as a template. Masked regions refer to those that produce statistically significant reductions in fractional modification by diazirine reagent **91** in the presence of AerJ.

didomain crystal structure (PDB accession = 2JGP).[219] The homology model was further refined by energy minimization using the ModRefiner webserver.[220] From masking of the residues at the C-terminus of the PCP domain, it is evident that the SLiM region is less solvent exposed in the presence of AerJ, evidencing its participation in the binding interface. Moreover, regions near the N-terminus of α-helix 2 and the linker between helices 1 and 2 show significant masking, consistent with formation of a binding interface with AerJ near to the conserved serine residue to which the phosphopantetheine arm would be appended in the *apo*-protein. This presumably allows the tethered substrate to access the halogenase active site. This is consistent with sequence alignment of the AerA PCP domain with other SLiM-bearing carrier protein domains in section 3.1.3, in which helix 2 of the carrier protein domains was found to be variable, potentially leading to

differences in halogenation efficiency by AerJ. It may be the case that interactions in this region are a primary determinant of the selectivity of AerJ towards the AerA PCP domain over other carrier proteins.

The binding mode for the AerA PCP domain in complex with AerJ suggested by carbene footprinting in this system is mostly in agreement to that of analogous experiments with the Bamb_5917 PCP domain and Bamb_5915 from enacyloxin biosynthesis.[123] In experiments with the Bamb_5917 PCP domain, both α-helices 1 and 2 were proposed to constitute a binding interface. This is consistent with the results described herein, although helix 1 appears to not be involved in the interaction of the AerA PCP domain and AerJ. Helices 3 and 4 of the Bamb_5917 PCP domain were also unmasked during interaction with Bamb_5915, but the corresponding peptides were not observed in tryptic digests of the AerA PCP domain. It may be of interest in future to repeat these experiments with the AerA PCP domain and AerB C domain. This would allow comparison of the PCP-C domain and PCP-halogenase binding modes and provide additional insights into carrier protein recruitment by different catalytic proteins.

Twenty-six tryptic peptides were detected for AerJ by UHPLC-ESI-Q-TOF-MS analysis, providing 47% sequence coverage (figure **3.21**, appendix 18). Mapping of the carbene footprinting data onto AerJ was performed using a homology model templated onto the crystal structure of MalA (PDB accession = 5WGY) provided by Phyre2[215] in the absence of a crystal structure of AerJ. MalA is a flavin-dependent halogenase from malbrancheamide biosynthesis in *Malbranchea aurantiaca* which acts upon tryptophan-like moieties,[97] (figure **3.22**) and was given the highest score for sequence and secondary structure alignment to AerJ by Phyre2. The excised AerJ βHDD had been crystalised and a structure resolved by Dr. Christopher Fage in parallel unpublished work within the Challis group, and this was used to map footprinting data onto the βHDD (figure **3.23**).

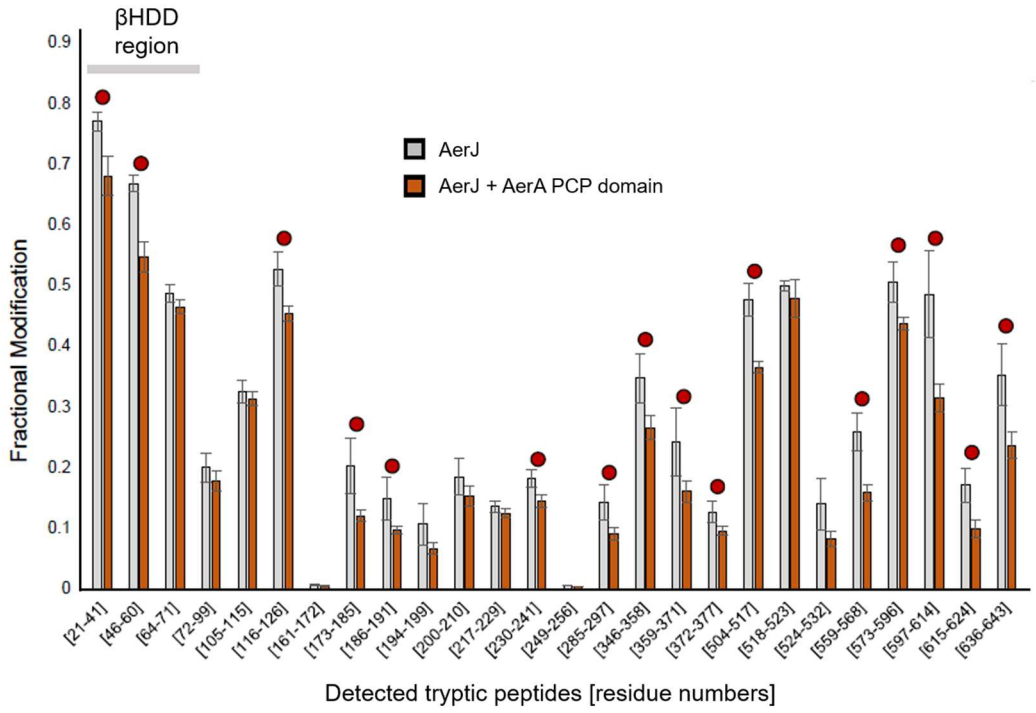


Figure 3.21 Fractional modification of tryptic peptides of AerJ by aryl diazirine **91** detected by UHPLC-ESI-Q-TOF-MS with and without the AerA PCP domain in the reaction mixture. Red circles indicate regions which were statistically significantly masked. Error bars represent the mean ± standard deviation from triplicate measurements. Graph provided by Dr. Matthew Jenner, University of Warwick.

As may be expected, several residues near to the active site channel of the halogenase are masked in the footprinting experiments, consistent with binding of the AerA PCP domain in an appropriate position to direct the tethered substrate towards the proposed catalytic lysine residue. However, there are masked residues on both faces of the active site tunnel, which are unlikely to both be related to binding of the AerA PCP domain. It is likely that the active site channel opposite to the $FADH_2$ binding site is where the PCP domain-tethered substrate binds, as this is where substrates have been found to be located in the active sites of the flavin-dependent halogenases PrnA, RebH and PltM.[91,180,211] It may be the case that one of these patches of masking is caused by positioning of the βHDD on the surface of the catalytic domain. Alternatively, this could be caused by global conformational changes

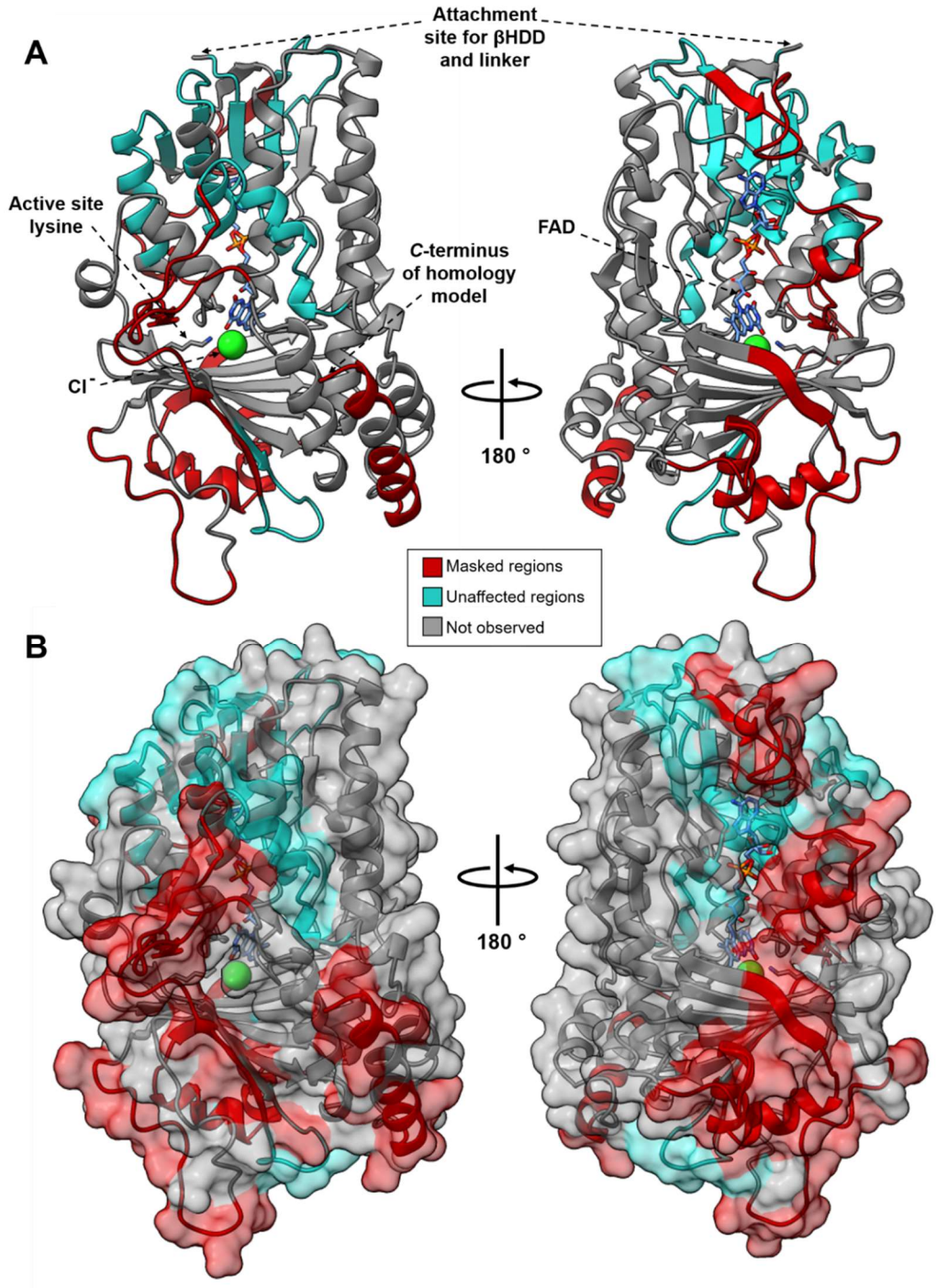


Figure 3.22 Homology model of AerJ showing how regions of the protein are affected during carbene footprinting when co-incubated with the AerA PCP domain. (**A**) Cartoon depiction of AerJ. (**B**) Surface representation of AerJ. Structures in (**A**) and (**B**) are shown in the same orientations. The homology model of AerJ is the output of Phyre2 homology modelling with the MalA crystal structure (PDB accession = 5WGY) as a template. Masked regions refer to those which produce statistically significant reduction in fractional modification by diazirine reagent **91** in the presence of AerJ.

of the active site channel upon binding of the carrier protein domain. Large patches of unaffected peptides were observed near to the binding site of FAD,

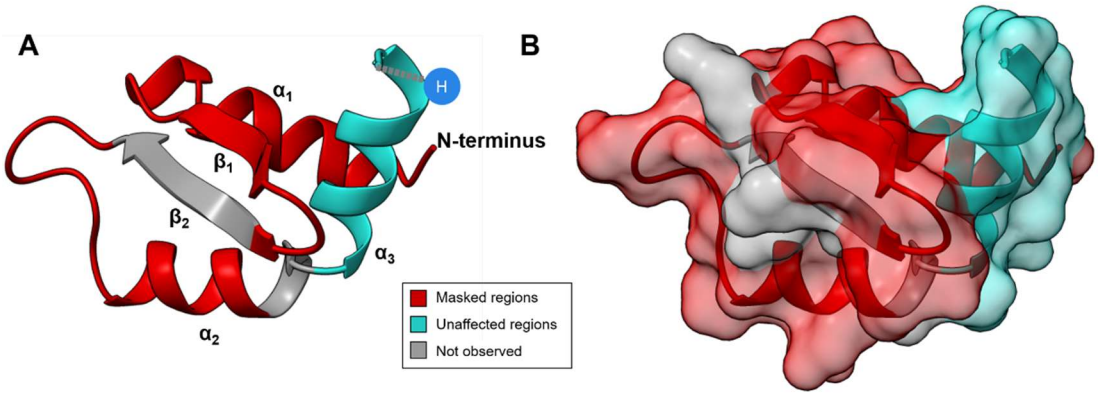


Figure 3.23 X-ray crystal structure of the excised AerJ β-hairpin docking domain showing how regions of the protein that are masked during carbene footprinting in the presence of the AerA PCP domain. (**A**) Cartoon depiction of the AerJ βHDD. (**B**) Surface representation of the AerJ βHDD. Structures in (**A**) and (**B**) are shown in the same orientations. Masked regions refer to those which produce statistically significant reduction in fractional modification by diazirine reagent **91** in the presence of the AerA PCP domain.

encompassing a large β-sheet and nearby helices, suggesting that the $FADH_2$ binding site is likely not affected by the interaction with the AerA PCP domain.

Considering the interactions of the AerJ βHDD, a digested peptide corresponding to the region across α-helix 2 and the linker between helix 2 and β-strand 2 was significantly masked (figure **3.23**). This gives further evidence, in addition to that of NMR titrations by Hacker *et al.*, that this groove is the site where the SLiM appended to the PCP domain binds.[126] The region encompassing α-helix 1 to β-strand 1 was also found to be masked. In previous studies, these have not been found to be involved in the interaction with the SLiM,[123,126] and instead may result from interaction of the βHDD with the surface of the catalytic domain of AerJ. Unfortunately, digested peptides coverage did not extend to β-strand 2 of the βHDD, but from previous studies it can be implied that this is involved in binding. Overall, these results are consistent with the βHDD of AerJ forming an interaction interface with the SLiM appended to the AerA PCP domain.

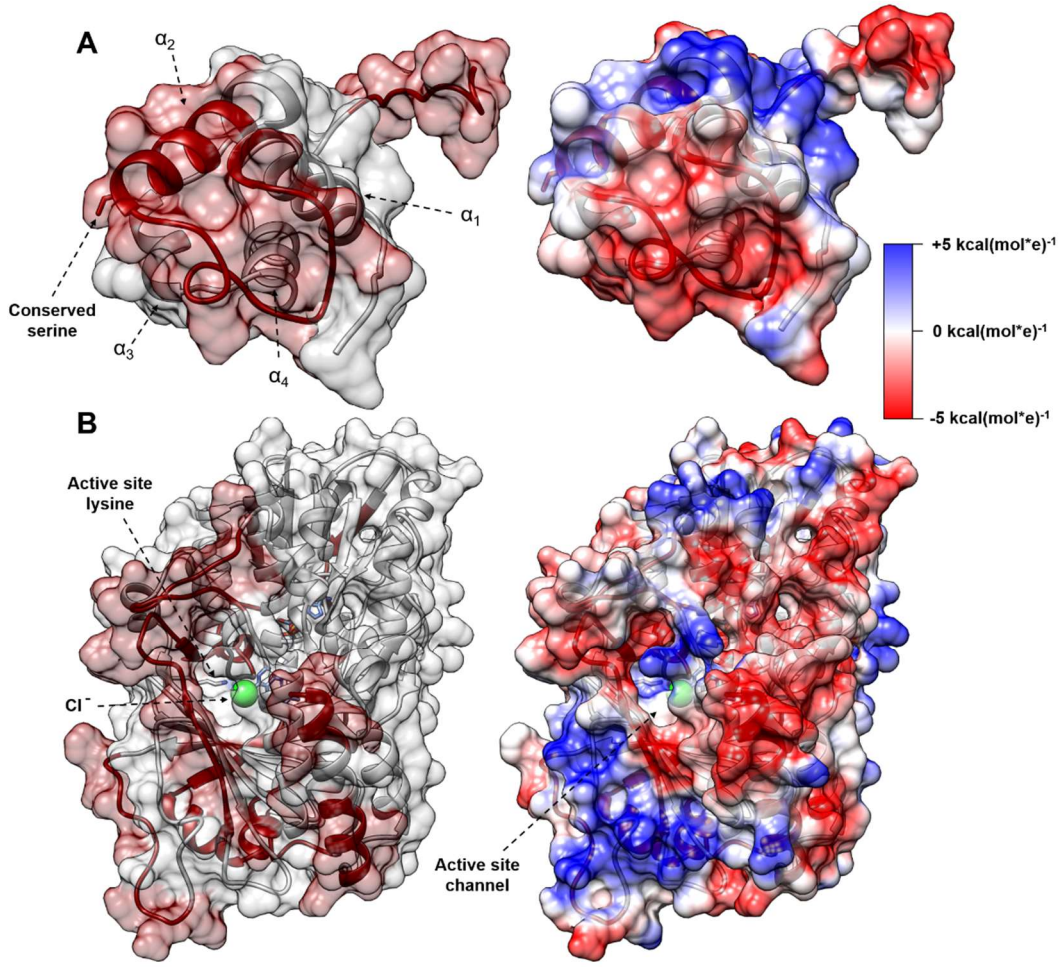


Figure 3.24 Structures comparing the regions masked in carbene footprinting experiments (*left*, with masked regions displayed in red) with the predicted coulombic surface potential calculated by UCSF Chimera (*right*) for: (**A**) a Phyre2 homology model of the AerA PCP domain; (**B**) a Phyre2 homology model of AerJ. Structures are displayed in the same orientation.

Coloumbic surface potentials were calculated for the AerA PCP domain and AerJ homology models using the program UCSF Chimera (figure **3.24**).[221] The masked regions of the AerA PCP domain were observed to mostly align with a negatively-charged region of the surface, with some positive charge on the surface of helix 2 (figure **3.24 A**). Meanwhile, regions of AerJ surface near the entrance to the active site channel of the halogenase are also masked, but comprise predicted patches of both negative and positive charge (figure **3.24 B**). It appears plausible that these two surfaces constitute part of the interaction interface between the two proteins, although further evidence of this interaction is required. Ongoing work in the Challis and Lewandowski groups is focused on investigating this binding model. In particular, some of the positively-charged regions near the active site channel are not covered by

detected tryptic peptides, which could provide crucial evidence for an interaction with the negatively-charged face of the PCP domain.

Unfortunately, there remains approximately 70 amino acids at the C-terminus of AerJ which are not covered by the homology model and do not map onto other flavin-dependent halogenase structures. Some residues in this region are masked when incubated with the AerA PCP domain, meaning they are likely implicated in interface formation. Optimisation of crystal trials for the full-length AerJ and homologues ApdC and McnD are ongoing. This could provide further insights into the structure and function of this C-terminal region. Additionally, coverage of the digested AerA PCP domain and AerJ was not complete, with large regions of the protein not observed in LC-MS analysis of tryptic digests. Repeating footprinting experiments with digestion of the proteins by proteases orthogonal to trypsin such as AspN or LysC, amongst others,[222] may provide access to further diagnostic peptides, allowing greater coverage and further localisation of the interaction interface between the AerA PCP domain and AerJ. This work is ongoing within the Challis group.

It was suggested that some of the masking of AerJ could be indicative of formation of a dimerisation interface. Although flavin-dependent halogenases are canonically monomeric, little is known about their structural dynamics. It is therefore plausible that binding of the AerA PCP domain may induce dimerisation. To test this hypothesis, the dimerisation state of AerJ in the presence of the AerA PCP was elucidated by analytical size-exclusion chromatography on a Superdex 200 Increase 10/30 GL column (figure **3.25**). The peak corresponding to AerJ had an elution volume corresponding to the monomer, with no peak for the dimeric state observed. Co-injection of the AerA PCP domain and AerJ in a 2:1 ratio as used for carbene footprinting experiments resulted in no shift in the AerJ peak, representative of no change in the oligomerisation state of the halogenase in the presence of the PCP domain. It is probable that the proteins were more dilute during gel filtration

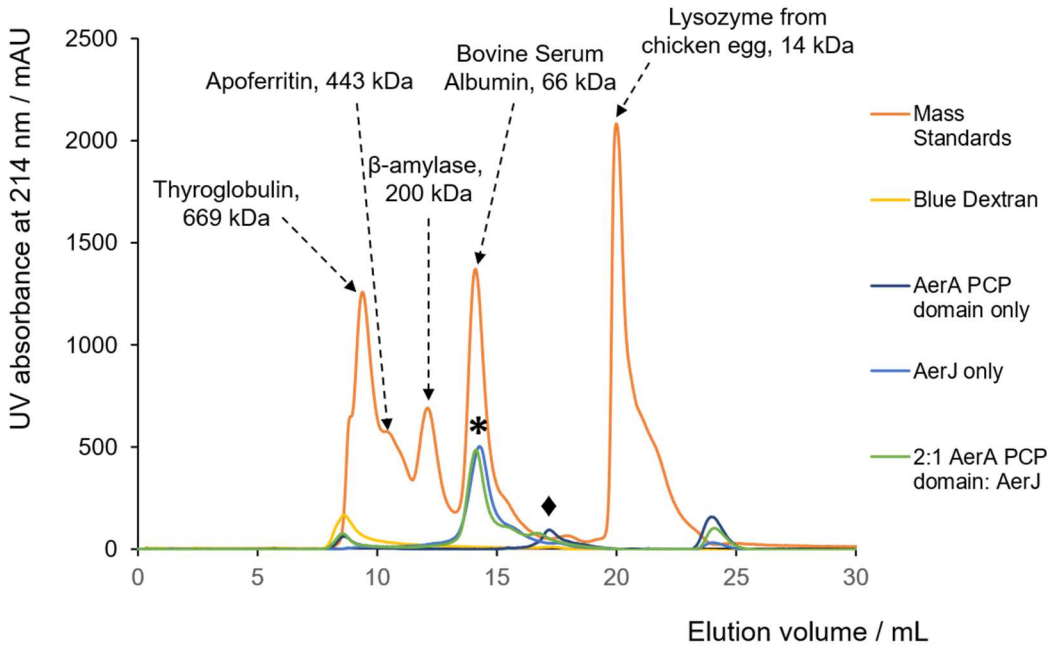


Figure 3.25 Analytical size-exclusion profile for AerJ, the AerA PCP domain and a 2:1 mixture of the AerA PCP domain:AerJ compared to standards of known mass. The asterisk (*) denotes the peak corresponding to AerJ (calculated = 74 kDa, observed = 87 kDa), and the diamond (♦) denotes the peak corresponding to the AerA PCP domain (calculated = 14 kDa, observed = 30 kDa).

than in the carbene footprinting reaction mixture, and higher concentrations may promote dimerization of the halogenase. However, current evidence suggests that masking of AerJ in carbene footprinting experiments is unlikely to be the result of a dimerisation interface. Interestingly, from these results the AerA PCP domain appears to exist as a dimer. However, this contradicts ongoing attempts to obtain a solution-state NMR structure of the AerA PCP domain by the Lewandowski group in unpublished work that shows the PCP domain to be monomeric.

3.3. Conclusions and future outlook

3.3.1. Summary

Several aspects of protein-protein interactions in aeruginosin biosynthesis have been explored. Experiments using the AerA PCP domain and the AerB PCP domain with 4-HPLA-*S*-pantheteine and 4-HPLA-Ile-*S*-panthetheine demonstrated that in aeruginosin biosynthesis both substrate selectivity and protein-protein interactions play a role in efficient halogenation by AerJ, but

that the nature of the carrier protein domain interaction partner appears to be the more important factor for AerJ activity.

Biochemical and biophysical experiments have been attempted to improve understanding of the SLiM-βHDD interfaces formed by AerJ and the AerB C domain with the AerA PCP domain. Inhibition of the AerB C domain's catalytic activity by excised AerJ or AerB βHDDs alongside K_D measurements and carbene footprinting clearly demonstrate that the AerA SLiM binds to the βHDDs. SLiM-βHDD binding has also been confirmed by biophysical techniques. Despite this, several experiments have suggested that removal or mutation of either the AerA SLiM or AerJ β-hairpin docking domain does not diminish catalytic activity of AerJ. Clearly, the role of SLiM-βHDD interfaces in aeruginosin biosynthesis is more complex than simply increasing catalytic efficiency of the βHDD domain-bearing domains.

Ongoing work in the Challis and Lewandowski groups is focused upon further carbene footprinting alongside solution state NMR titrations to further localise the interaction interface between the AerA PCP domain and AerJ, with the goal of producing a more complete model of binding. This information will guide the engineering of protein-protein interactions to develop AerJ and other βHDD-bearing catalytic proteins into biosynthetic engineering tools. Moreover, a crystal structure of AerJ is being sought to allow more detailed examination of the interaction of the AerA PCP domain-AerJ interface. This would also allow characterisation of the C-terminal subdomain of AerJ which produced masked regions in carbene footprinting experiments but is not observed in homology models.

3.3.2. Implications for biosynthetic engineering utilising AerJ

Flavin-dependent halogenase AerJ is a particularly promising tool for engineering of biosynthetic pathways, as carbon-halogen bonds can be utilised as chemical handles for further functionalisation by transition metal-catalysed cross coupling reactions.[223] Introduction of chlorines into natural products through introduction of AerJ by pathway engineering, therefore, represents an attractive strategy for accessing libraries of natural product

analogues which may harbour improved bioactivity or other drug-like properties. As discussed in section 3.1.2, AerJ can chlorinate L-Ile-4-HPLA-*S*-AerA(PCP), albeit with reduced efficiency compared to 4-HPLA-*S*-AerA(PCP), suggesting this halogenase is able to tolerate various phenolic substrates. Similarly, AerJ has been shown to halogenate substrates tethered to heterologous carrier proteins, albeit less efficiently than with the cognate partner. These properties are useful because they can be exploited to introduce the halogenase into other biosynthetic systems to engineer pathways. Synthesis and evaluation of halogenation of further pantetheine-tethered substrate analogues and additional progress in understanding the PCP domain-halogenase binding determinants will be critical for development of AerJ into a tool for pathway engineering applications.

3.3.3. The role of SLiM-βHDD interfaces in aeruginosin biosynthesis

The role played by SLiM-βHDD binding in the aeruginosin biosynthetic pathway remains ambiguous. However, one hypothesis that could be concluded from the data presented herein is that instead of increasing efficiency of chlorination of the AerA PCP domain-tethered substrate, the AerJ βHDD may be present to assist in inhibition of the interaction of the PCP domain with the AerB C domain by binding to the AerA SLiM. This would allow the halogenase to outcompete the AerB subunit for binding to the PCP domain, allowing halogenation of the PCP domain-tethered 4-HPLA group to occur before the AerB C domain can condense it, after which the formed 4-HPLA-L-Ile-*S*-AerB can no longer be chlorinated. This is consistent with the inhibition of the AerB C domain catalytic activity by the excised AerJ βHDD whilst also explaining why chlorination efficiency was not significantly affected by removal of the βHDD from the halogenase. The role of the AerB βHDD is likely to facilitate selectivity for enhanced pathway fidelity, similar to that concluded for previously investigated docking domains in megasynth(et)ases.

To investigate the validity of this hypothesis further, several experiments could be carried out. Using the halogenation-condensation cascade described in

section 2.3.2, in which 4-HPLA-*S*-AerA(PCP) is halogenated by AerJ and subsequently condensed with L-Ile-*S*-AerB(PCP) by the AerB C-A didomain, the halogenation efficiency of AerJΔβHDD mutant could be compared with the wild-type halogenase. A shift in chlorination pattern of the 4-HPLA-L-Ile-AerB(PCP) product to give overall less chlorinated species with the AerJΔβHDD mutant would be indicative of the AerJ βHDD playing a role in allowing the halogenase to outcompete the AerB C domain.

Additionally, binding kinetics of the of the PCP domain-halogenase interaction could be determined by BLI or related techniques such as surface plasmon resonance. Observing a higher on-rate (k_{on}) for interaction of the AerA PCP domain and the halogenase with wild-type AerJ than for the ΔβHDD mutant could suggest that the AerJ βHDD allows faster association of the proteins. This would further evidence the role of the AerJ βHDD in aiding competition with the AerB C domain.

Chapter 4: Utilising enzymatic crosstalk to generate novel natural product analogues

4.1. *In Vitro* Crosstalk

*Work in this section was conducted in collaboration with Challis group master's student Daniel Bahia, who synthesised 4-HPPA-S-pantetheine (**96**) under the supervision of the author.*

4.1.1. Crosstalk condensation of AerA(PCP)-tethered substrates with DHCCA analogues catalysed by Bamb_5915

As described in section 1.5.2.2.3, the standalone condensation domain Bamb_5915 from enacyloxin biosynthesis has high substrate tolerance, accepting a range of acyl donors and acceptors.[29] It has also been shown to act upon substrates tethered to non-native interaction partners that bear a conserved SLiM.[123] The former has proved invaluable in rational engineering of enacyloxin IIa production in *B. ambifaria* BCC0203 using mutagenesis and mutasynthesis approaches to access novel analogues.[224,225] However, the ability of Bamb_5915 to crosstalk with proteins from evolutionary divergent species has the potential to diversify assembly line engineering beyond the scope of the inherent chemistry of the producing biosynthetic pathway. As a model system to investigate both the metabolite engineering and biocatalytic potential of Bamb_5915, the productive crosstalk of Bamb_5915 with the AerA PCP domain was explored. Through this approach, we aimed to generate novel small molecule hybrids combining chemical moieties found in aeruginosin and enacyloxin IIa.

Challis group stocks of pET151 vector containing the *bamb_5915* gene from *B. ambifaria* AMMD were used to transform *E. coli* C43(DE3) cells and the protein was overproduced and purified by nickel-affinity chromatography according to literature procedures.[29,123] An acetyl unit was first used as the condensation acceptor, as this was known to be readily accepted by the Bamb_5915 C domain. Acetyl-CoA was loaded onto the carrier protein by incubation with Sfp and $MgCl_2$. Residual acetyl-CoA from the loading reaction was removed by dilution with reaction buffer and concentration through a 3 kDa molecular weight cut-off centrifugal concentrator with a minimum of one-

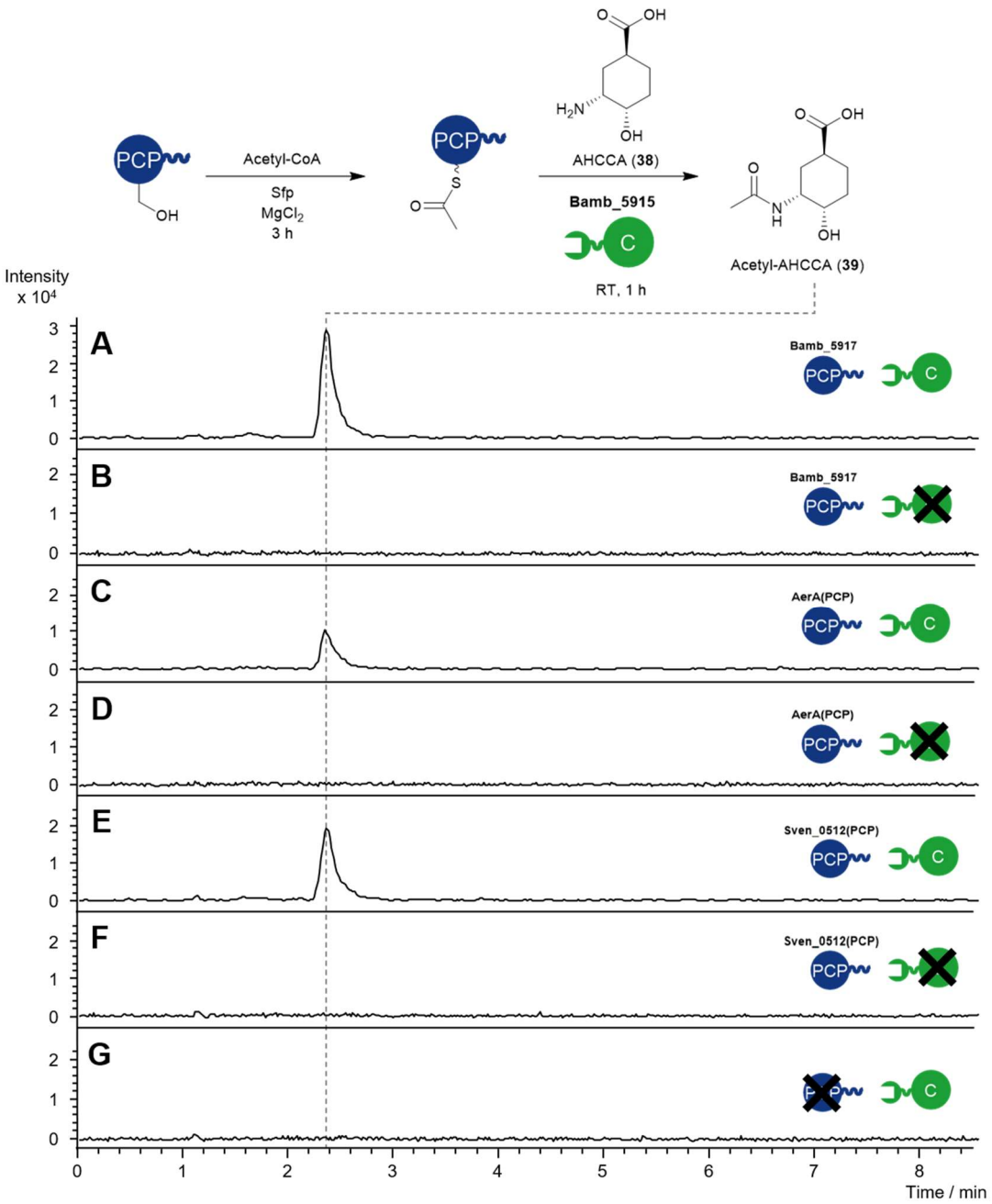


Figure 4.1 Extracted ion chromatograms from UHPLC-ESI-QTOF MS analysis following condensation of acetyl-*S*-PCPs and AHCCA (**38**) catalysed by Bamb_5915. (**A**) acetyl-*S*-Bamb_5917(PCP) with Bamb_5915. (**B**) acetyl-*S*-Bamb_5917(PCP) control without Bamb_5915. (**C**) acetyl-*S*-AerA(PCP) with Bamb_5915. (**D**) acetyl-*S*-AerA(PCP) control without Bamb_5915. (**E**) acetyl-*S*-Sven_0512(PCP) with Bamb_5915. (**F**) acetyl-*S*-Sven_0512(PCP) control without Bamb_5915. (**G**) a control reaction where the acetyl-CoA loading reaction was carried out in the absence of a carrier protein domain and then introduced to Bamb_5915 and AHCCA (**38**). In all cases, the m/z were calibrated to a sodium formate internal standard and the extracted ion chromatograms shown are the sum of *m/z* = 202.1079 and 224.0899 ± 0.02 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of **39** respectively.

hundred reaction volumes. Acetyl-AerA(PCP) was then incubated with Bamb_5915 and an amino analogue of the natural acyl acceptor substrate, AHCCA (**38**), previously synthesised and utilised as a condensation donor in the Challis group.[29,123] Production of acetyl-AHCCA (**39**) was monitored by

UHPLC-ESI-Q-TOF-MS analysis. In parallel, the same assay was conducted with the native interaction partner, the Bamb_5917 PCP domain (figure **4.1 A-B**). The assay results provided preliminary evidence that the AerA PCP domain can crosstalk with Bamb_5915 (figure **4.1 C-D**). However, the amount of **39** formed using the AerA PCP domain was approximately three times lower than that with the Bamb_5917 PCP domain. This was calculated by comparison of the ion intensities of **39** detected in the extracted ion chromatogram. This was consistent with the results observed for crosstalk between Bamb_5915 and Sven_0512(PCP),[123] which were also repeated here (figure **4.1 E-F**).

Assays in which Bamb_5915 was replaced with reaction buffer did not produce detectable amounts of **39**, indicating that conversion to acetyl-AHCCA was not a result of non-enzymatically catalysed nucleophilic attack of the pantetheinyl-thioester by AHCCA (**38**). Similarly, a reaction where acetyl-CoA loading was conducted without a PCP domain present also showed no product formation (figure **4.1 G**). This confirms that Bamb_5915 cannot directly accept residual acetyl-CoA as a condensation acceptor.

We then tested whether Bamb_5915 can accept PCP domain-tethered 4-HPLA as an acyl donor substrate. Initial trials of condensation of 4-HPLA-*S*-Bamb_5917(PCP) and 4-HPLA-*S*-AerA(PCP) with AHCCA (**38**), catalysed by Bamb_5915, showed formation of the expected 4-HPLA-AHCCA (**93**) by UHPLC-ESI-Q-TOF-MS analysis (figure **4.2 A-D**). An authentic standard of **93** was not synthesised, but the molecular formula of the expected product was confirmed by high-resolution mass spectrometry (figure **4.3**). Because racemic 4-HPLA-*S*-pantetheine was loaded onto the carrier protein, a mixture of diastereomeric products was observed. However, the ion intensity of the peaks corresponding to the product were very low. This is likely due to Bamb_5915 not readily accepting α-branched substrates. This is perhaps expected as initial *in vitro* studies of Bamb_5915 demonstrated it was unable to accept α-branched isobutyryl-, pivaloyl- and serinyl-*S*-NAC substrates as acyl donors.[29]

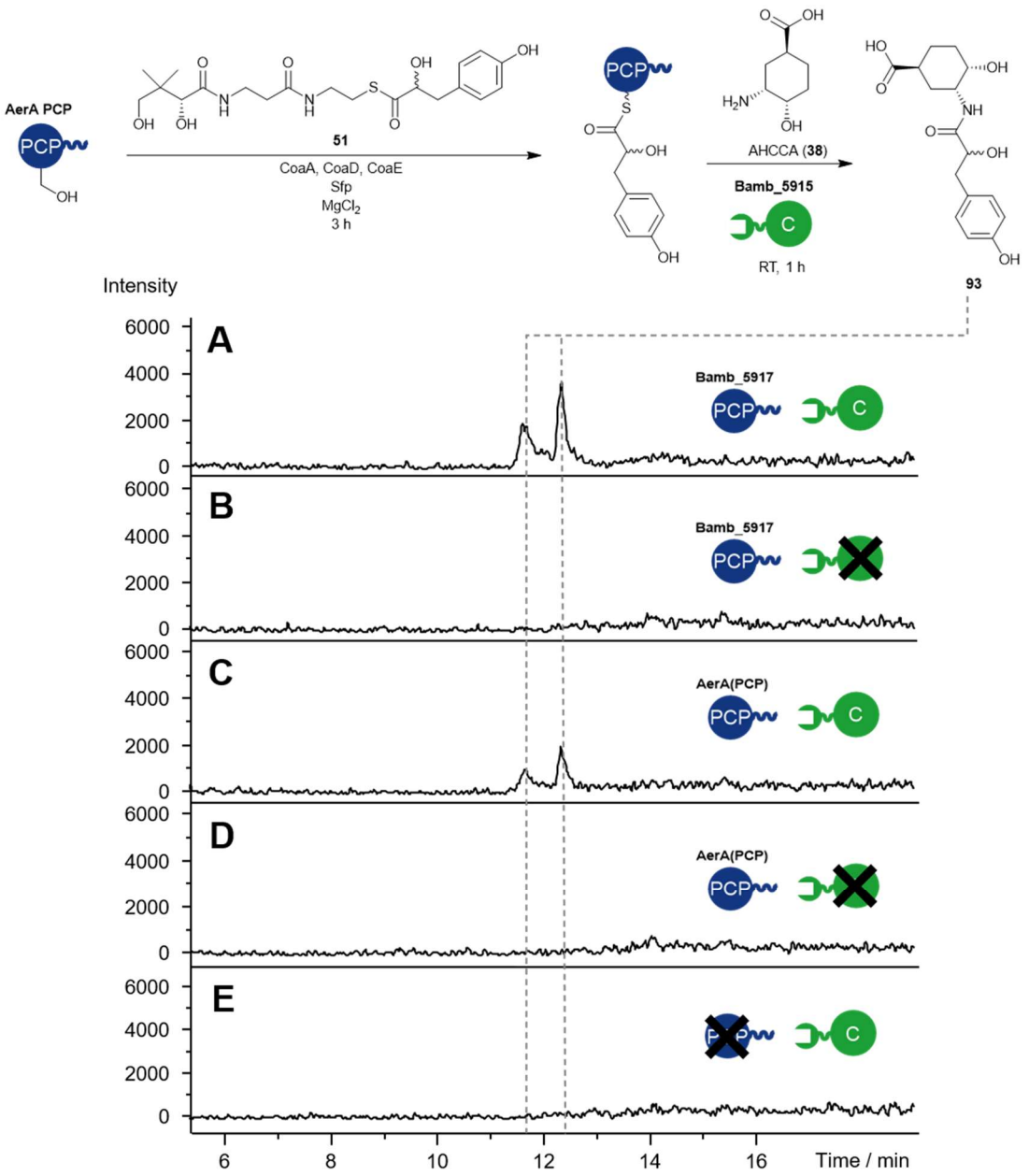


Figure 4.2 Extracted ion chromatograms from UHPLC-ESI-QTOF MS analysis of the reaction of 4-HPLA-*S*-PCPs and AHCCA (**38**) catalysed by standalone C domain Bamb_5915. (**A**) 4-HPLA-*S*-Bamb_5917(PCP) with Bamb_5915; (**B**) 4-HPLA-*S*-Bamb_5917(PCP) control without Bamb_5915; (**C**) 4-HPLA-*S*-AerA(PCP) with Bamb_5915; (**D**) 4-HPLA-*S*-AerA(PCP) control without Bamb_5915; (**E**) a control reaction where the 4-HPLA-*S*-pantetheine loading reaction was carried out in the absence of any carrier protein and then introduced to Bamb_5915 and AHCCA. In all cases, masses were calibrated to a sodium formate internal standard and the extracted ion chromatograms shown are the sum of *m/z* = 324.1442 and 346.1261± 0.01 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of **93** respectively.

Again, the reaction with the Bamb_5917 PCP domain produced more of the expected product than the reaction with the AerA PCP domain. This is likely to be because Bamb_5915 has a higher affinity for its native interaction partner than the non-native AerA PCP domain.

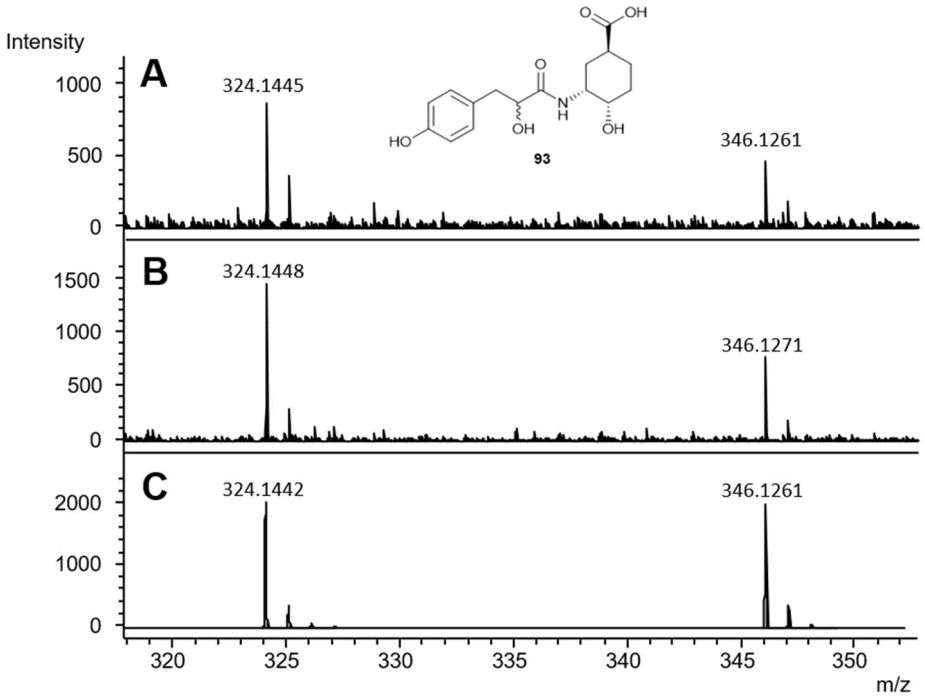
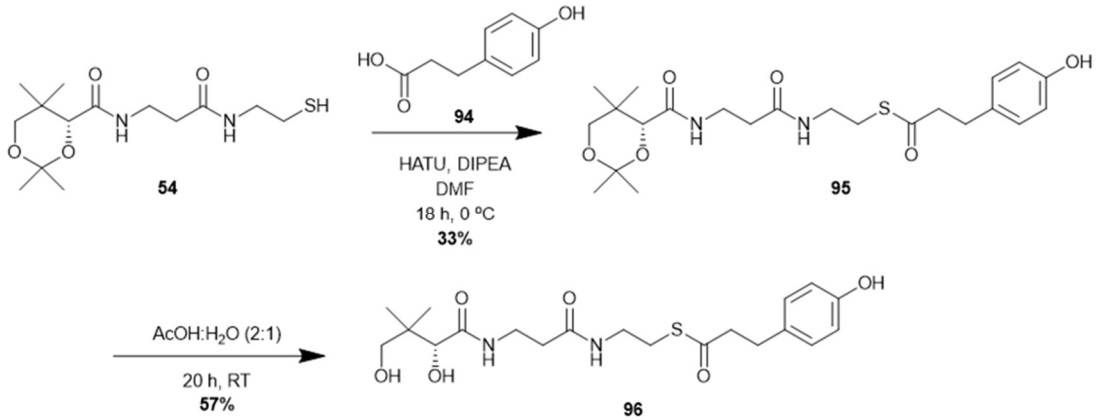


Figure 4.3 Mass spectra for the diastereomeric products produced by condensation of 4-HPLA-*S*-PCPs and AHCCA (**38**) catalysed by Bamb_5915. Mass spectra show $[M+H]^+$ and $[M+Na]^+$ masses from: (**A**) the first eluted diastereomer of **93** from UHPLC chromatography (11.6 mins); (**B**) the second eluted diastereomer of **93** from UHPLC chromatography (12.4 mins); (**C**) a simulated mass spectra of **93** with a peak width of 0.1 Da and ion intensity of 2000 A.U.

The poor acceptance of 4-HPLA-*S*-PCP domains by Bamb_5915 and the generation of two diastereomeric products, which if halogenation by AerJ was introduced could yield a complex mixture of up to six different molecules, meant that this was not a useful substrate for further study. A substrate analogue lacking the α-hydroxyl group was therefore investigated as this should both be more readily accepted by Bamb_5915 and would not produce diastereomeric products. A synthetic pantetheine with a 3-(4-hydroxyphenyl)propionic acid (4-HPPA, **94**) moiety appended as a thioester



Scheme 4.1 Synthesis of 4-HPPA-*S*-pantetheine (**96**).

was synthesised *via* a similar preparation described for 4-HPLA-*S*-pantetheine (**51**) (scheme **4.1**). The thioester linkage was synthesised by coupling of acetonide-protected pantothenic acid (**54**) and commercially available 3-(4-hydroxyphenyl)propionic acid (**94**) with amide coupling reagent HATU. Deprotection of acetonide **95** by stirring in acetic acid and water lead to formation of the expected pantetheine thioester **96**.

Initially, the reactions were trialled with AHCCA (**38**), as the synthesis of an authentic standard of an ester of the natural DHCCA (**37**) substrate was envisioned to be challenging. Loading of 4-HPPA-*S*-pantetheine (**96**) onto the Bamb_5917 PCP domain or AerA PCP domain and subsequent incubation with AHCCA (**38**) and Bamb_5915 lead to appearance of a product peak in the extracted ion chromatograms for the $[M+H]^+$ and $[M+Na]^+$ ions corresponding to 4-HPPA-AHCCA (**97**) following UHPLC-ESI-QTOF-MS analysis (figure **4.4**). Ion intensities of the peaks corresponding to **97** were approximately 3 times higher than the combined intensities of the two epimers obtained with 4-HPLA, suggesting that, as expected, the substrate lacking α-branching was more readily accepted by Bamb_5915. Interestingly, in the two enzymatic reactions a second peak with similar retention time (approximately 12.2 mins) and approximately 5% of the area of the product peak was detected. This had the same calculated molecular formula as **97** ($[M+H]^+$ found = 308.1488, calculated = 308.1498), however the identity of this compound remains unclear. Unlike the major peak in which both the protonated and sodiated adducts are observed, only an *m/z* value corresponding to the $[M+H]^+$ ion is observed for this secondary peak. This peak could correspond to a diastereomer of the product, resulting from epimerisation of one of the stereocentres under the assay conditions or a regioisomer resulting from nucleophilic attack of the thioester by the hydroxyl group of AHCCA (**38**). Because the product is not observed in controls without Bamb_5915, the former seems more likely.

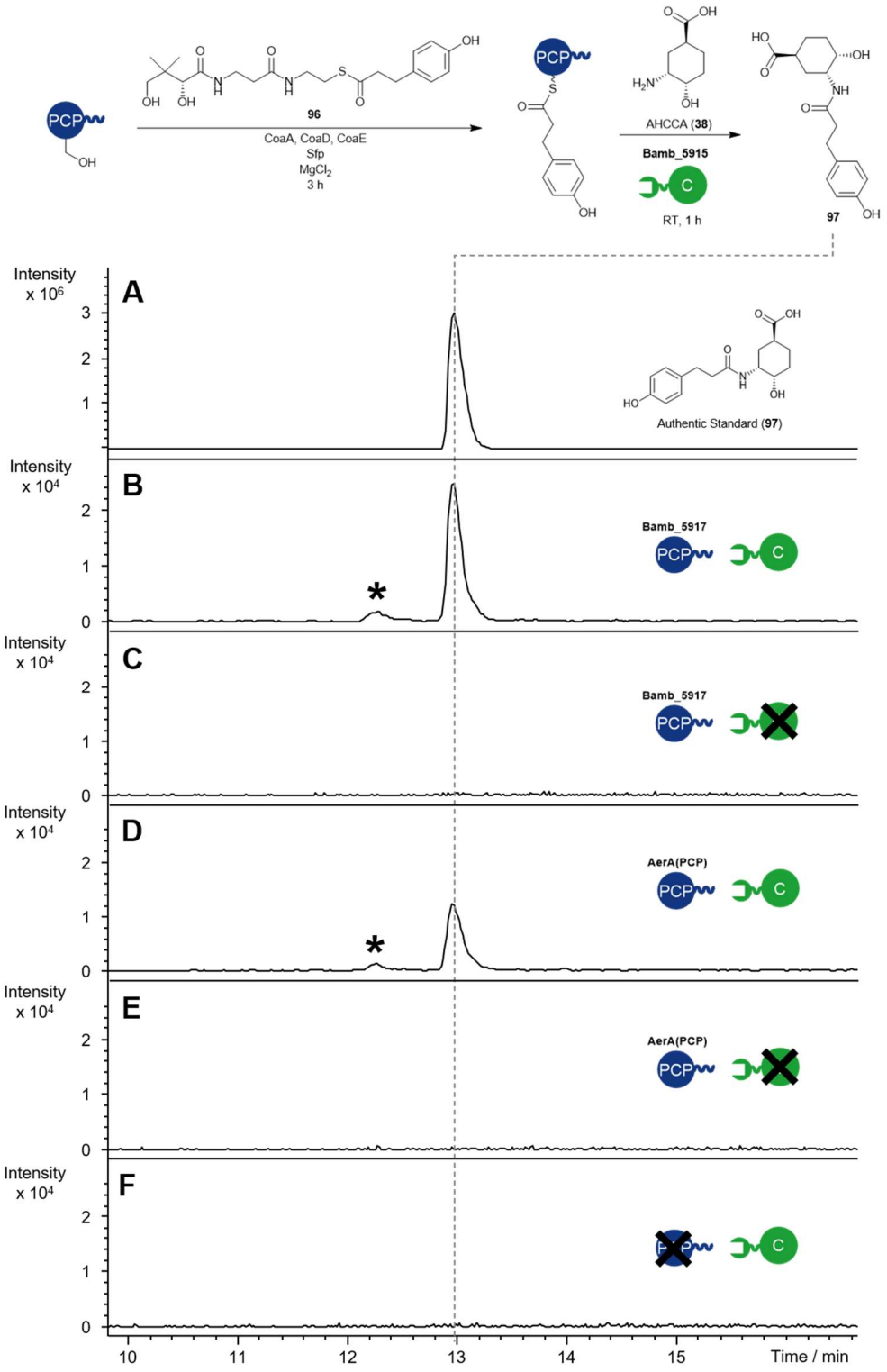
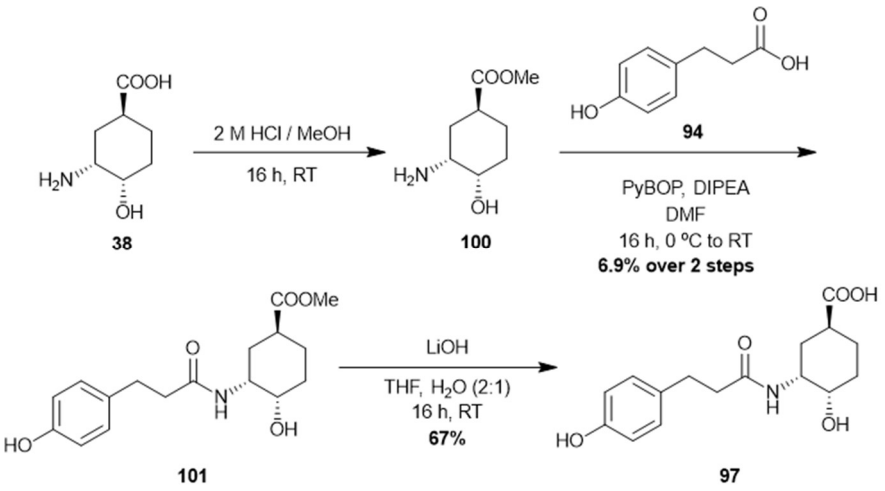


Figure 4.4 Extracted ion chromatograms from UHPLC-ESI-QTOF MS analysis of reaction of 4-HPPA-*S*-PCPs and AHCCA (**38**) catalysed by Bamb_5915. (**A**) a synthetic standard of 4-HPPA-AHHCA (**97**); (**B**) 4-HPPA-*S*-Bamb_5917(PCP) with Bamb_5915; (**C**) 4-HPPA-*S*-Bamb_5917(PCP) control without Bamb_5915; (**D**) 4-HPPA-*S*-AerA(PCP) with Bamb_5915; (**E**) 4-HPPA-*S*-AerA(PCP) control without Bamb_5915; (**F**) a control reaction where the 4-HPLA-*S*-pantetheine loading reaction was carried out in the absence of any carrier protein and then introduced to Bamb_5915 and AHCCA (**38**). In all cases, the masses were calibrated to a sodium formate internal standard and the extracted ion chromatograms shown are the sum of *m/z* = 308.1498 and 330.1317 ± 0.005 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions respectively. Secondary peaks observed in the extracted ion chromatograms are highlighted with an asterisk (*).



Scheme 4.2 Synthesis of an authentic standard of 4-HPPA-AHCCA (**97**).

An authentic standard of **97** was synthesised by protecting AHCCA (**38**), previously synthesised in the Challis group,[29] as the corresponding methyl ester (**100**) by reaction with thionyl chloride in methanol. Coupling of **100** to 4-HPPA (**94**) with PyBOP yielded **101.** Hydrolysis of the methyl ester **101** with lithium hydroxide afforded the authentic standard of **97** (scheme **4.2**). Due to the hydrophilicity of **97**, purification by semi-preparative HPLC was required. Comparison of the retention times of the authentic standard and the product of the *in vitro* condensation reaction confirmed the enzymatic synthesis of 4-HPPA-AHCCA (**97**) (figure **4.4 A**).

Whilst it has been shown that feeding of AHCCA (**38**) into a *B. ambifaria* mutant strain lacking the genes responsible for DHCCA (**37**) biosynthesis led to incorporation of AHCCA (**38**) into enacyloxin IIa (**35**) analogues,[224] to simplify engineering of the enacyloxin IIa pathway *in vivo*, using the natural condensation donor DHCCA (**37**) would be preferable. Repeating the reactions described for AHCCA with synthetic DHCCA (**37**), originally synthesised in the Challis group,[29,123] lead to observation of the 4-HPPA-DHCCA (**102**) ester product by UHPLC-ESI-Q-TOF-MS (figure **4.5**). A synthetic standard was not pursued due to the difficulty of discriminating between the two hydroxyl groups of DHCCA (**37**) for formation of the ester bond. However, taking these results together with the AHCCA experiments, it is highly likely that the expected product was formed. Only one peak was

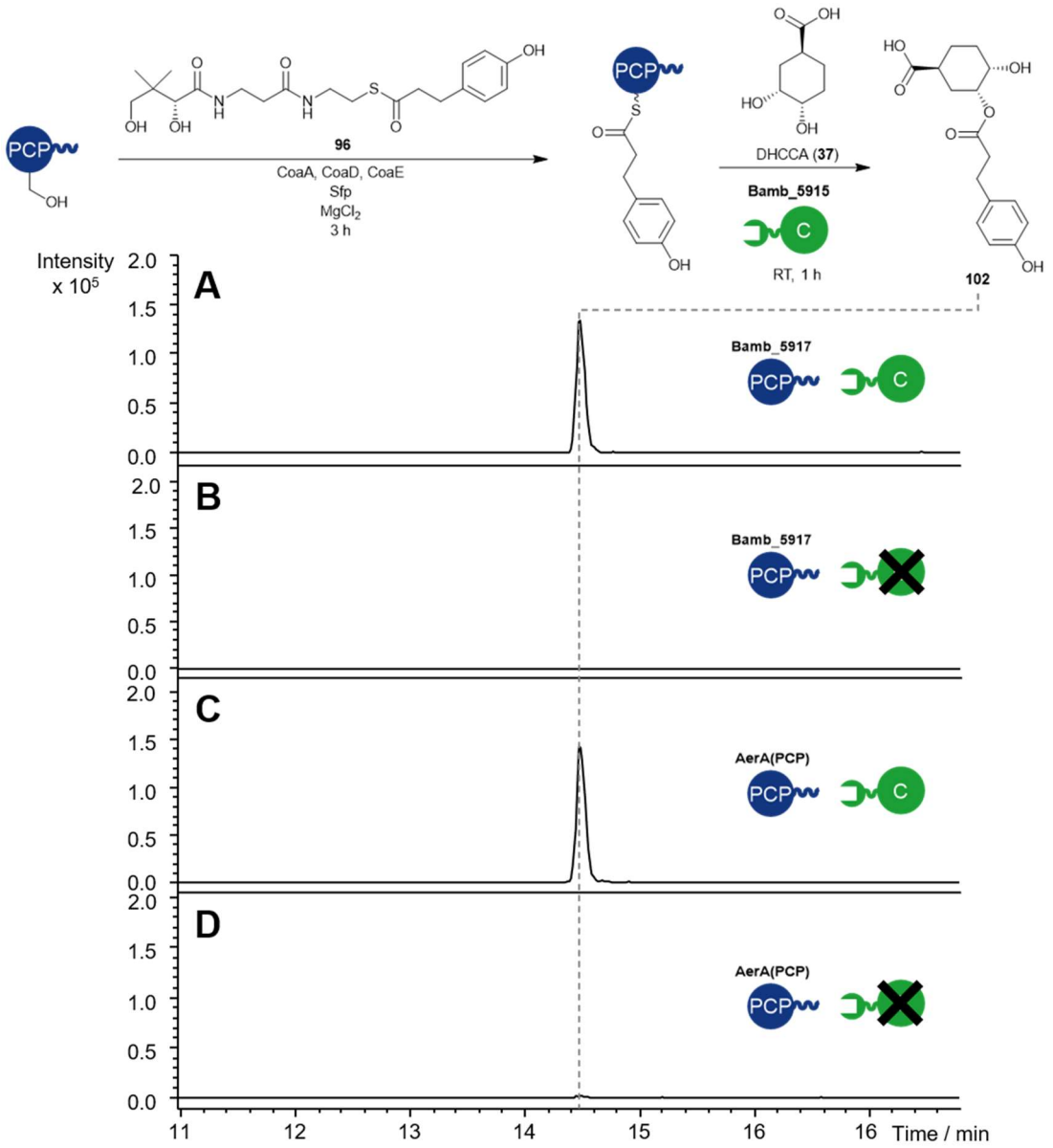


Figure 4.5 Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analysis of reaction 4-HPPA-*S*-PCP domains and DHCCA (**38**) catalysed by Bamb_5915. (**A**) 4-HPPA-*S*-Bamb_5917(PCP) with Bamb_5915. (**B**) 4-HPPA-*S*-Bamb_5917(PCP) control without Bamb_5915. (**C**) 4-HPPA-*S*-AerA(PCP) with Bamb_5915. (**D**) 4-HPPA-*S*-AerA(PCP) control without Bamb_5915. In all cases, the *m/z* values were calibrated to a sodium formate internal standard and the extracted ion chromatograms shown are the sum of *m/z* = 309.1338 and 331.1158 ± 0.01 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ ions of **102** respectively.

observed in the extracted ion chromatogram, suggesting only a single regiochemistry of the product was formed.

4.1.2. Sequential halogenation by AerJ and condensation by Bamb_5915

The extent of halogenation of 4-HPPA-*S*-AerA(PCP) by AerJ was first monitored by intact-protein mass spectrometry as described in chapter 2 (figure **4.6**). Comparison of the peak heights of the three species showed

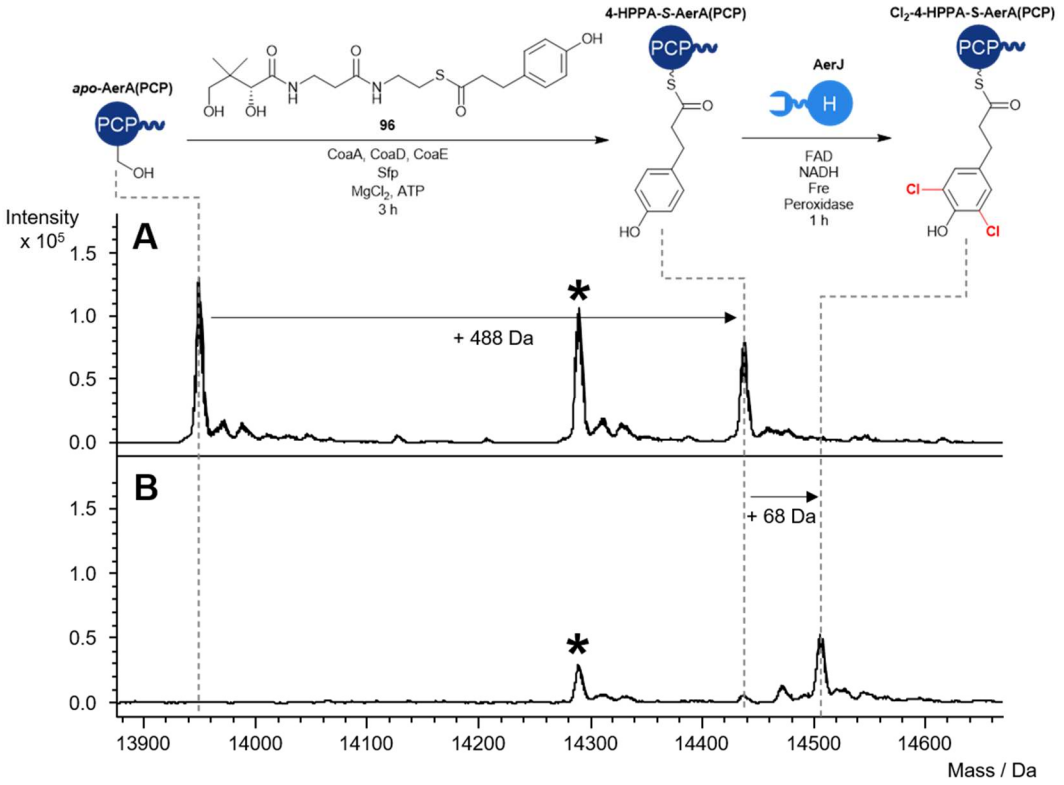


Figure 4.6 UHPLC-ESI-Q-TOF analysis of halogenation of 4-HPPA-*S*-AerA(PCP) by AerJ. Charge state-deconvoluted mass spectra show: (**A**) Product of chemoenzymatic conversion of *apo*-AerA(PCP) to 4-HPPA-*S*-AerA(PCP). (**B**) Product of halogenation with AerJ and relevant cofactors. Asterisks (*) indicate the presence of *holo*-AerA(PCP) resulting from hydrolysis of the thioester linkage.

>75% conversion to the di-chlorinated product in one hour, approximately consistent with the degree of chlorination observed with 4-HPLA-*S*-AerA(PCP). This suggested 4-HPPA would be an acceptable substitute for 4-HPLA to generate a chemoenzymatic cascade with both halogenation by AerJ and condensation by Bamb_5915.

To investigate competition between Bamb_5915 and AerJ for the PCP domain-tethered substrates, Bamb_5915 and AerJ were introduced to 4-HPPA-*S*-Bamb_5917(PCP) or 4-HPPA-*S*-AerA(PCP), along with the appropriate cofactors for the halogenase and AHCCA (**38**). Following simultaneous incubation for 1 h and analysis by UHPLC-ESI-Q-TOF-MS, 4-HPPA-*S*-Bamb_5917(PCP) produced approximately 78% non-chlorinated **97** and 22% mono-chlorinated product **98** (figure **4.7 C**) whilst 4-HPPA-*S*-AerA(PCP) produced approximately 6% non-chlorinated, 20% mono-chlorinated and 74% di-chlorinated product **99** (figure **4.7 F**). This can be rationalised by considering that the Bamb_5917 PCP domain

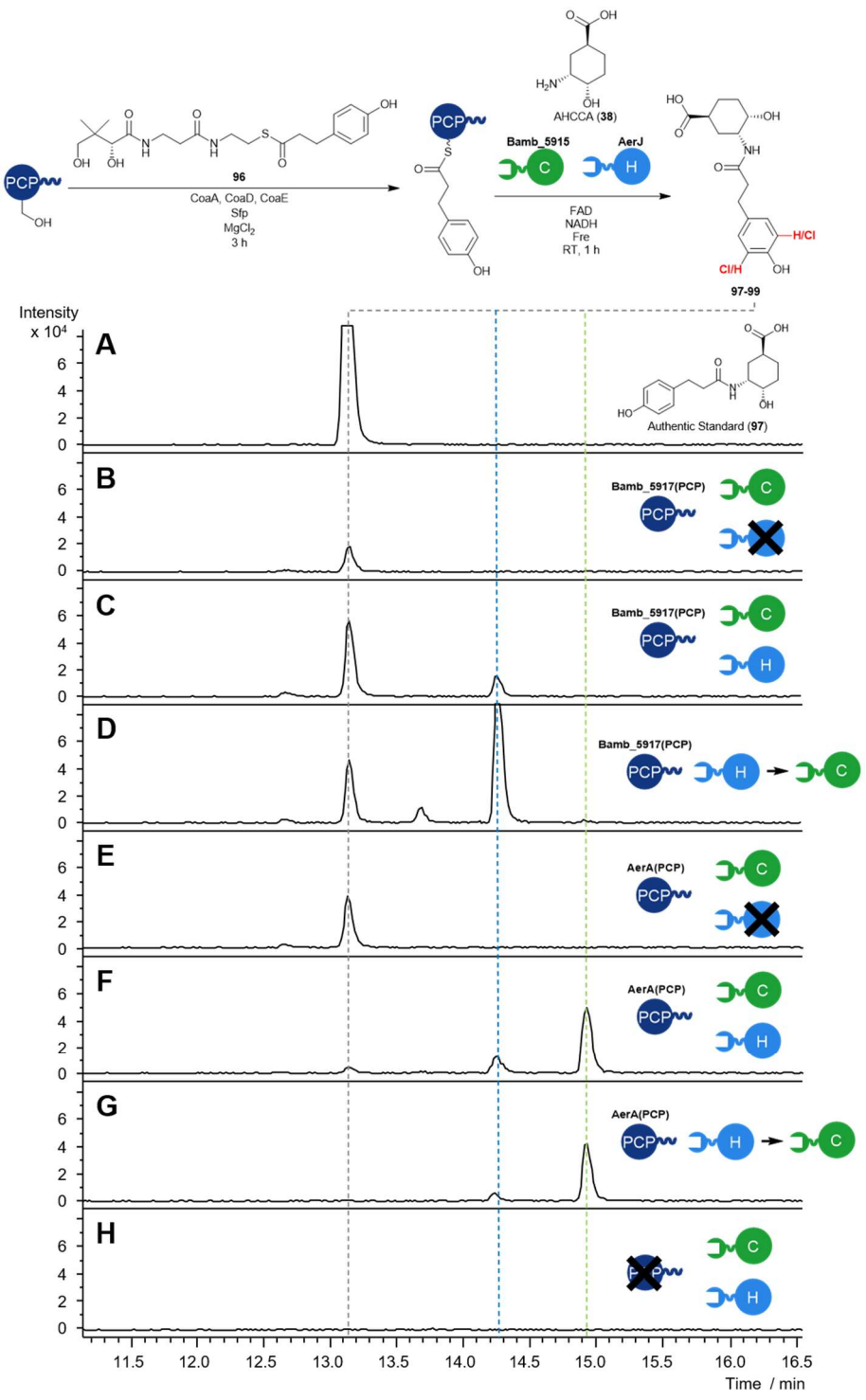


Figure 4.7 Extracted ion chromatograms from UHPLC-ESI-QTOF MS analysis of reaction of 4HPPA-*S*-PCPs with AerJ and subsequent condensation with AHCCA (**38**) catalysed by Bamb_5915. (**A**) Non-chlorinated authentic standard (**97**). (**B**) 4-HPPA-*S*-Bamb_5917(PCP) control with no halogenase. (**C**) Simultaneous reaction of 4-HPPA-*S*-Bamb_5917(PCP) with AerJ and Bamb_5915. (**D**) Sequential reaction of 4-HPPA-*S*-Bamb_5917(PCP) with AerJ and Bamb_5915. (**E**) 4-HPPA-*S*-AerA(PCP) control with no halogenase. (**F**) Simultaneous reaction of 4-HPPA-*S*-AerA(PCP) with AerJ and Bamb_5915. (**G**) Sequential reaction of 4-HPPA-*S*-Bamb_5917(PCP) with AerJ and Bamb_5915. (**H**) A control reaction where the 4-HPPA-*S*-pantetheine loading reaction was carried out in the absence of any carrier protein and then introduced to AerJ and cofactors, Bamb_5915 and AHCCA. In all cases, the *m/z* values were calibrated to a sodium formate standard and the extracted ion chromatograms shown are the sum of *m/z* = 308.1498, 330.1317, 342.1108, 364.0928, 376.0719 and 398.0538 ± 0.02 Da, corresponding to the $[M+H]^+$ and $[M+Na]^+$ of the non-, mono- and di-chlorinated molecular ions respectively. Blue dashed line = mono-chlorinated product **98**, green dashed line = di-chlorinated product **99**.

interacts more productively with its native interaction partner Bamb_5915 than non-native AerJ, whilst the opposite is likely true for the AerA PCP domain. Thus, Bamb_5915 likely outcompetes AerJ for interaction with 4-HPPA-*S*-Bamb_5917(PCP), leading to less turnover by the halogenase. In the case of 4-HPPA-*S*-AerA(PCP), the halogenase likely outcompetes the condensation domain, leading to near complete conversion to the di-chlorinated species.

In an attempt to increase the degree of chlorination observed, the halogenation and condensation reactions were also performed sequentially, with halogenase and cofactors introduced and incubated for one hour before introduction of Bamb_5915 and AHCCA (**38**). The product profile of the reaction with 4-HPPA-*S*-Bamb_5917(PCP) shifted dramatically, with now roughly 30% non-chlorinated **97** and 70% mono-chlorinated product **98** observed (figure **4.7 D**). The product profile for the parallel reaction with the AerA PCP domain did not change significantly, as the initial enzymatic cascade demonstrated near complete conversion to di-chlorinated product **99**. Nevertheless, sequential reactions did lead to even higher conversion (approximately 11% mono-chlorinated **98** and 89% di-chlorinated **99**), with no detectable non-chlorinated product remaining (figure **4.7 G**). These results together suggest that careful choice of PCP domain, reaction times and stoichiometries of proteins could allow exclusive access to the mono- and di-chlorinated products in future experiments.

It was clear from analysis of the mass spectra that chlorinated species had been formed due to the diagnostic chlorine isotopic patterns. For the UHPLC peaks corresponding to mono-chlorinated 4-HPPA-AHCCA (**98**), an approximate 3:1 ratio of the molecular ion and a +2 Da *m/z*, corresponding to species containing a ^{35}Cl and ^{37}Cl atom at natural abundances respectively, was observed for both the $[M+H]^+$ and $[M+Na]^+$ adducts (figure **4.8 A-B**). For the peaks corresponding to di-chlorinated 4-HPPA-AHCCA, a 9:6:1 ratio of the molecular ion, +2 Da and +4 Da m/z were observed, corresponding to $^{37}Cl^{37}Cl$,

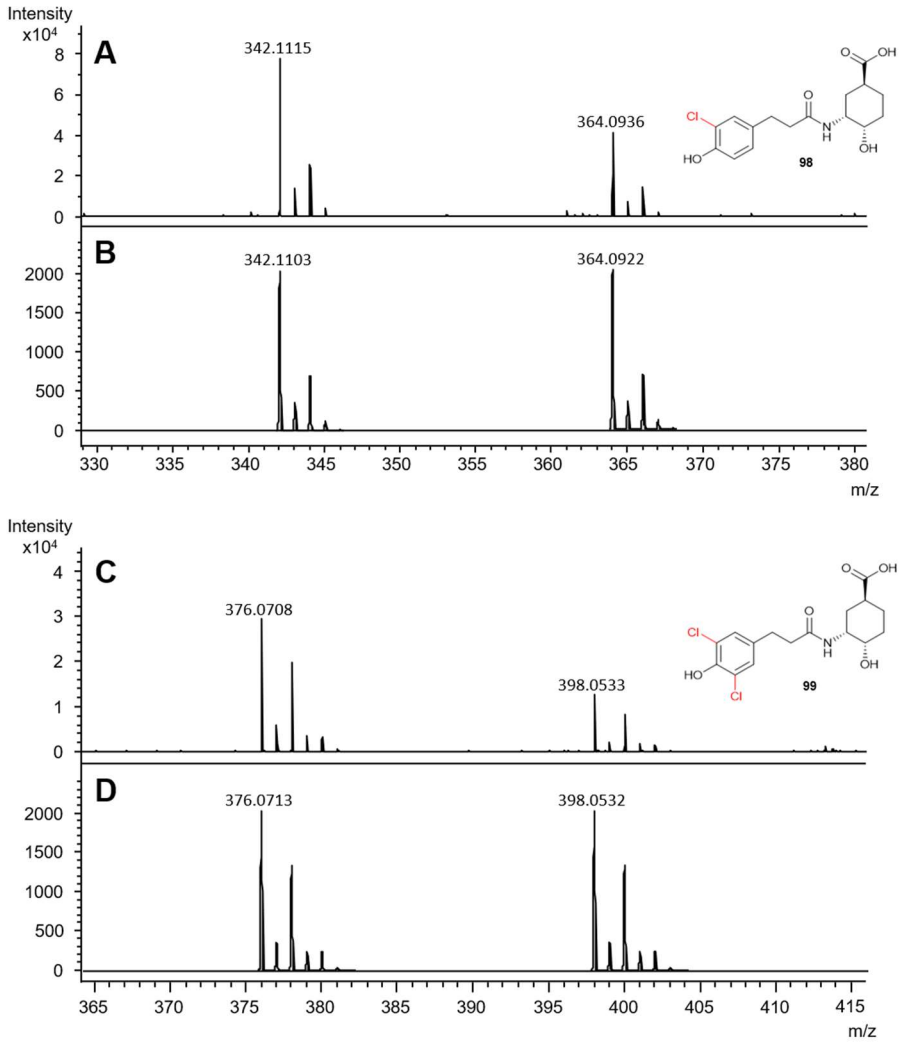


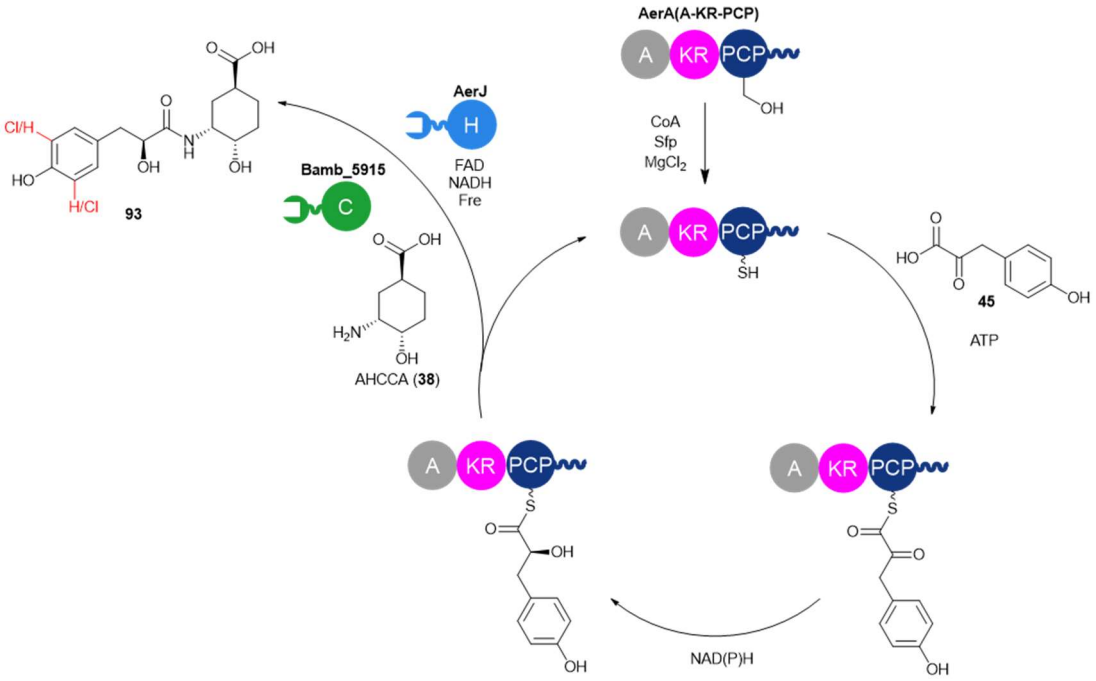
Figure 4.8 Mass spectra of chlorinated 4-HPPA-AHCCA (**97**) products from tandem halogenation and condensation of 4HPPA-*S*-PCPs by AerJ and Bamb_5915. Mass spectra show $[M+H]^+$ and $[M+Na]^+$ masses from: (**A**) UHPLC-ESI-Q-TOF analysis of mono-chlorinated product **98**; (**B**) a simulated mass spectra of mono-chlorinated product **98** with a peak width of 0.1 Da and ion intensity of 2000 A.U.; (**C**) UHPLC-ESI-Q-TOF analysis of di-chlorinated product **99**; (**D**) a simulated mass spectra of di-chlorinated product **99** with a peak width of 0.1 Da and ion intensity of 2000 A.U.

$^{37}Cl^{35}Cl$ and $^{35}Cl^{35}Cl$ containing species at natural chlorine isotopic abundances (figure **4.8 C-D**). Whilst authentic standards would be desirable for confirming the regiochemistry of chlorination by the halogenase, comparison with the non-chlorinated standard **97** and the distinct chlorination patterns observed give strong evidence that simultaneous halogenation and condensation by AerJ and Bamb_5915 was viable for generation of these novel natural product hybrids.

4.1.3. Attempted multiple turnover using AerA(A-KR-PCP)

Having established that AerJ and Bamb_5915 could act in tandem towards SLiM-bearing carrier protein domains to generate novel natural product hybrids, larger titres were pursued. Accessing milligrams of compounds would allow structural elucidation of the product by NMR to confirm the chlorination regiochemistry, assess biological activity and allow further chemical functionalisation.

However, the maximum amount of product produced by the approach described in section 4.1.2. was fundamentally limited. The chemoenzymatic synthesis of acyl-*S*-PCPs only allows a single turnover as following condensation, the PCP is left in the *holo*-form, from which the acyl-*S*-PCP cannot be regenerated unless non-enzymatically catalysed transthioesterification of the acyl-*S*-pantetheine occurs. As a consequence, to scale up production of acyl-AHCCA derivatives, a large amount of acyl-*S*-PCP domain would need to be generated as a starting material. This would likely require significant optimisation of the reaction conditions and would be limited by protein stability and yields from protein purification.



Scheme 4.3 Proposed chemoenzymatic synthesis of chlorinated 4-HPLA-AHCCA analogues (**97-99**) using the full-length AerA(A-KR-PCP) subunit.

To circumvent these limitations the full AerA(A-KR-PCP) subunit was targeted for overproduction. Using the adenylation and ketoreductase domains to load and reduce the 4-hydroxyphenylpyruvic acid (**45**) with cofactors NADPH and ATP respectively, would allow multiple turnovers (scheme **4.3**). Although the acyl donor would be 4-HPLA, which has been shown to be poorly accepted as a substrate of Bamb_5915, some condensation was observed in experiments described in section 4.1.1. Therefore, this was thought to be an acceptable model system to observe product formation with a view to changing the acyl donor by exploring or engineering the substrate scope of the AerA A domain in future experiments.

A gene encoding AerA(A-KR-PCP) for cloning into pET24a-(+) was designed using the deposited sequence (GenBank accession: ACM68682.1) with a His_8-tag and TEV cleavage site. The gene was synthesised and cloned NdeI to XhoI into pET24a-(+) by Epoch Life Sciences. Overproduction in BL21(DE3) cells and sequential purification by Ni-affinity and size-exclusion chromatography yielded the desired protein. SDS-PAGE and mass spectrometry analysis demonstrated the identity of this protein (figure **4.9**). An impurity of approximately 80 kDa was observed in the SDS-PAGE gel (figure **4.9 A**). Attempts were made to determine the identity of this impurity by UHPLC-ESI-Q-TOF-MS analysis but were unsuccessful. This could be determined in future work by peptide mass fingerprinting, where the SDS-PAGE gel band is proteolytically digested and the resulting peptides are analysed by MALDI-TOF mass spectrometry.

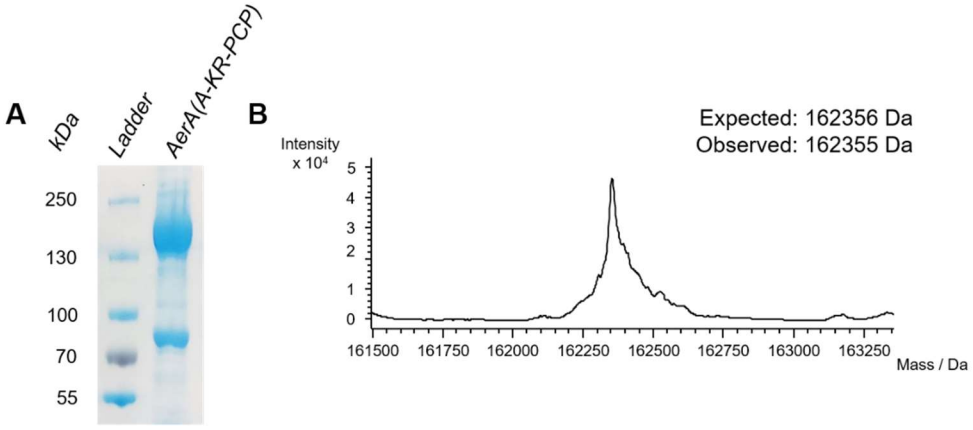


Figure 4.9 Confirmation of identity of recombinant AerA(A-KR-PCP). (**A**) SDS PAGE analysis of AerA(A-KR-PCP) resulting from Ni-affinity and size-exclusion chromatography. (**B**) Mass spectrum of AerA(A-KR-PCP) from UHPLC-ESI-Q-TOF-MS analysis.

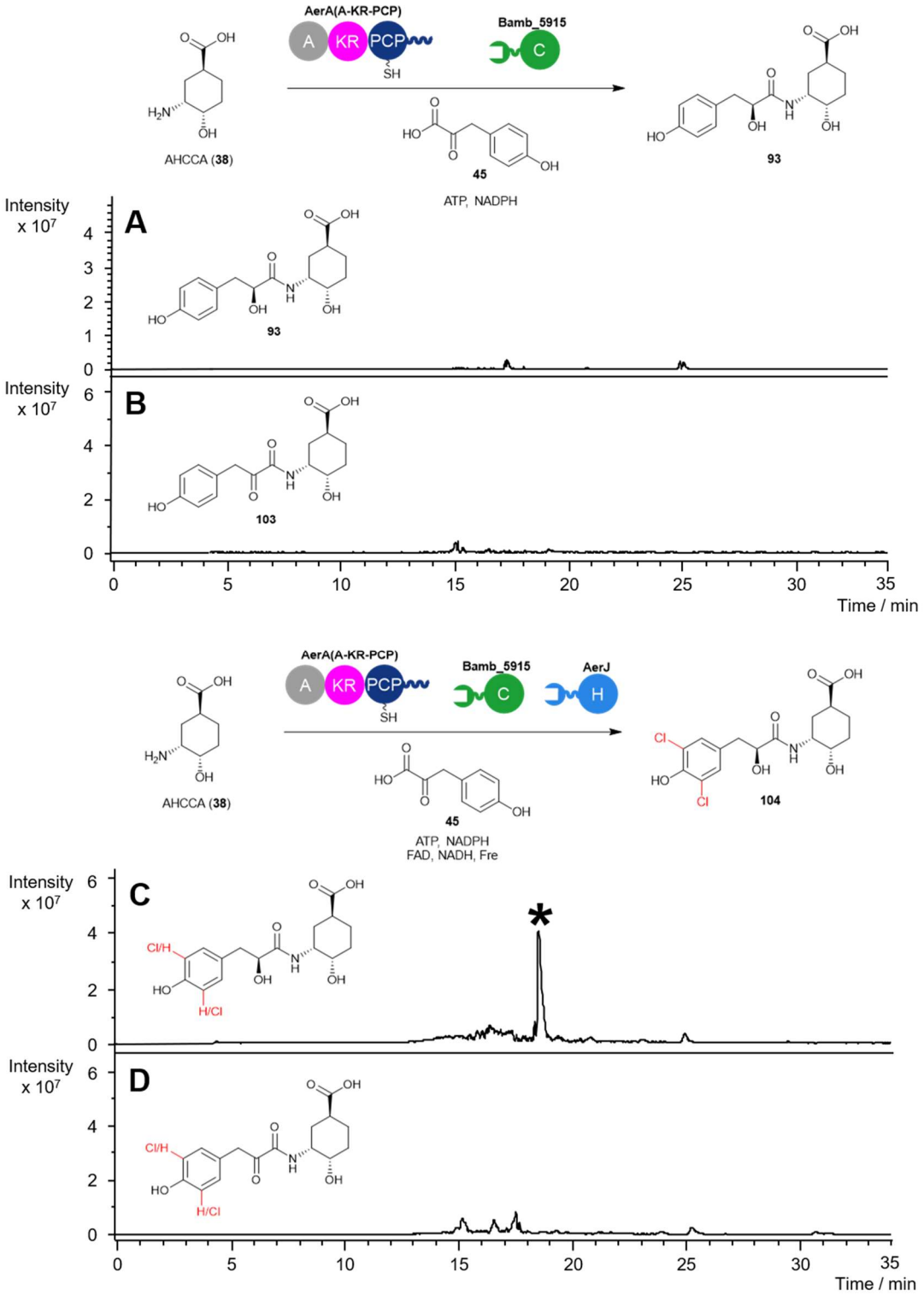


Figure 4.10 Extracted ion chromatograms from UHPLC-IT-MS analysis of products of *in vitro* condensation reaction of AHCCA (**38**) and **45** using AerA(A-KR-PCP) and Bamb_5915. (**A**) No formation of the expected 4-HPLA-AHCCA product **93**. EIC is the sum of 324.14 and 346.13 ± 0.2 Da. (**B**) No formation of potential product **101** where the α-keto group had not been reduced by the KR domain. EIC is the sum of 322.13 and 344.11 ± 0.2 Da. (**C**) Formation of di-chlorinated **93** following introduction of AerJ and cofactors. EIC is the sum of 324.14, 346.13, 358.11, 380.09, 392.07 and 414.05 ± 0.2 Da. An asterisk (*) denotes the potential product peak. (**D**) No formation of a potential product **101** where the α-keto group had not been reduced by the KR domain EIC is the sum of 322.13, 344.11, 356.09, 378.07, 390.05 and 412.03 ± 0.2 Da. In (**A**)-(**D**) cases EIC masses correspond to both the $[M+H]^+$ and $[M+Na]^+$ adducts of the expected products. The mass spectrum of the peak highlighted with an asterisk (*) is provided in figure **4.11**.

To trial the multiple turnover reaction, AerA(A-KR-PCP) was first converted to the *holo*-form by incubation with coenzyme A, Sfp and $MgCl_2$ for 1 h. Then, preincubation with ATP, 4-hydroxyphenylpyruvic acid (**45**) and NADPH for 3 h was carried out to load and reduce the pyruvate substrate. Following this, Bamb_5915 and AHCCA (**38**) were introduced to the reaction and incubated for 1 h in an attempt to generate the condensation product **93** (figure **4.10**). Concurrently, the same reaction with AerJ and cofactors introduced was also attempted. Analysis of the reaction mixture was carried out on an UHPLC-IT-MS instrument.

No formation of expected product **93** was observed in LC-MS analysis of the reaction mixture (figure **4.10 A**). In addition to searching for masses corresponding to **93**, the LCMS trace was also examined for mass of **101**, the product of condensation of the 4-hydroxyphenylpyruvic acid substrate with AHCCA (**38**) before reduction of the α-keto group by the AerA KR domain, however this was not seen either (figure **4.10 B**). Interestingly, in the example with AerJ and cofactors introduced, a peak corresponding to the formation of di-chlorinated **93** was found (figure **4.10.C**). Although an authentic standard was not available, the expected *m/z* and isotopic pattern for di-chlorination were detected (figure **4.11**), giving preliminary evidence of product formation. The reason for only observing a product with introduction of the halogenase

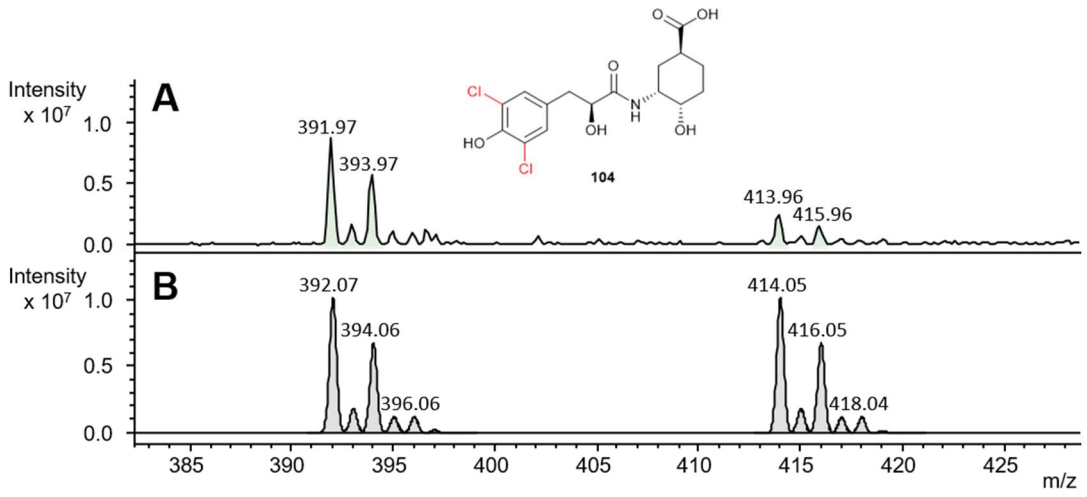


Figure 4.11 Mass spectra of the di-chlorinated 4-HPLA-AHCCA (**93**) product from tandem halogenation and condensation of AHCCA (**38**) and **45** using AerA(A-KR-PCP) with AerJ and Bamb_5915. Mass spectra show $[M+H]^+$ and $[M+Na]^+$ m/z of: (**A**) the dichlorinated product corresponding to the peak labelled with an asterisk (*) in figure **4.10**; (**B**) a simulated mass spectra of the product with a peak width of 0.2 Da and ion intensity of 10^7 A.U.

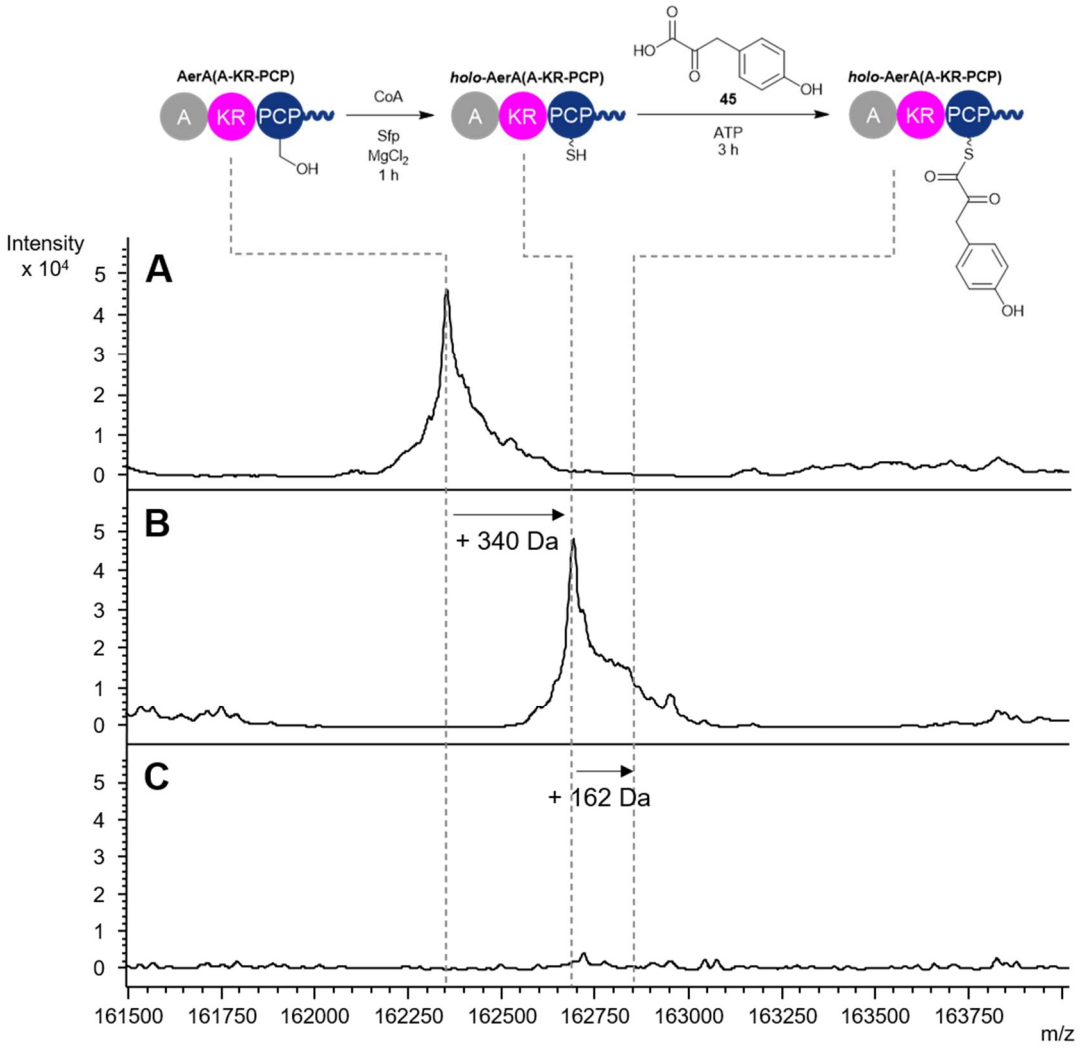


Figure 4.12 UHPLC-ESI-Q-TOF-MS monitoring of intact AerA(A-KR-PCP) demonstrating attempted loading of 4-hydroxyphenylpyruvic acid (**45**). Charge state-deconvoluted spectra show: (**A**) AerA(A-KR-PCP); (**B**) *holo*-AerA(A-KR-PCP) following incubation with CoA, Sfp and $MgCl_2$; (**C**) product of incubation of *holo*-AerA(A-KR-PCP) with ATP and 4-hydroxyphenylpyruvic acid (**45**) following centrifugation and sample preparation.

remains unclear and further replicates are necessary to validate these results. Unfortunately, due to low yields of the protein, there was insufficient AerA(A-KR-PCP) to perform a control reaction without Bamb_5915. Therefore, it is possible that non-enzymatic condensation could be occurring to generate the observed product, although this is unlikely given results from previous experiments with the AerA PCP domain.

The low yield of **93** could be attributed to the poor stability of AerA(A-KR-PCP) in solution, as some precipitation was observed during the *in vitro* reaction. When the reaction was repeated with monitoring by intact protein mass spectrometry, conversion to *holo*-form could be observed, but the protein again

precipitated upon incubation with ATP and 4-hydroxyphenylpyruvic acid (**45**). Following removal of the precipitate by centrifugation, AerA(A-KR-PCP) could no longer be observed in UHPLC-ESI-Q-TOF analysis of the supernatant (figure **4.12**). In future, optimisation of buffer conditions will be necessary to develop this into a viable chemoenzymatic system for generation of natural product hybrids. Alternatively, there are several known fusion tags which can improve the yield and solubility of recombinant proteins which could be introduced at the N-terminus of AerA(A-KR-PCP) to aid solubility of the protein.[226]

An alternative approach was to use AerA PCP domain as the donor for condensation, with intermolecular loading and reduction of 4-hydroxyphenylpyruvic acid (**45**) by a construct containing the A and KR domains. Mutagenesis of the pET24a-(+)-AerA(A-KR-PCP) plasmid to remove the AerA PCP domain was successful, but the AerA(A-KR) didomain was not soluble. Instead, point mutation of the PCP domain's conserved serine residue to alanine was conducted. This would allow intermolecular loading of the standalone AerA PCP domain, without loading the intramolecular PCP domain in the subunit. AerA(A-KR-PCP)S1352A was overproduced and purified as soluble protein by Ni-affinity and size-exclusion chromatography (figure **4.13**).

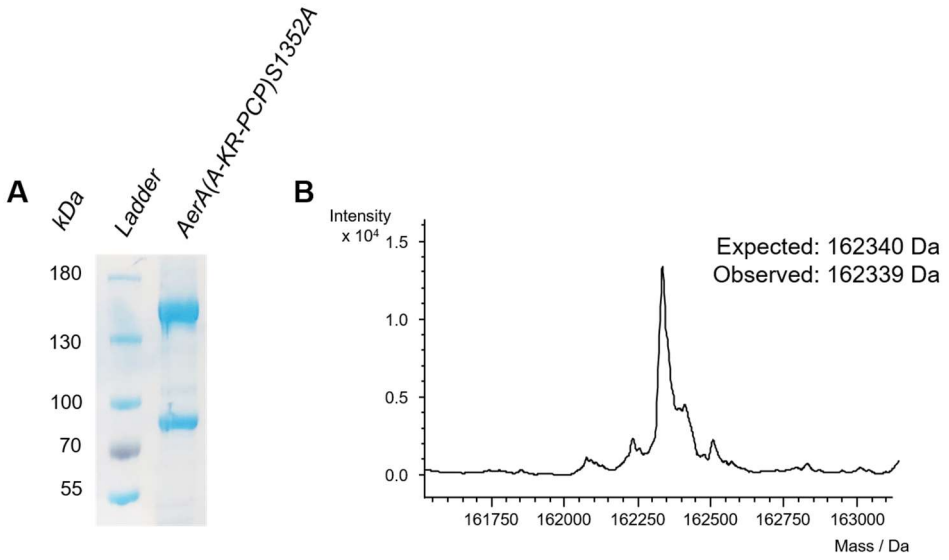


Figure 4.13 Confirmation of identity of recombinant AerA(A-KR-PCP) S1352A. (**A**) SDS PAGE analysis of AerA(A-KR-PCP) resulting from Ni-affinity and size-exclusion chromatography. (**B**) Mass spectrum of AerA(A-KR-PCP) from UHPLC-ESI-Q-TOF-MS analysis of the purified protein.

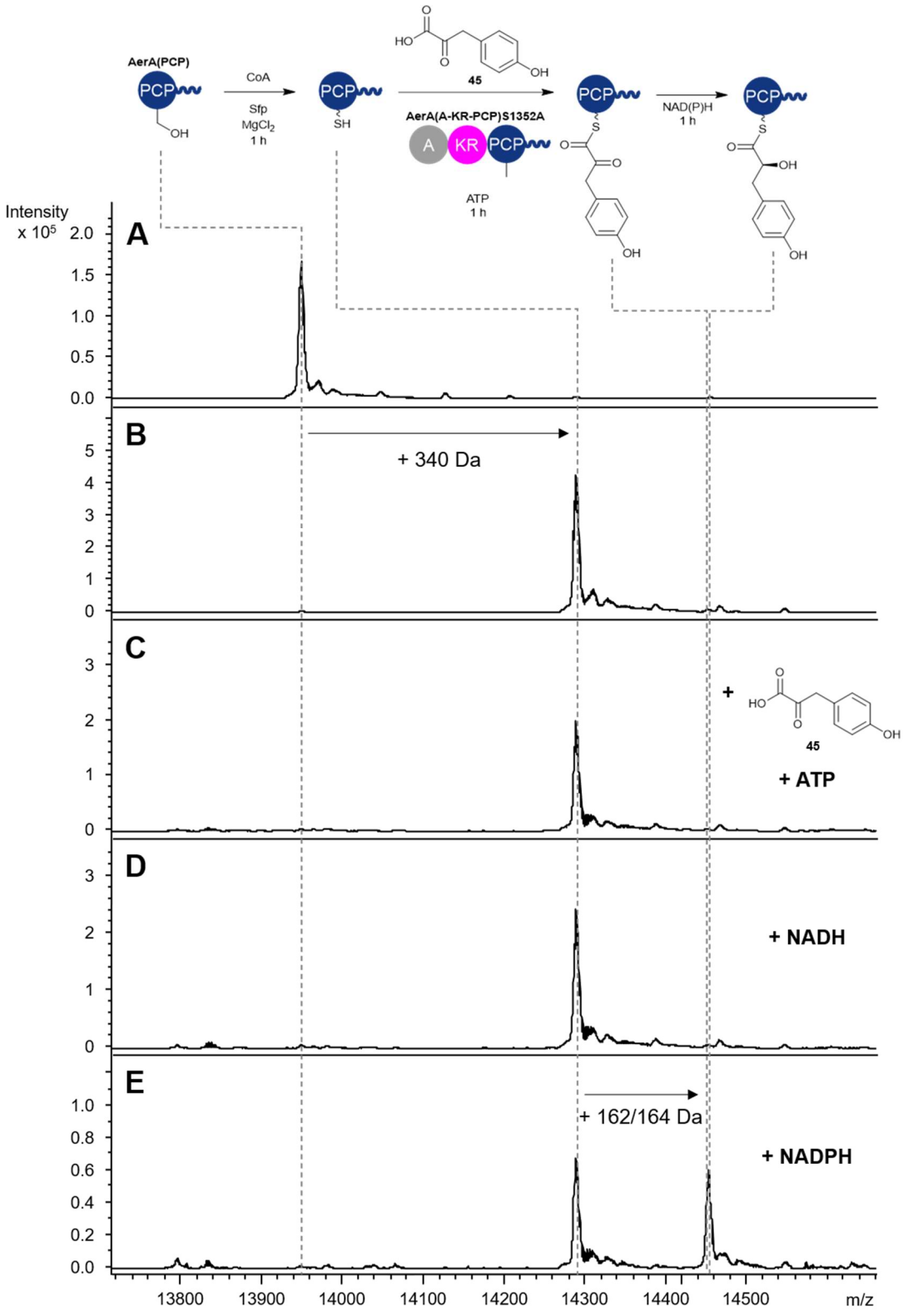


Figure 4.14 UHPLC-ESI-Q-TOF-MS monitoring of intact AerA(PCP) demonstrating attempted loading of 4-hydroxyphenylpyruvic acid (**45**) using AerA(A-KR-PCP)S1352A. Charge state-deconvoluted spectra show: (**A**) *apo*-AerA(PCP); (**B**) *holo*-AerA(PCP) following incubation with CoA, Sfp and $MgCl_2$; (**C**) product of incubation of *holo*-AerA(PCP) with AerA(A-KR-PCP)S1352A, 4-hydroxyphenylpyruvic acid (**45**) and ATP; (**D**) as (**C**) with NADH added; (**E**) as (**C**) with NADPH added.

Conversion of AerA(PCP) to the *holo* form was achieved using coenzyme A, Sfp and $MgCl_2$ as previously described (figure **4.14 B**). Then, a catalytic amount of AerA(A-KR-PCP)S1352A was introduced with ATP and 4-hydroxyphenylpyruvic acid (**45**). No loading of the substrate was observed, so the loading was repeated with the addition of NADH or NADPH (figure **4.14 D-E**). Bioinformatics analysis of the AerA KR domain could not conclusively predict cofactor specificity towards NADH or NADPH, so both were trialled. Interestingly, addition of NADPH did promote loading of the substrate. Presumably, the substrate was also reduced to the 4-hydroxyphenyllactic acyl group, although this could not be distinguished as the 2 Da difference is too small to be clearly observed at the resolution of the Q-TOF instrument. This could be confirmed in future experiments by LC-MS analysis of the product following condensation of the formed acyl-*S*-AerA(PCP) with AHCCA (**38**) catalysed by Bamb_5915. Work is ongoing to scale up this reaction towards allowing milligram scale production of the hybrid natural product derivative. It is currently unclear as to why co-incubating with NADPH allowed loading of the substrate by the A domain. It may be the case that binding of the NADPH cofactor to the KR domain allows global conformational changes in the protein which activates the A domain towards catalysing adenylation and subsequent thioesterification of **45**.

4.2. Progress towards *in vivo* crosstalk of AerA and Bamb_5915 for enacyloxin pathway engineering

An alternative approach to generate aeruginosin-enacyloxin hybrids on a milligram scale is through engineering of the enacyloxin IIa biosynthetic pathway in *B. ambifaria*. An in-frame *bamb_5917* knockout mutant of *B. ambifaria* BCC0203 had previously been constructed *via* homologous recombination.[29] Introduction of *aerA* into *B. ambifaria* BCC0203 Δ*bamb_5917* would allow it to load and process 4-hydroxyphenylpyruvic acid (**45**), naturally produced in the cell as a precursor in tyrosine biosynthesis. The loaded pyruvate could then be condensed with DHCCA (**37**), biosynthesised from shikimic acid, by Bamb_5915. There are two benefits to choosing the *bamb_5917* knockout. Firstly, the strain does not generate background

enacyloxin IIa which will aid detection and purification of any novel molecules, and secondly Bamb_5917 would not be present to compete with AerA for binding to Bamb_5915, which could inhibit the condensation reaction.

To test this approach to *in vivo* crosstalk, an AerA(A-KR-PCP) pMLBAD plasmid construct was generated (figure **4.15 A**). pMLBAD is a plasmid vector designed for regulated gene expression in *Burkholderia*.[227] It contains the *E. coli araBAD* operon consisting of the P_{BAD} promoter and its regulatory gene *araC*.[228] In the absence of L-arabinose, dimeric AraC binds to two regions of the *araBAD* operon, forming a loop in the DNA. This represses gene

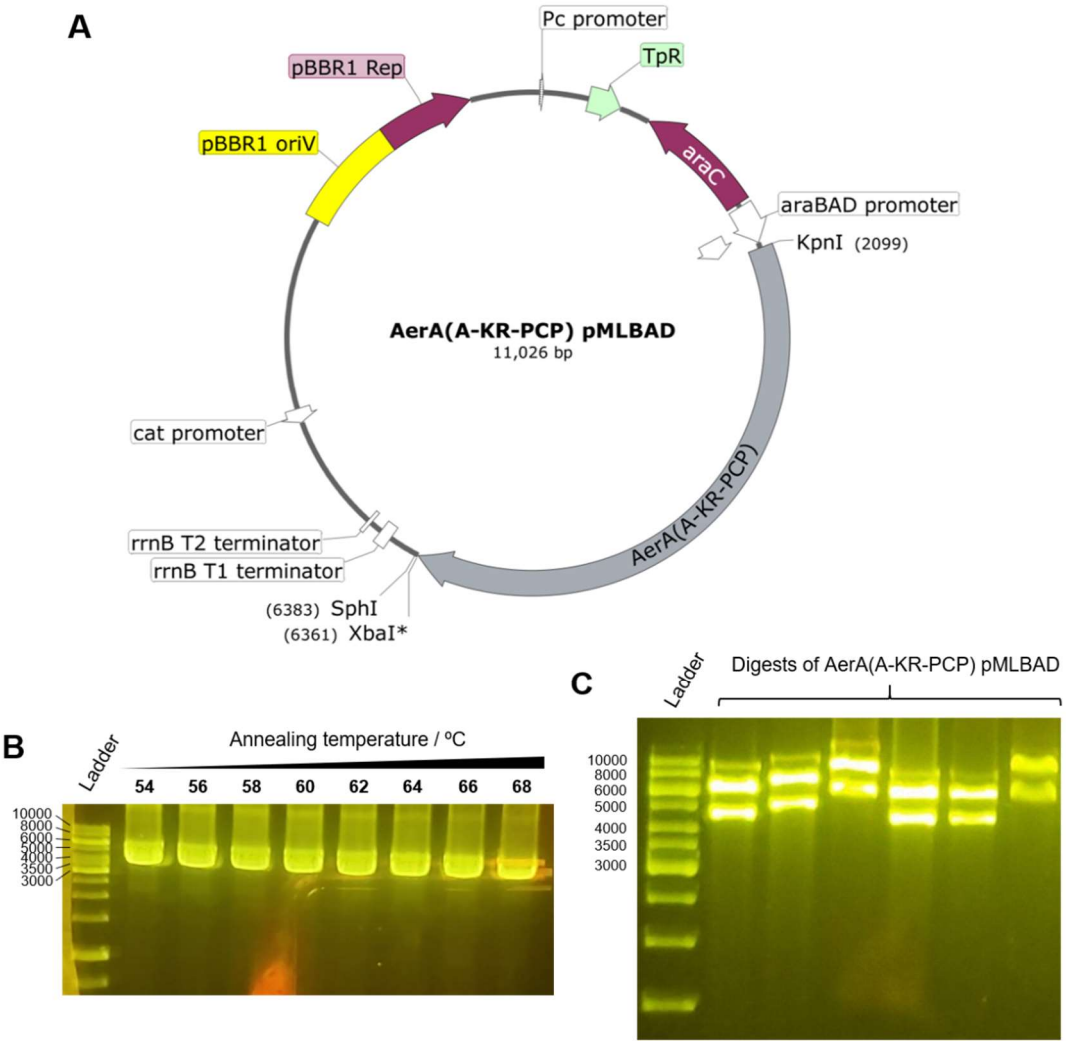


Figure 4.15 Genetic manipulation to generate pMLBAD-AerA(A-KR-PCP). (**A**) Plasmid map of pMLBAD-AerA(A-KR-PCP) demonstrating key translated features and restriction sites used in this work. (**B**) Agarose gel electrophoresis of PCR products resulting from amplification of the *aerA* gene from pET24a-(+)-AerA(A-KR-PCP). Expected product size = 4.3 kb. (**C**) Digest of pMLBAD-AerA(A-KR-PCP) plasmids with KpnI and SphI to check fidelity of final ligation products. Expected product sizes = 6.7 kb, 4.3 kb. The plasmid from the first lane was further verified by sequencing.

expression by preventing RNA polymerases from binding the P_{BAD} promoter. When L-arabinose is present, it binds to an allosteric site of the AraC dimer, causing a conformational change which releases the looped DNA. This allows the P_{BAD} promoter to recruit polymerases to the plasmid DNA, enabling transcription of the insert and subsequent expression of the protein of interest. pMLBAD also houses the *tpR* gene which encodes a dihydrofolate reductase that is not competitively inhibited by trimethoprim, and thus confers trimethoprim resistance.[229]

The *aerA* gene was amplified from pET24a-(+)-AerA(A-KR-PCP) with introduction of 5' KpnI and 3' XbaI restriction sites *via* primer overhangs (figure **4.15 B**). The insert and empty pMLBAD vector were digested using KpnI and XbaI, purified and ligated using T4 DNA ligase. The ligation mixture was introduced into chemically-competent *E. coli* TOP10 cells and transformants selected on LB agar plates supplemented with trimethoprim (50 μg mL^{-1}). The plasmids of transformants were isolated and their fidelity was confirmed by restriction digest (using restriction enzymes KpnI and SphI) and agarose gel electrophoresis analysis (figure **4.15 C**) and subsequent Sanger sequencing of the insert.

Plasmids were then introduced in *B. ambifaria* BCC0203 Δ*bamb_5917 via* triparental mating. This was achieved by co-incubation of the *Burkholderia*

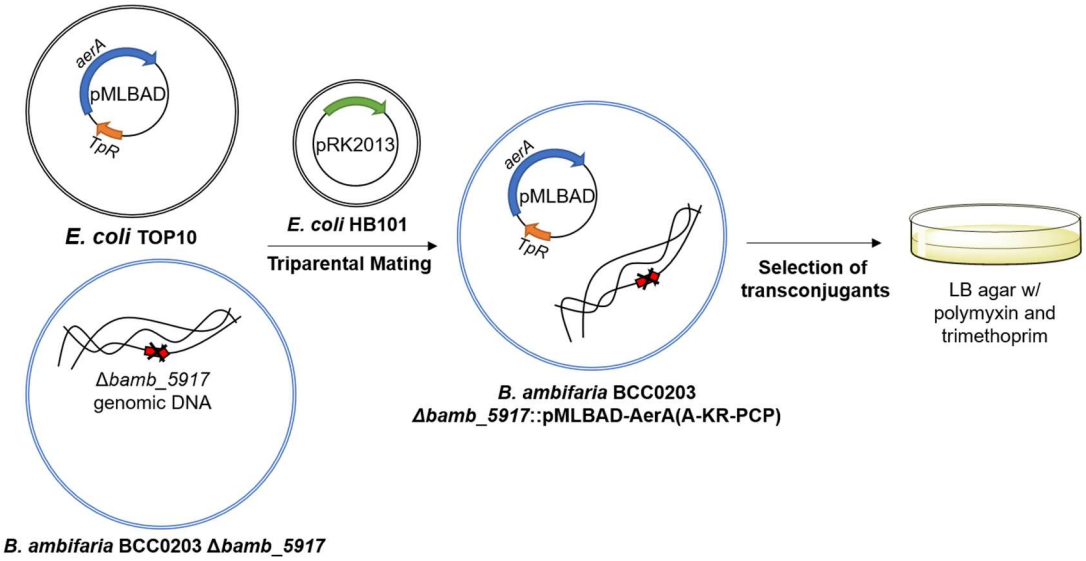


Figure 4.16 Workflow for generation of *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP) transconjugants by triparental mating.

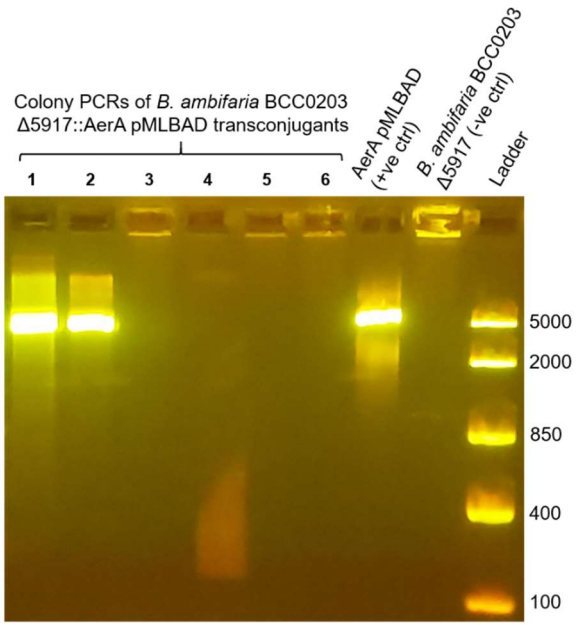


Figure 4.17 Agarose gel electrophoresis demonstrating colony PCR screening of generated *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP) transconjugants. Expected size = 4.3 kb. A positive control where the PCR was conducted with 15 ng of AerA(A-KR-PCP) pMLBAD plasmid and negative control where colony PCR had been performed on *B. ambifaria* BCC0203 Δ*bamb_5917* were also performed.

strain, *E. coli* TOP10 cells containing the plasmid of interest and *E. coli* HB101 containing the pRK2013 plasmid. The latter was included as a helper strain, with the pRK2013 plasmid encoding the RK2 conjugal transfer genes allowing for self-mobilisation into the *E. coli* TOP10 cells, which can then facilitate conjugal transfer of the otherwise non-transmissible plasmid of interest to the *B. ambifaria* BCC0203 acceptor strain (figure **4.16**).[230,231] Transconjugants were selected on LB agar supplemented with trimethoprim, to select for the pMLBAD plasmid, and polymyxin, to which *B. ambifaria* BCC0203 is naturally resistant[232] and allows selection against the *E. coli* strains. Transconjugant colonies were then screened for uptake of the AerA(A-KR-PCP) pMLBAD plasmid by colony PCR (figure **4.17**). Both positive transconjugants identified in figure **4.17** were used to screen for metabolite production.

To screen for metabolite production, transconjugants were plated onto basal salts medium-agar (BSM-agar).[233] BSM-agar is a typical growth medium for *Burkholderia* metabolite production[234] and when supplemented with glycerol as a carbon source enables enacyloxin IIa (**35**) production in *B. ambifaria* BCC0203.[29] Although expression of *aerA* in the pMLBAD plasmid is regulated

by an L-arabinose-inducible promoter, previous complementation studies have shown that L-arabinose-supplemented media is not necessary for complementation with pMLBAD constructs in *B. ambifaria* BCC0203.[29] As a result, initial screens were conducted on media both with and without 0.2% L-arabinose.

Following 3 days of incubation on BSM-agar at 30 °C, cells were removed from the agar. The agar was then cut into small pieces and extracted with organic solvent. EtOAc is typically used as a common extraction solvent, however this was thought to be unsuitable for extracting the polar compound expected to be produced. Instead, agar was extracted with acetonitrile and the solvent removed *in vacuo*. Due to the miscibility of acetonitrile and water, some water was also extracted from the agar. The excess water was further removed by lyophilisation. The resulting extracts were then analysed by UHPLC-ESI-Q-TOF-MS.

As positive controls, wild-type *B. ambifaria* BCC0203 and the Δ*bamb_5917* strain complemented with pMLBAD-Bamb_5917, a transconjugant previously created in the Challis lab,[29] were also grown in parallel (figure **4.18 A** & **C**). These both showed production of a compound with the expected enacyloxin IIa (**35**) *m/z*, demonstrating that the enacyloxin biosynthetic gene cluster was being expressed and productive even with *bamb_5917* provided on a pMLBAD plasmid. This shows that metabolite production was not being suppressed by any unintentional differences in metabolite growth conditions as compared to Masschelein *et al.*[29] It should be noted that multiple peaks are observed extracted ion chromatograms corresponding to enacyloxin IIa. This is typical, as light-induced isomerisation is known to occur due to the highly-conjugated polyene region of enacyloxin, often resulting in a complex mixture of stereoisomers.[235,236]

Extracted ion chromatograms searching for masses of 323.1131, 345.0950, 325.1287 and 347.1107 Da (±0.01 Da), corresponding to the $[M+H]^+$ and $[M+Na]^+$ of 4-hydroxyphenyllactic acid-DHCCA **105** and 4-

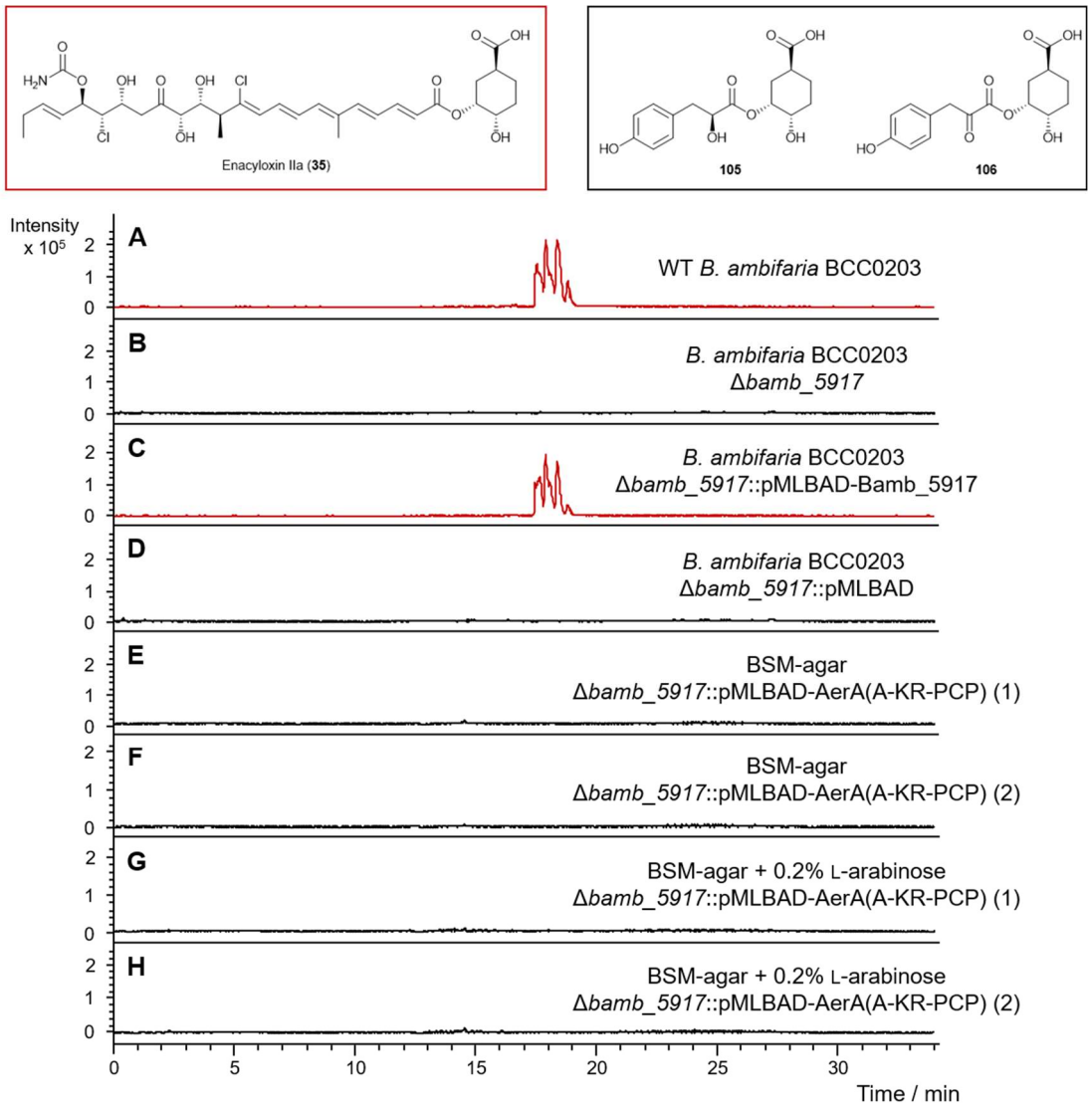


Figure 4.18 Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analysis of metabolite production screens of *Burkholderia ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP) transconjugants grown on BSM agar. (**A**) *WT B. ambifaria* BCC0203. (**B**) *B. ambifaria* BCC0203 Δ*bamb_5917*. (**C**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-Bamb_5917. (**D**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD. (**E**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP), replicate 1. (**F**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP), replicate 2. (**G**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP), grown on BSM-agar supplemented with 0.2% L-arabinose, replicate 1. (**H**) *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP), grown on BSM-agar supplemented with 0.2% L-arabinose, replicate 2. The red chromatograms correspond to the *m/z* for $[M+H]^+$ and $[M+Na]^+$ of enacyloxin IIa (**35**), EIC = sum of 702.2448 and 724.2267 ± 0.01 Da. The black chromatograms correspond to the *m/z* for $[M+H]^+$ and $[M+Na]^+$ of **105** and **106**, EIC = sum of 323.1131, 345.0950, 325.1287 and 347.1107 ± 0.01 Da.

hydroxyphenylpyruvate-DHCCA **106**, of extracts from the two replicate transconjugants displayed no corresponding product peaks (figure **4.18 E-H**).

The reason for the lack of production of compound **105** or **106** was unclear. One explanation is that the concentration of 4-hydroxyphenylpyruvic acid (**45**) in the cell was too low. To probe this, growth of the transconjugants was

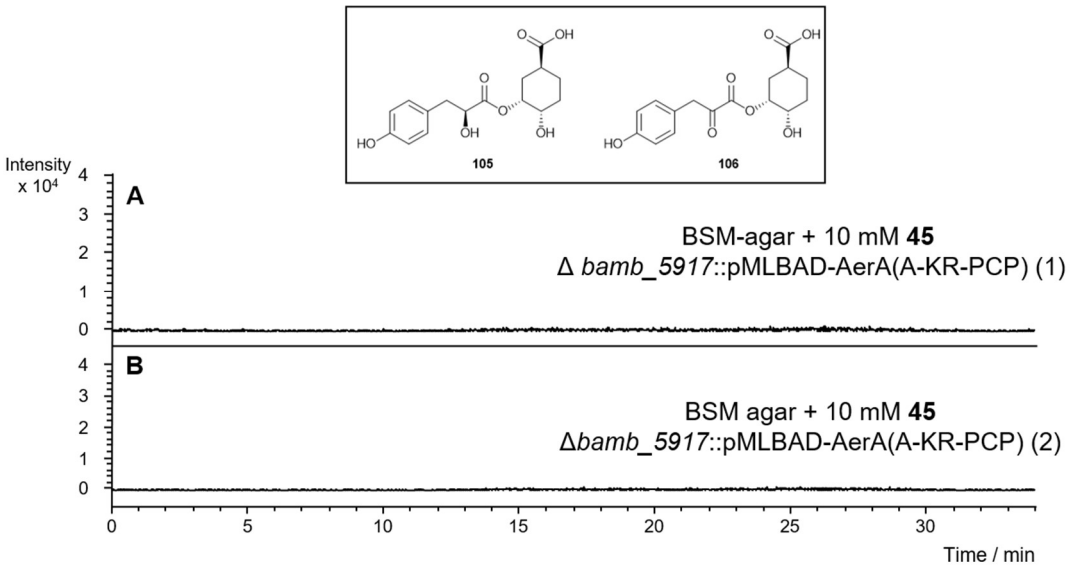


Figure 4.19 UHPLC-ESI-Q-TOF-MS analysis of metabolite production screens of *Burkholderia ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP) transconjugants grown on BSM-agar supplemented with 10 mM 4-hydroxyphenylpyruvic acid **45**. Extracted ion chromatograms show analysis of extracts from (**A**) replicate 1 and (**B**) replicate 2. Chromatograms correspond to the masses of **102** and **103**, EIC = sum of 323.1131, 345.0950, 325.1287 and 347.1107 ± 0.05 Da.

repeated on BSM-agar supplemented with **45** at a concentration of 10 mM. Because of poor solubility of **45** in common solvents, including water and DMSO, it was dissolved directly into warm liquid BSM-agar before pouring of plates. However, UHPLC-ESI-Q-TOF-MS analysis of the extracts, showed no production of **105** or **106** (figure **4.19**).

It remains unclear as to why aeruginosin-enacyloxin IIa hybrids were not produced by *B. ambifaria* BCC0203 Δ*bamb_5917*::pMLBAD-AerA(A-KR-PCP) transconjugants. It may be that *aerA* is not expressed well in *B. ambifaria*. As *aerA* is from a *Microcystis* strain and no codon optimisation was carried out during gene synthesis, the codons of *aerA* may not be compatible with the codon bias of *Burkholderia*, which is known to affect heterologous protein expression.[237,238] Indeed, comparing the codon usage of genes within the enacyloxin gene cluster of related *Burkholderia* species *B. ambifaria* AMMD,[235] for which an annotated enacyloxin IIa gene cluster is available in the MiBiG repository,[239] with codons found in *aerA* from *M. aeruginosa* NIES-98 revealed large differences in codon bias between the nucleotide sequences (table **4.1**).[240] A codon-optimised sequence of *aerA* for expression in *Burkholderia* may need to be designed for optimal expression of the gene for productive *in vivo* crosstalk. It may also also be the case that the

Table 4.1 Comparison of codon usage percentages of degenerate codons in the genes from the enacyloxin biosynthetic gene cluster in *Burkholderia ambifaria* AMMD (MiBiG accession: BGC0001094) and the *aerA* gene used in this study from Microcystis aeruginosa NIES-98. Percentages reported relate to the percentage that each codon is used to encode the related amino acid compared to the other degenerate codons that encode the same amino acid. the given codon appears in coding genes. Colouring of percentages represents the magnitude of the value, where white = 0% and dark green = 100%.

1st Amino Acid of Codon		2nd Amino Acid of Codon: T				C				A				G			
		Amino Acid	Codon	Enacyloxin BGC codon abundance	*aerA* Codon abundance	Amino Acid	Codon	Enacyloxin BGC codon abundance	*aerA* Codon abundance	Amino Acid	Codon	Enacyloxin BGC codon abundance	*aerA* Codon abundance	Amino Acid	Codon	Enacyloxin BGC codon Abundance	*aerA* Codon Abundance
T	T	Phe	TTT	4%	78%	Ser	TCT	1%	28%	Tyr	TAT	25%	87%	Cys	TGT	5%	71%
			TTC	96%	22%		TCC	11%	19%		TAC	75%	13%		TGC	95%	29%
		Leu	TTA	0%	49%		TCA	1%	14%	Stop	TAA	13%	0%	Stop	TGA	83%	100%
			TTG	4%	14%		TCG	48%	6%		TAG	4%	0%	Trp	TGG	100%	100%
	C	Leu	CTT	1%	13%	Pro	CCT	2%	38%	His	CAT	36%	77%	Arg	CGT	6%	28%
			CTC	30%	5%		CCC	28%	34%		CAC	64%	23%		CGC	71%	7%
			CTA	1%	11%		CCA	1%	17%	Gln	CAA	10%	75%		CGA	1%	28%
			CTG	64%	9%		CCG	69%	11%		CAG	90%	25%		CGG	20%	7%
	A	Ile	ATT	3%	70%	Thr	ACT	1%	31%	Asn	AAT	22%	76%	Ser	AGT	2%	28%
			ATC	96%	17%		ACC	47%	31%		AAC	78%	24%		AGC	37%	5%
			ATA	1%	12%		ACA	1%	25%	Lys	AAA	8%	83%	Arg	AGA	0%	23%
		Met	ATG	100%	100%		ACG	51%	13%		AAG	92%	17%		AGG	2%	7%
	G	Val	GTT	2%	47%	Ala	GCT	2%	45%	Asp	GAT	29%	84%	Gly	GGT	5%	22%
			GTC	37%	21%		GCC	45%	14%		GAC	71%	16%		GGC	80%	8%
			GTA	2%	16%		GCA	3%	26%	Glu	GAA	19%	91%		GGA	2%	49%
			GTG	60%	16%		GCG	49%	15%		GAG	81%	9%		GGG	13%	20%

4'-phosphopantetheinyl transferase in *B. ambifaria* BCC0203 is not compatible with the AerA PCP domain, meaning the expressed NRPS module may not be efficiently converted to the *holo*-form *in vivo* and, therefore, remains catalytically inactive. The 4'-phosphopantetheinyl transferase from *B. ambifaria* BCC0203 could be cloned, overproduced and purified and tested for activity towards AerA *in vitro*. If it is found to be inactive towards AerA, the *B. subtilis* 4'-phosphopantetheinyl transferase Sfp could be provided on a separate plasmid to AerA. Alternatively, minimal uptake of 4-hydroxyphenylpyruvic acid (**45**) by the cell due to poor membrane permeability could mean the available concentration of **45** was still too low. Further investigation and optimisations are clearly required before attempting larger scale growth.

4.3. Conclusions and Future Outlook

In this study, crosstalk between proteins from evolutionary divergent bacteria has been explored with a view to exploiting their promiscuity to generate novel natural product analogues. The ability of both βHDD-bearing Bamb_5915 and AerJ to interact with the SLiM-bearing Bamb_5917 and AerA PCP domains has been exploited to generate halogenated and condensed small molecule hybrids incorporating moieties from both the enacyloxin IIa and aeruginosin biosynthetic pathways. Whilst these molecules have not been investigated for biological activity yet, these results provide a platform for generating novel chemical entities which cannot necessarily be easily accessed by traditional chemical synthesis. As more information is uncovered regarding the protein-protein interactions essential for crosstalk through structural and biochemical characterisation, engineering of other proteins to crosstalk in this manner will become viable and further increase the scope of biosynthetic engineering efforts.

A pMLBAD-AerA(A-KR-PCP) construct for complementation has been generated and introduction into a *B. ambifaria* BCC0203 Δ*bamb_5917* mutant has been demonstrated. Whilst production of the expected product **105** was not identified in initial screening of metabolite production, there is further scope

to optimise growth conditions toward production of hybrid natural products in *B. ambifaria* by exploiting the broad substrate scope of Bamb_5915. Once production of the expected metabolite has been established, there is scope to diversify the number of metabolites accessible by this approach. Feeding of synthetic DHCCA analogues into a *B. ambifaria* Δ*bamb5912-14* mutant strain lacking the genes necessary for DHCCA biosynthesis has enabled mutasynthesis of enacyloxin IIa derivatives.[224,225] A similar approach could be used in parallel with introduction of plasmid-borne *aerA* to generate diversity of assembled molecules. Similarly, replacement of the AerA A and KR domains with loading cassettes from other NRPS and PKS systems may enable introduction of alternative moieties to the native 4-HPLA substrate. A domain engineering of the AerA A domain may also be able to achieve this objective. There is also the potential to produce halogenated derivatives *via* introduction of *aerJ*. This could be achieved by providing both *aerA* and *aerJ* on a single plasmid or alternatively integrating one or both of the genes into the *B. ambifaria* genomic DNA.

Chapter 5: Conclusions and perspectives

5.1. Halogenation and condensation in aergunosin biosynthesis

In vitro reconstitution of the initial steps of aeruginosin biosynthesis in *Microcystis aeruginosa* NIES-98 has been described with monitoring by intact protein mass spectrometry. Synthesis and enzymatic loading of 4-HPLA-*S*-pantetheine (**51**) onto the recombinant AerA PCP domain allowed generation of 4-HPLA-*S*-AerA(PCP), which was used to probe the activity of both the flavin-dependent halogenase AerJ and the AerB condensation domain. Incubation of 4-HPLA-*S*-AerA(PCP) with AerJ led to di-chlorination of the carrier protein-tethered 4-HPLA moiety. Incubation of 4-HPLA-*S*-AerA(PCP) with AerB(C-A-PCP-E) loaded with L-isoleucine demonstrated condensation of the 4-HPLA group. These reactions demonstrate that both AerJ and the AerB C domain can productively interact with 4-HPLA-*S*-AerA(PCP), providing a biochemical basis for chlorine incorporation consistent with the presence of SLiM-βHDD interfaces between the PCP domain and both catalytic domains directing the biosynthesis.

Splitting of the AerB subunit allowed monitoring of concurrent halogenation and condensation, resulting a mixture of non-, mono- and di-chlorinated condensation products. Synthesis of 4-HPLA-Ile-*S*-pantetheine (**72**) and loading onto the AerB PCP domain confirmed that halogenation by AerJ cannot occur following condensation. These results together reveal that competition between AerJ and AerB for binding to the AerA PCP domain likely results in the mixture of halogenated congeners of aeruginosin produced by *M. aeruginosa* NIES-98.

Two additional flavin-dependent halogenases with appended N-terminal βHDD domains were identified in previous bioinformatics work; McnD and ApdC from cyanopeptolin and anabaenopeptide biosynthesis, respectively.[123] Interestingly, these halogenases are proposed to act upon either carrier protein-tethered tyrosinyl or hexapeptidyl substrates, and the biosynthetic pathways produce mixtures of mono- and non-chlorinated congeners. Future work could involve examination of the activity of these halogenases *in vitro* to determine substrate specificity and chlorination activity relative to AerJ,

granting a more comprehensive understanding of halogenase recruitment of this type.

5.2. Role of SLiM-βHDD interactions in recruitment of AerJ and AerB(C) by AerA(PCP)

The role of SLiM-βHDD interfaces in the aeruginosin biosynthetic system remains cryptic. Inhibition of the activity of the AerB C domain towards 4-HPLA-*S*-AerA(PCP) by excised docking domains suggests that the docking domain interface is important for the activity of the catalytic domain, consistent with literature results.[123] However, swapping the SLiM region of the AerA PCP domain with that of heterologous carrier proteins, point mutation of the AerJ βHDD and removal of the AerJ βHDD were inconclusive in providing evidence for the necessity of SLiM-βHDD interactions, with little observable difference in productivity found for these constructs *in vitro* compared to the wild-type proteins. It remains unclear as to whether this is a real effect or an artefact of difficulties in equating concentrations of the different recombinant proteins. Nevertheless, current biochemical evidence shows that docking domain binding is not a significant determinant of protein-protein interactions and suggests that other interactions between the PCP and cognate catalytic domains are more important for productive binding.

Biophysical studies on these interfaces concur with this hypothesis. Biolayer Interferometry demonstrated that AerJ and AerB C domain constructs lacking the βHDD interacted with the AerA PCP domain with a lower affinity than the wild-type proteins with an intact docking domain. However, the measured K_D for the ΔβHDD constructs remained on the low μM scale, evidencing that the majority of the interaction interface is determined by binding of the globular portions of both proteins. Moreover, carbene footprinting experiments evidence SLiM-βHDD interface formation between the AerA PCP domain and AerJ, but also highlights regions on the surface of the PCP domain and halogenase are involved in formation of a binding interface. However, a crystal structure of the halogenase is necessary for more in-depth understanding of this binding mode, as homology models lack sequence coverage and may lead

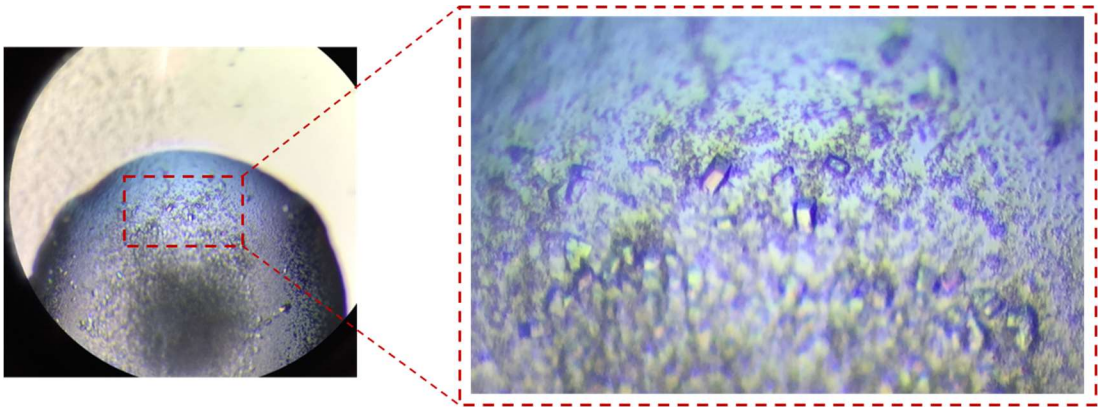


Figure 5.1 Microscope image of initial crystals obtained for flavin-dependent halogenase AerJ. Crystal screening was carried out by Dr. Christopher Fage, University of Warwick.

to inaccurate prediction of the binding site. Optimisation of crystal growth conditions of AerJ are ongoing within the Challis and Lewandowski groups (figure **5.1**).

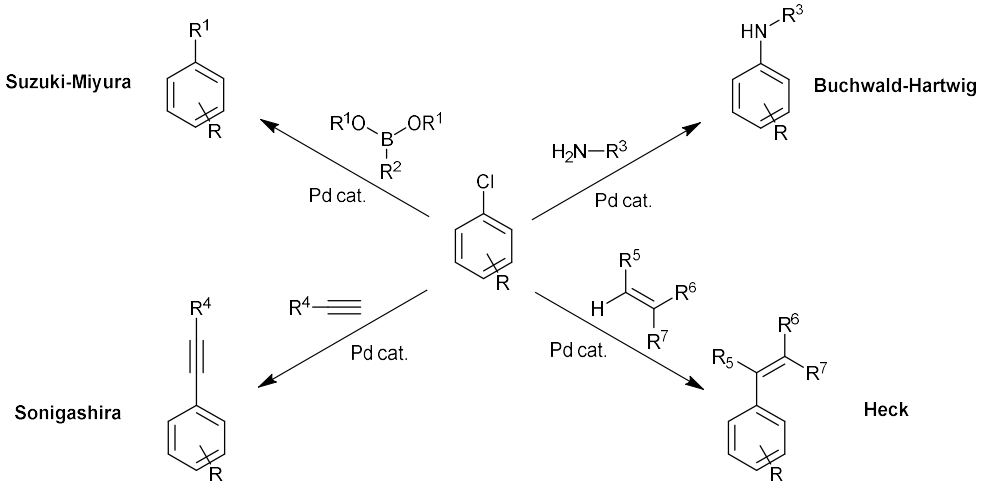
Ongoing solution state NMR titration experiments will provide a more complete insight into the interaction of the AerA PCP domain with AerJ and should give more localised detail regarding the residues of the PCP domain involved in interface formation. In combination with the carbene footprinting results, this could inform targeted mutagenesis studies of the PCP domain to probe residues involved in key interactions with the halogenase, and link protein-protein interactions to the efficiency of the biochemical reactions. Optimisation of the UHPLC-ESI-Q-TOF-MS biochemical reactions developed in this study will help to facilitate such experiments. Further to this, more rigorous biophysical studies of AerA PCP domain-AerB C domain binding using similar methods would allow comparison of the PCP-halogenase and PCP-C domain interfaces. These characterisations will direct future efforts to engineer productive interactions between carrier protein domains and catalytic domain for development of pathway engineering approaches.

For BLI experiments, the was K_D measured with the His_8-affinity tags in place. Although the dissociation constant is comparable to those predicted by ongoing NMR titration experiments, it may be necessary to repeat the BLI measurements with the affinity tag removed to give binding affinities more representative of the native interaction. This could also be explored for the *in*

vitro reactions, especially those where comparison of different PCP domains or different halogenase mutants was attempted, in which the intact affinity tag may influence the protein-protein interactions either positively or negatively and influence the outcomes of these experiments.

5.3. Applicability of AerJ to pathway engineering

Recruitment of halogenases towards carrier protein-tethered substrates represents an attractive strategy for simple introduction of chlorine atoms into natural products *via* pathway engineering. The halogenase AerJ has demonstrated several qualities suggesting it could be an effective tool for such applications. AerJ appears to have a relatively relaxed tolerance towards carrier protein domain-tethered phenols and could allow generation of chlorinated derivatives of complex natural products with minimal active site optimisation, as has been necessary for flavin-dependent halogenases that act upon non-megasynthetase-tethered substrates.[90,170,241] Furthermore, the propensity of AerJ to crosstalk with non-cognate carrier protein domains from divergent biosynthetic pathways, demonstrated in this study by chlorination of Bamb_5917(PCP), Sven_0512(PCP) and BlmVIII(ACP)-tethered substrates, could enable such an approach. This is of particular interest as chlorides make excellent substituents for improving drug-like properties, offering both enhanced physicochemical properties and improved active site binding through halogen bonding. In addition, aryl chlorides are attractive handles for further functionalisation through transition metal-catalysed cross-couplings



Scheme 5.1 Examples of molecular diversity produced by metal-catalysed cross-couplings of aryl halides.

(scheme **5.1**),[242-245] which could allow even greater diversity of accessible natural product analogues. It may be of interest in future work to attempt to construct an *in vitro* system which produces a chlorinated product which could then be subsequently derivatised by cross-coupling to demonstrate the viability of this approach.

AerJ has shown tolerance towards some non-native substrates tethered to the AerA PCP domain, halogenating non-native 4-HPLA-Ile-*S*-AerA(PCP) substrates as well as 4-HPPA-*S*-AerA(PCP). These observations merit a more comprehensive investigation of its substrate tolerance towards fully understanding the utility of AerJ. This could be achieved by synthesis of a range of pantetheine-tethered substrates which could then be loaded onto AerA(PCP) and the products of incubation with AerJ monitored by UHPLC-ESI-Q-TOF-MS analysis using methods described in this study. Interesting targets for such studies would be phenolic moieties found in other natural products, such as tyrosinyl, salicyloyl, 4-hydroxyphenylglycoyl and 2,3-

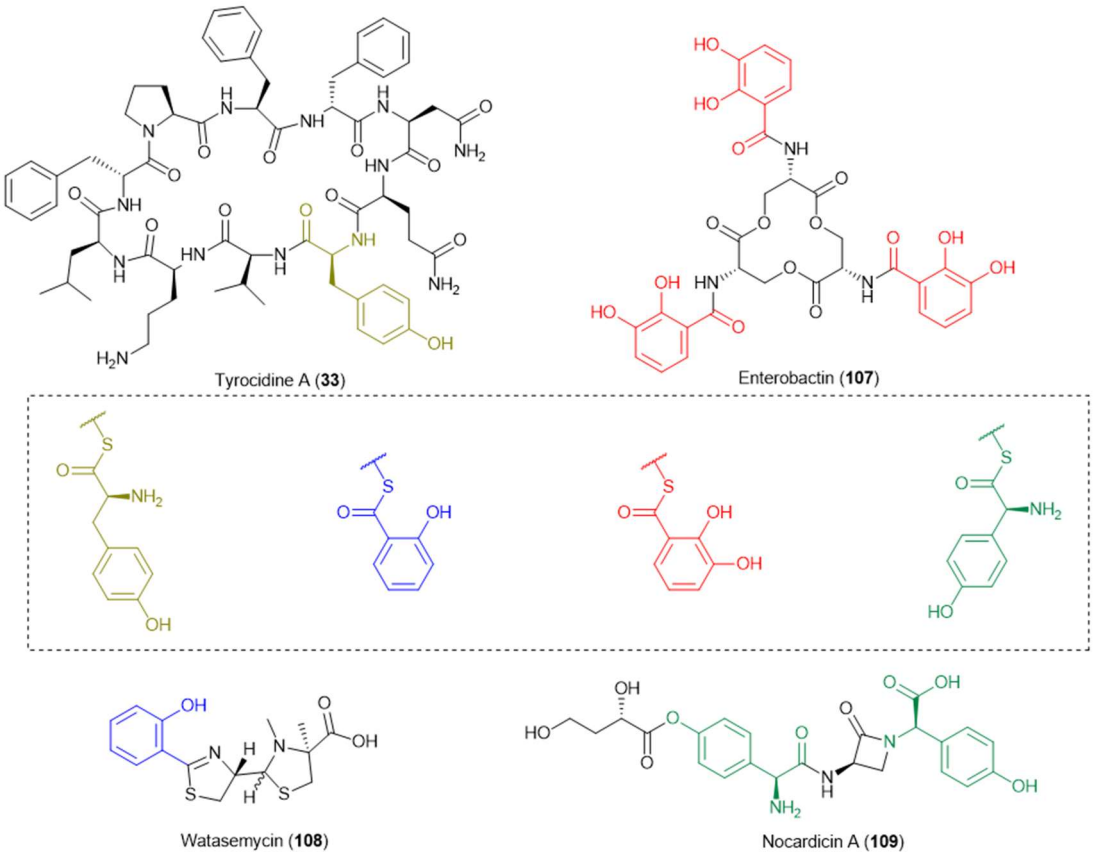


Figure 5.2 Potential phenol-containing substrates to be synthesised and evaluated for chlorination by AerJ and example natural products containing those moieties.

dihydroxybenzoyl groups, amongst others (figure **5.2**). If a crystal structure of AerJ is achieved, co-crystallisation or molecular docking of the native substrate and analogues may give further insights into the range of chemistries that this halogenase may tolerate.

Given that flavin dependent halogenases that can halogenate indole-containing natural product intermediates are known, there is scope to expand halogenase recruitment described in this study to indole-containing carrier protein-tethered substrates. This could be achieved by engineering of the AerJ active site to accept indole substrates, or, conversely, engineering of indole-selective halogenases to act upon carrier protein-tethered substrates. For the former, ongoing efforts to crystallise AerJ and homologues will help to understand the differences in active site architecture of AerJ and halogenases selective for indoles to facilitate active site engineering. For the latter, introduction of a SLiM-βHDD interface may aid recruitment, and carbene footprinting of AerJ modelled onto a crystal structure should provide key regions on the surface of the halogenase which can be engineered to allow productive interactions with carrier protein domains to be realised.

5.4. Rational design of biosynthetic pathways using SLiM-βHDD interfaces

Interaction of SLiM-βHDD interfaces has previously been demonstrated in the literature to mediate productive crosstalk between non-cognate interaction partners.[123,132] In this study, this concept was utilised to generate complex natural products hybrids combining chemistries from aeruginosin and enacyloxin IIa. The broad substrate tolerance of the Bamb_5915 condensation domain from enacyloxin biosynthesis was exploited to generate the 4-HPPA-AHCCA conjugate **97.** This could be achieved with both 4-HPPA-*S*-Bamb_5917(PCP) or 4-HPPA-*S*-AerA(PCP) as the acyl donor. Additionally, introduction of halogenase AerJ produced chlorinated derivatives, giving further diversity in the product scope of this system, and use of the Bamb_5917 PCP or AerA PCP domains gave mostly mono- and di-chlorinated products, respectively. Whilst these studies are very much proof-of-principle, they give

precedent for the generation of robust *de novo* biosynthetic pathways by exploiting protein-protein interaction promiscuity to produce small molecules with potential biological activity.

Generation of a plasmid construct for heterologous expression of AerA(A-KR-PCP) in *Burkholderia* was successful, and introduction to *Burkholderia ambifaria* BCC0203 Δ*bamb_5917* by triparental mating was also achieved. However, production of the expected metabolite 4-HPLA-DHCCA (**105**), was not observed *in vivo* using this system. It remains unclear as to the cause of this. Future work involving this system could involve codon optimisation of *aerA* to see if this allows expression of the protein *in vivo*. Alternatively, optimisation of feeding conditions of the *Burkholderia* transconjugant may prove valuable. Successful reconstitution of this reaction could facilitate several unique expansions of this work, including mutasynthesis, introduction of the halogenase AerJ, A domain engineering of AerA(A-KR-PCP) and semi-synthesis of derivatives of **105**. This could lead to generation a library of molecules which could then be trialled for biological activity, the results of which could direct future target compounds.

Recruitment of AerJ to the AerA PCP domain to the best of our knowledge represents the first characterised example of a docking domain pair directing activity of a structure-tailoring enzyme, rather than an interaction between PKS or NRPS subunits. The implications of this for pathway engineering are vast, as engineering of protein-protein interactions could hypothetically enable recruitment of a wide range catalytic enzymes to act upon growing polyketide or nonribsomsal peptide chains during biosynthesis. This could allow access to a large diversity of otherwise inaccessible natural product analogues. For this approach to be successful, however, more information on the key protein-protein interactions involved in recruitment of catalytic domains by carrier protein domains is essential.

Chapter 6: Experimental

6.1. Biological Procedures

6.1.1. Equipment and Instruments

UHPLC-ESI-Q-TOF-MS analysis of intact proteins were performed on a Bruker MaXis II ESI-Q-TOF-MS connected to a Dionex 3000 RS UHPLC fitted with an ACE C_4-300 RP column (100 × 2.1 mm, 5 µm, 30 °C). 2.0 µL of sample was injected. The column was eluted with a linear gradient of 5–100% MeCN containing 0.1% formic acid in H_2O containing 0.1% formic acid over 30 mins at a flow rate of 0.2 mL min^{-1}. The mass spectrometer was operated in positive ion mode with a scan range of 200–3,000 m/z. Source conditions: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.8 bar; dry gas (N_2) at 9.0 L min^{-1}; dry temperature at 200 °C. Ion transfer conditions: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 300 *m/z*; collision energy at 8.0 eV; collision RF at 2,000 Vpp; transfer time at 110.0 µs; pre-pulse storage time at 10.0 µs.

UHPLC-ESI-Q-TOF-MS analysis of digested peptide fragments in carbene footprinting experiments were performed on a Bruker MaXis II ESI-Q-TOF-MS connected to a Dionex 3000 RS UHPLC fitted with a ZORBAX Eclipse Plus C_{18} column (100 × 2.1 mm, 1.8 µm, 30 °C). 2.0 µL of sample was injected. The column was eluted with a linear gradient of 5–100% MeCN containing 0.1% formic acid in H_2O containing 0.1% formic acid over 30 mins at a flow rate of 0.2 ml min^{-1}. The mass spectrometer was operated in positive ion mode with a scan range of 50–3,000 m/z. Source conditions: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.8 bar; dry gas (N_2) at 8.0 L min^{-1}; dry temperature at 180 °C. Ion transfer conditions: ion funnel RF at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 *m/z*; collision energy at 5.0 eV; collision RF at 600 Vpp; transfer time at 121.0 µs; pre-pulse storage time ats 1.0 µs. Calibration of the MS was performed by injection of 20 µL of 1 mM sodium formate solution prior to each run.

UHPLC-ESI-Q-TOF-MS analyses of small molecules were performed on a Bruker MaXis Impact ESI-Q-TOF-MS coupled to a Dionex Ultimate 3000 HPLC instrument fitted with a ZORBAX eclipse plus C_{18} column (100 x 2.1

mm, 1.8 µm). 2.0 µL of sample was injected. The column was eluted with a linear gradient of 5-100% MeCN containing 0.1% formic acid in H_2O containing 0.1% formic acid over 34 minutes at a flow rate of 0.2 ml min^{-1}. The mass spectrometer was operated in positive ion mode with a scan range of 50-2500 *m/z*. Source conditions were: end plate offset at -500 V; capillary at -4500 V; nebulizer gas (N_2) at 1.4 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 200 °C. Calibration of the MS was performed by injection of 20 µL of 1 mM sodium formate solution prior to each run.

UHPLC-ESI-IT-MS of small molecules were performed on a Bruker AmaZon ETD speed ESI-IT-MS coupled to an Dionex 3000 RS UHPLC fitted with an ZORBAX eclipse plus C_{18} column (100 × 2.1 mm, 1.8 µm, 30 °C). 5.0 µL of sample was injected. The column was eluted with a linear gradient of 5-100% MeCN containing 0.1% formic acid in H_2O containing 0.1% formic acid over 28 mins with a flow rate of 0.2 mL min^{-1}. The mass spectrometer was operated in positive ion mode with a scan range of 70-3000 m/z. Source conditions: end plate offset at -500 V; capillary at -4500 V; nebulizer gas (N_2) at 1.7 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 °C.

HPLC-ESI-IT-MS of small molecules were performed on a Bruker AmaZon X ESI-IT-MS coupled to an Agilent 1260 Infinity HPLC instrument fitted with an ZORBAX eclipse plus C18 column (150 x 4.6 mm, 3.5 µm). 20 µL of sample was injected. The column was eluted with a linear gradient of 5-100% MeCN containing 0.1% formic acid in H_2O containing 0.1% formic acid over 45 mins at a flow rate of 1.0 mL min^{-1}. The mass spectrometer was operated in positive ion mode with a scan range of 100-3000 m/z. Source conditions: end plate offset at -500 V; capillary at -4500 V; nebulizer gas (N_2) at 2.8 bar; dry gas (N_2) at 10 L min^{-1}; dry temperature at 200 °C.

Semi-preparative HPLC was performed on an Agilent 1260 instrument fitted with a BetaSil C_{18} column (150 x 21.2 mm, 5 µm particle size). Absorbances were monitored at 210, 254 and 280 nm. Compounds were eluted with a linear gradient of HPLC grade acetonitrile or MeOH (as stated for each compound)

in H_2O containing 0.1% formic acid at a flow rate of 10 mL min^{-1}. 2 mL elution fractions were collected.

Media was sterilised in a Priorclave autoclave at a temperature of 121 °C and pressure of 1 bar for 10 min. Buffers and solutions were sterile-filtered by passing through a 0.22 µm Minisart syringe filter or 1 L Nalgene 0.2 µm filter unit (Thermo Scientific), as appropriate.

Liquid cultures of *E. coli* were incubated in New Brunswick Scientific Innova shakers. Solid cultures of *E. coli* and *Burkholderia* were incubated in Thermo Scientific Heraus static incubators *Burkholderia* strains were handled in a BSL-2 level lab fitted with a Bassaire laminar flow hood. Optical densities of bacterial cultures were measured using a Biochrom WPA CO8000 spectrophotometer with readings at a wavelength of 600 nm.

Benchtop centrifugation of Eppendorf tubes was carried out using Eppendorf models 5415D, 5424, 5804 or 5810 centrifuges. Centrifugation of bacterial cultures and cell lysates was conducted using a Hitachi Koki Himac CR22N using R9A or R20A2 rotors, as appropriate.

Cell lysis was carried out using a Constant Systems TS Cabinet cell disruptor (high-pressure cell disruption) or using a Sonics & Materials Inc. Vibra-Cell™ VCX 130 Ultrasonic liquid processor fitted with a 6 mm probe in a Rosett glass cooling cell in an ice bath (sonication).

DNA and protein concentrations were determined by measurements at wavelengths of 260 nm and 280 nm, respectively, with a Thermo Scientific NanoDrop Lite instrument unless otherwise stated. Concentration determination by Bradford assay was conducted using a Varian Cary 50 UV-Visible Spectrophotometer using a wavelength of 595 nm.

Lyophilisation was conducted using a LaboGene ScanVac Coolsafe 15 L freeze dryer.

PCRs were carried out using an Eppendorf Mastercycler nexus X2. SDS-PAGE and agarose gel electrophoresis of DNA was conducted using a PowerPac 300. Visualisation of agarose gels was conducted on a UVP BioDoc Imaging System UV Transilluminator (GelRed stain) or Nippon Genetics FastGene B/G LED Transilluminator (Midori Green stain).

Sequencing was performed by Eurofins Genomics (formerly GATC Biotech) *via* Sanger sequencing methodology.

6.1.2. Materials

6.1.2.1. Reagents

LB broth powder (Miller) was purchased from Fisher Scientific or Sigma-Aldrich.

GelRed nucleic acid stain was purchased from Cambridge Scientific. Midori Green nucleic acid stain and Midori Green direct DNA stain were purchased from Scientific Laboratories Supplies Ltd. GeneRuler 1 kb DNA ladder and FastRuler Middle Range DNA ladder were purchased from Thermo Fisher Scientific. GeneJET Plasmid Miniprep Kit and GeneJET Gel Extraction Kit were purchased from Thermo Fisher Scientific. Primers for mutagenesis and sequencing were designed with the aid of NEBasechanger and purchased from Sigma-Aldrich purified by desalting. Q5 Hot Start High Fidelity 2 x polymerase mastermix and OneTaq Hotstart quick-load polymerase mastermix were purchased from Thermo Fisher Scientific. KLD reaction mix and the appropriate buffer were purchased from New England Biolabs. Restriction enzymes and their appropriate buffers were purchased from New England Biolabs and stored at -20 °C. T4 ligase and appropriate buffer was purchased from Thermo Fisher Scientific.

Pageruler and Pageruler Plus protein molecular weight reference ladder was purchased from Thermo Scientific and used as provided. PAGEmark protein Blue PLUS protein reference ladder was purchased from G-Biosciences InstantBlue Coomassie Protein Stain was purchased from Expedeon.

FAD, ATP and methylamine (40% wt in water) were purchased from Sigma-Aldrich. CoA, NADH and NADPH were purchased from VWR. L-isoleucine was purchased from Alfa-Aesar. acetyl-*S*-CoA was purchased from Fisher Scientific. All reagents were used with no further purification.

6.1.2.2. Growth Media

LB medium was prepared using LB broth powder (Miller) to a final concentration of 25 g L^{-1} and was autoclaved prior to inoculation.

Basal salts medium-agar (BSM-agar) was prepared by dissolving NaH_2PO_4 (1.0 g/L), $K_2HPO4.3H_2O$ (4.3 g/L), NH_4Cl (2.0 g/L), nitriloacetic acid (0.1 g/L), Bacto casamino acids (0.5 g/L), Bacto yeast extract (0.5 g/L), and glycerol (4.0 g/L) in dH_2O. A 50 x solution of metal salts was prepared by dissolving $MgSO_4.7H2O$ (10 g/L), $FeSO_4.7H_2O$ (0.5 g/L), $MnSO_4.H_2O$ (0.15 g/L), $ZnSO_4.H_2O$ (0.15 g/L), $CoSO_4.7H_2O$ (0.05 g/L) in dH_2O, and was added to the solution to give a final 1 x concentration. Bacto agar (15 g L^{-1}) was added and the medium was sterilised by autoclaving.

6.1.2.3. Antibiotics

Antibiotics used in this project are described in table 6.1. The working concentration used is described in the applicable section.

Table 6.1 Antibiotic stocks used in this project.

Antibiotic	Stock Concentration	Solvent	Supplier
Kanamycin	50 mg mL^{-1}	dH_2O	Sigma-Aldrich
Ampicillin	100 mg mL^{-1}	dH_2O	Melford
Trimethoprim	50 mg mL^{-1}	Sterile DMSO	Sigma-Aldrich
Polymyxin	30,000 U mL^{-1}	dH_2O	Sigma-Aldrich

6.1.2.4. Plasmids

Table 6.2. Origins of plasmids used in this study.

Plasmid	Origin of Plasmid	Use in this study	Resistance Marker
pET24a-(+)-AerA(PCP)	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of AerA(PCP)	Kan
pET24a-(+)-AerA(A-KR-PCP)	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of AerA(A-KR-PCP)	Kan
pET24a-(+)-AerJ	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of AerJ	Kan
pET24a-(+)-AerB(C)	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of AerB(C)	Kan
pET24a-(+)-AerB(C-A-PCP-E)	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of AerB(C-A-PCP-E)	Kan
pET24a-(+)-McnD	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of McnD	Kan
pET24a-(+)-ApdC	Gene synthesised and cloned by Epoch Life Sciences Inc.	Overproduction of ApdC	Kan
pET151-Bamb_5915	Challis group stock, cloned from *B. ambifaria* AMMD genomic DNA described by Masschelein *et al.*[29]	Overproduction of Bamb_5915	Amp
pET151-Bamb_5917 (PCP)	Challis group stock, cloned from *B. ambifaria* AMMD genomic DNA described by Masschelein *et al.*[29]	Overproduction of Bamb_5917(PCP)	Amp

pET151-Sven_0512(PCP)	Challis group stock, cloned from *S. venezuelae* genomic DNA as described by Kosol *et al.*[123]	Overproduction of Sven_0512(PCP)	Amp
pET28a-SsuE	Challis group stock, cloned from *E. coli* K12 genomic DNA as described by Dorrestein *et al.*[40]	Overproduction of SsuE	Kan
pET45b-Fre	Dr. Binuraj Menon, cloned from *E.coli* BL21 genomic DNA as described by Shepherd *et al.*[170]	Overproduction of Fre	Kan
pET29b-PanK	Dr. Manuela Tosin, cloned from *E. coli* K12 genomic DNA as described by Tosin *et al.*[166]	Overproduction of CoaA	Kan
pET29b-Ppat	Dr. Manuela Tosin, cloned from *E. coli* K12 genomic DNA as described by Tosin *et al.*[166]	Overproduction of CoaD	Kan
pET20b-(+)-DPCK	Dr. Manuela Tosin, cloned from *E. coli* K12 genomic DNA as described by Tosin *et al.*[166]	Overproduction of CoaE	Amp
pET28a-Sfp	Dr. Manuela Tosin, cloned from *Bacillus subtilis* genomic DNA.	Overproduction of Sfp	Kan
pMLBAD	Challis group stocks, originally sourced from Valvano and co-workers.[227]	Empty vector, for complementation of *Burkholderia* strains	Tpr

6.1.2.5. Proteins

Peroxidase from horseradish type IV was purchased from Sigma-Aldrich. A 1 mM stock solution was made by dissolving in storage buffer with 10% glycerol, flash-freezing in liquid N_2 and storing at -80 °C until required.

Aliquots of Sven_0512(PCP) and Bamb_5917(PCP)-AerA SLiM were kindly provided by Dr. Shanshan Zhou. Aliquots of BlmVIII(ACP) and TycA(PCP) were kindly provided by Dr. Christopher Fage. Aliquots of PcdC(ACP) were kindly provided by Munro Passmore. These were all provided as N-terminal His_6-tagged proteins and the identity and purity of each protein was confirmed by SDS-PAGE and UHPLC-ESI-Q-TOF-MS analysis prior to use.

6.1.2.6. Bacterial Strains

The bacterial strain used in this study are provided in table **6.3**. For *E. coli* BL21(DE3), C43(DE3), TOP10 and NEB5α strains, commercial strains were used as parent cultures to make competent cells stocks, rather than used directly for transformations.

Table 6.3 Bacterial strains used within this project.

Strain	Use of strain in this study	Origin of Strain
One Shot™ *E. coli* TOP10	Plasmid propagation	Invitrogen
E. coli NEB5α	Plasmid propagation	New England Biolabs
One Shot™ *E. coli* BL21(DE3)	Protein overproduction	Invitrogen
Overexpress™ *E. coli* C43(DE3)	Protein overproduction	Sigma-Aldrich
E. coli HB101 containing pRK2013 plasmid[246]	Helper strain for conjugation between *E.coli and Burkholderia*	Challis group strain

Burkholderia ambifaria BCC0203 WT[235]	Enacyloxin IIa producing strain	Challis group strain
Burkholderia ambifaria BCC0203 Δ*bamb_5917*[29]	*bamb_5917* knockout mutant of *B. ambifaria* BCC0203	Challis group strain
Burkholderia ambifaria BCC0203 Δ*bamb_5917::bamb5917*[29]	*bamb_5917* knockout mutant of *B. ambifaria* BCC0203 complemented with pMLBAD-Bamb_5917	Challis group strain

6.1.2.7. Buffers

10 x TBE buffer was prepared by dissolving trisaminomethane.HCl (108 g) and Boric acid (55 g) in 900 mL dH_2O and adding 0.5 M Ethylenediaminetetraacetic acid solution (40 mL). The volume was adjusted to 1 L by addition of dH_2O. The buffer was diluted to 1 x concentration prior to use.

10 x SDS-PAGE running buffer was prepared by dissolving trisaminomethane.HCl (30 g), glycine (144 g) and sodium dodecylsulfate (10 g) in 1 L dH_2O. The buffer was diluted to 1 x concentration prior to use.

5 x SDS-PAGE sample loading buffer was prepared by dissolving sodium dodecyl sulfate (10 g), 2-mercaptoethanol (5 mL), bromophenol blue (0.02 g), glycerol (30 g) and trisaminomethane.HCl (3.9 g) in 100 mL dH_2O. The pH was adjusted to pH 6.8 by careful addition of 6 M HCl. The buffer was diluted to 1 x concentration with protein sample prior to use.

Loading buffer for Ni-NTA chromatography was prepared by dissolving trisaminomethane-HCl (2.42 g), NaCl (8.76 g), imidazole (1.36g) and glycerol (10 g) in 1 L dH_2O. The pH was adjusted to 7.5 by careful addition of 6 M HCl. The buffer was sterile-filtered prior to use.

Elution buffer (500 mM imidazole) for Ni-NTA chromatography was prepared by dissolving trisaminomethane-HCl (2.42 g), NaCl (8.76 g), imidazole (34.0

g) and glycerol (10 g) in 1 L dH_2O. The pH was adjusted to 7.5 by careful addition of 6 M HCl. The buffer was sterile-filtered prior to use. 50 mM, 100 mM, 200 mM and 300 mM imidazole elution buffers were prepared by mixing of this buffer with loading buffer to give the desired concentration.

Storage buffer was prepared by dissolving trisaminomethane-HCl (2.42 g), NaCl (8.76 g), imidazole (1.36g) and 100 mL glycerol in 900 mL dH_2O. The pH was adjusted to 7.5 by careful addition of 6 M HCl. The buffer was sterile-filtered prior to use.

Gel filtration buffer for size-exclusion chromatography of the full AerB subunit was prepared by dissolving 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (2.38 g), NaCl (8.76 g) and glycerol (5 g) in 1 L dH_2O. The pH was adjusted to 7.5 by careful addition of 6 M HCl. The buffer was sterile-filtered prior to use.

Gel filtration buffer for determining dimerization state of proteins by size-exclusion chromatography prepared by dissolving $Na_2HPO_4.7H_2O$ (20.2 g), $NaH_2PO_4.7H_2O$ (3.4 g) in 1 L dH_2O. The pH was adjusted to 7.5 by careful addition of 6 M HCl. The buffer was sterile-filtered prior to use.

6.1.3. DNA Manipulation

6.1.3.1. Preparation of chemically competent *E. coli*

E .coli BL21(DE3), C43(DE3), TOP10 or NEB5α from -80 °C freezer stocks were streaked onto LB agar and grown overnight at 37 °C. A single colony was used to inoculate 10 mL LB and incubated overnight (37 °C, 180 rpm). The resulting culture was diluted 1:100 to a final volume of 1 L and incubated (37 °C, 180 rpm) until the OD_{600} measure approximately 0.4. The cells were pelleted (15 min, 3000 rpm, 4 °C). The supernatant was removed, the pellet resuspended in ice-cold sterile 0.1 M $MgCl_2$ (100 mL) and incubated on ice for 20 mins. The suspension was pelleted (15 min, 3000 rpm, 4 °C). The pellet was resuspended in ice-cold sterile 0.1 M $CaCl_2$ (20 mL). Sterile glycerol was added to a final concentration of 10%. The solution was aliquoted into 50 µM aliquots, flash frozen in liquid N_2 and stored at -80 °C until use.

6.1.3.2. Transformation of chemically competent *E. coli* with plasmid DNA

An aliquot of competent *E. coli* TOP10 or NEB5α was thawed on ice. 100 ng of plasmid DNA (as determined by A_{260} measurements) or 5 µL of ligation reaction was added and mixed by gentle pipetting. The mixture was incubated on ice for 20 mins and then heat-shocked for 30-45s at 42 °C. The mixture was incubated on ice for 5 mins. 950 µL of sterile LB was added and the mixture incubated at 37 °C, 180 rpm for 1 h to allow the cells to develop resistance from the plasmid-borne resistance marker. 100 µL of the transformation culture was plated onto LB agar supplemented with the appropriate antibiotic (1 in 1000 dilution from the stock concentration). The plates were incubated at 37 °C overnight.

6.1.3.3. Plasmid Isolation

A single colony of *E. coli* containing the plasmid of interest was used to inoculate 10 mL of LB supplemented with the appropriate antibiotic and incubated overnight at 37 °C, 180 rpm. The resulting culture was centrifuged (4000 rpm, 10 mins). From this cell pellet, the cells were lysed and the plasmid purified using the GeneJET plasmid miniprep kit according to the manufacturer's guidelines. Concentration of plasmid was determined by A_{260} measurements. Plasmid solutions were stored at -20 °C and thawed before use in downstream applications.

6.1.3.4. Agarose gel electrophoresis

1.0 g of agarose was suspended in 100 mL 1 x TBE buffer. The mixture was heated in a microwave until the agarose had dissolved. The solution was allowed to cool to RT and 1 µL GelRed or Midori Green dye was added, if appropriate. The solution was poured into a mould, a comb added and allowed to set. The set gel was placed into an electrophoresis tank, covered with 1 x TBE and the comb removed. DNA samples were prepared by addition of loading dye or Midori Green direct DNA stain (if no stain was added to the gel) to a final concentration of 1 x. The DNA samples were loaded into the gel along with the appropriate size reference ladder and electrophoresed at 120 V for 45-75 mins.

6.1.3.5. Mutagenesis for the generation of mutated pET24a-(+) plasmids

Point mutations, deletion mutations and insertion mutants were generated from the appropriate pET24a-(+) template and Q5 site-directed mutagenesis kit according to the manufacturer's recommended guidelines. Primer and templates used are described in table **6.4**, reaction composition in table **6.5**, and thermal cycling conditions shown in table **6.6**. PCR products were checked by agarose gel electrophoresis, and the sample with the best purity and yield resulting from the gradient PCR was used for blunt end ligation using KLD reaction mixture as per the manufacturer's protocol. *E. coli* TOP10 or NEB5α cells were transformed with the ligation products *via* the methodology described in section 6.1.3.2, plated onto LB agar plated supplemented with kanamycin (50 μg mL^{-1}) and incubated at 37 °C overnight. Single colonies were used to inoculate 10 mL LB supplemented with kanamycin (50 μg mL^{-1}). Plasmids were isolated using a GeneJET plasmid miniprep kit according to the manufacturer's protocol. The identity of all constructs was confirmed *via* Sanger sequencing using a T7 universal forward primer, pET-RP universal reverse primer and internal primers as applicable. Internal sequencing primers used are described in table **6.7**.

Table 6.4 Primer pairs and templates used in this study for site-directed mutagenesis. Underlined nucleotides in primer sequences correspond to changes introduced from the with respect to the template.

Plasmid Generated	Template Plasmid	Primer pairs used to introduce mutation (underlined are non-annealing bases)	Annealing temperature from gradient chosen for ligation (gradient) / °C
pET24a-(+)-AerJ βHDD	pET24a-(+)-AerJ	**F:** 5'-AGCTTGTTCACGCAATAAAAC -3' **R:** 5'-TAATGACTCGAGCACCAC-3'	63 (no gradient)
pET24a-(+)-AerJΔβHDD	pET24a-(+)-AerJ	**F:** 5'-GGATCCCTGGAAGTACAG -3' **R:** 5'-GACAATTTTAGTTCGGAAACAG -3'	60 (no gradient)

pET24a-(+) AerJ(K156A)	pET24a-(+)-AerJ	**F:** 5'-TAATCCCAATGCAGGCAAATGTTG -3' **R:** 5'-CGCTATTTTTTCAAACCTC -3'	58 (58-72)
pET24a-(+)-AerB(βHDD)	pET24a-(+)-AerB(C)	**F:** 5'-TGCACTTCTTAAAAAAGTGAGAATTTC -3' **R:** 5'-TGATGACTCGAGCACCAC -3'	62 (no gradient)
pET24a-(+)-AerB(C)ΔβHDD	pET24a-(+)-AerB(C)	**F:** 5'-GGATCCCTGGAAGTACAG -3' **R:** 5'-ACAACACCATCTAAACCC -3'	60 (no gradient)
pET24a-(+)-AerB(C-A-PCP)	pET24a-(+)-AerB(C-A-PCP-E)	**F:** 5'-TAATCCCTGGTCAATTTC -3' **R:** 5'-CTTTCAGAACTTGATTTTTAAGG -3'	57 (53-63)
pET24a-(+)-AerB(C-A)	pET24a-(+)-AerB(C-A-PCP)	**F:** 5'-TTCTAATTCATTAACATCAATTTCTG -3' **R:** 5'-TAAGGATCCGAATTCGAG -3'	59 (55-65)
pET24a-(+)-AerB(C-A) H222A	pET24a-(+)-AerB(C-A)	**F:** 5'-CGATAATAATCGCATGGAAGACGTTGACTAAAATATAATTTTC -3' **R:** 5'-ATGGTTGGTCAAAAGGCA -3'	59 (no gradient)
pET24a-(+)-AerB(A-PCP)	pET24a-(+)-AerB(C-A-PCP)	**F:** 5'-AGATCCATGGGAACCACG -3' **R:** 5'-TATCCTGATCATTACTGTATTCATC -3'	61 (no gradient)
pET24a-(+)-AerB(C)'	pET24a-(+)-AerB(C-A-PCP)	**F:** 5' - AGACGTTTGAGTTTGATTC - 3' **R:** 5' - TAAGGATCCGAATTCGAG - 3'	57 (no gradient)

pET24a-(+)-AerB(A)	pET24a-(+)-AerB(A-PCP)	**F:** 5'-TTCTAATTCATTAACATTCAATTTCTG -3' **R:** 5'-TAAGGATCCGAATTCGAG -3'	53 (50-65)
pET24a-(+)-AerB(PCP)	pET24a-(+)-AerB(A-PCP)	**F:** 5'-AGATCCATGGGAACCAC G -3' R: 5'-AATTATCTTCTTCCTAGTA CCG -3'	53 (50-65)
pET24a-(+)-AerJ(R23A)	pET24a-(+)-AerJ	**F:** 5'-ATAGCGCAGCGCATCCC CTTCC -3' **R:** 5'-CGTTCTCCTCAAGGTGTT ATG -3'	67 (52-70)
pET24a-(+)-AerJ(R25A)	pET24a-(+)-AerJ	**F:** 5'-GAGAACGATAAGCCAGCT TATCCCCTTC -3' **R:** 5'-CTCAAGGTGTTATGACTC -3'	58 (52-70)
pET24a-(+)-AerJ(R27A)	pET24a-(+)-AerJ	**F:** 5'-TTGAGGAGAAGCATAGC GCAGCTTATC -3' **R:** 5'-GGTGTTATGACTCCCGAT TTAC -3'	67 (52-70)
pET24a-(+)-AerJ(K23A/R25A)	pET24a-(+)-AerJ(R25A)	**F:** 5'-ATAAGCCAGCGCATCCCC TTCC -3' **R:** 5'-CGTTCTCCTCAAGGTGTT ATG -3'	60 (58-72)
pET24a-(+)-AerJ(K23A/R27A)	pET24a-(+)-AerJ(R27A)	**F:** 5'-ATAGCGCAGCGCATCCC CTTCC -3' **R:** 5'-GCTTCTCCTCAAGGTGTT ATG -3'	62 (58-72)
pET24a-(+)-AerJ(R25A/R27A)	pET24a-(+)-AerJ(R25A)	**F:** 5'-TTGAGGAGAAGCATAAGC CAGCTTATCC -3' **R:** 5'-GGTGTTATGACTCCCGAT TTAC -3'	66 (58-72)

pET24a-(+)-(K23A/R25A/R27A)	pET24a-(+)-AerJ(R25A/R27A)	F: 5'-ATAAGCCAGCGCATCCCCTTCC -3' R: 5'-GCTTCTCCTCAAGGTGTTATG -3'	70 (58-72)
pET24a-(+)-AerA(A-KR-PCP)S1352A	pET24a-(+)-AerA(A-KR-PCP)	F: 5'-GCCAGCAAGGCATGTCCTCCT -3' R: 5'-TAGTCAAGTCATCTCTCGTTTAC -3'	64 (52-70)
pET24a-(+)-AerA(A-KR)	pET24a-(+)-AerA(A-KR-PCP)	F: 5'-TGTTGTCATAACTTTACCACTTTTG -3' R: 5'-TGATGACTCGAGCACCAC -3'	58 (52-70)
pET24a-(+)-AerA(PCP) ΔSLiM	pET24a-(+)-AerA(PCP)	F: 5'-TGTTGTCATAACTTTACCACTTTTG -3' R: 5'-TGATGACTCGAGCACCAC -3'	62 (no gradient)
pET24a-(+)-AerA(PCP) Bamb_5917 SLiM	pET24a-(+)-AerA(PCP) ΔSLiM	F: 5'-TTCGTACTCCGATGCTGTTGTCATAACTTTACCACTTTTG -3' R: 5'-GAAGGCGTGATCCGCTGATGACTCGAGCACCAC -3'	58 (52-70)
pET24a-(+)-AerA(PCP) Sven_0512 SLiM	pET24a-(+)-AerA(PCP) ΔSLiM	F: 5'-ATATTCAACTAAGGCTGTTGTCATAACTTTACCACTTTTG -3' R: 5'-GAAGAAGGAGAATTATGATGACTCGAGCACCAC -3'	61 (52-70)
pET24a-(+)-AerA(PCP) ArfA SLiM	pET24a-(+)-AerA(PCP) ΔSLiM	F: 5'-ATTTAATGAGGGAATTGTTGTCATAACTTTACCACTTTT - 3' R: 5' - CGTAAACGTCGTTTATGATGACTCGAGCACCAC - 3'	64 (52-70)

pET151-Sven_0512 (PCP) AerA SLiM	pET151-Sven_0512 (PCP) [123]	**F:** 5' - GAACTAGAGGAGGGGGAGCTGTG - 3' **R:** 5' - TCGGACGCGTCGGACGGGTCACC - 3'	66 (58-72)

Table 6.5 Reaction components used for Q5 mutagenesis PCR.

PCR component	Volume per 25 µL reaction / µL
Q5 Hot Start High-Fidelity Master Mix x2	12.5
Forward primer (10 µM)	1.25
Reverse primer (10 µM)	1.25
Template DNA	10-20 ng DNA (usually 0.5-2 µL)
Autoclaved dH_2O	Remaining volume up to 25 µL

Table 6.6 Thermocycler conditions used for mutagenesis using Q5 site-directed mutagenesis mastermix.

Step		Temperature / °C	Time / s
Initial Denaturation		98	30
x 25 cycles	Denaturation	98	10
	Annealing	See table **6.4**	30
	Extension	72	30 s kb^{-1}
Final Extension		72	120

Table 6.7 Primers used for Sanger sequencing of site-directed mutants and cloned plasmids.

Plasmid Generated	Primer pairs for Sanger Sequencing
All pET24a-(+) constructs	5'- TAATACGACTCACTATAGGG-3' (GATC Biotech/Eurofins T7 primer) 5'- CTAGTTATTGCTCAGCGG-3' (GATC Biotech/Eurofins pET-RP primer)
pET24a-(+)-AerB(C-A-PCP)	5'- GAGGAGTGACTCGTTTAGTCTG -3' 5'- ATGATAGCGACCTATTAACTCCTG -3' 5'- CCCGAATTACCTATACAGTATGCG-3'
pET24a-(+)-AerB(C-A)	5'- ATGATAGCGACCTATTAACTCCTG -3' 5'- CCCGAATTACCTATACAGTATGCG-3'
pET24a-(+)-AerB(A-PCP)	5'- GAGGAGTGACTCGTTTAGTCTG -3'
pET24a-(+)-AerA(S1352A)	5'- TTTTTGATAGCTGTTCAAAGACTCA -3' 5'- GGCAATTTCATCGGTTTCTA -3' 5'- TCCACTACAGTTTCAATTTCATGG -3' 5'- CAAATGATGTTTGTTCCTCGC -3'
pET24a-(+)-AerA(A-KR)	5'- TTTTTGATAGCTGTTCAAAGACTCA -3' 5'- GGCAATTTCATCGGTTTCTA -3' 5'- TCCACTACAGTTTCAATTTCATGG -3' 5'- CAAATGATGTTTGTTCCTCGC -3'
pMLBAD-AerA(A-KR-PCP)	5'- GATTTAATCTGTATCAGG -3' 5'- TCAGACCTAGATGTGAAACGA -3' 5'- TTTCGGTTTGATCCAGAGGT -3' 5'- TCCACTACAGTTTCAATTTCATGG -3' 5'- TTAAAGAACGCAATATCATCCG -3 5'- CCAAGCTCCAGAGGAAACAA -3'

6.1.3.6. Construction of the pMLBAD-AerA(A-KR-PCP) construct

The *aerA* gene was amplified from the pET24a-(+)-AerA(A-KR-PCP) construct with introduction of a 5' KpnI restriction site and 3' XbaI restriction site using the OneTaq Hotstart quick-load polymerase mastermix with the reaction conditions described in table **6.8** and thermocycling conditions described in table **6.9** with forward primer 5'- ATATGGTACCATGAGGACGGAACACAA –

3’ and reverse primer 5’- GCGCTCTAGATCATAATTCTCCTTCTTCATAAT – 3’.

Table 6.8 Reaction components used for OneTaq PCR amplification.

PCR component	Volume per 25 µL reaction / µL
OneTaq Hot Start Master Mix x2	12.5
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Template DNA	15 ng DNA
Autoclaved dH_2O	Up to 25

Table 6.9 Cycling conditions used for OneTaq Hotstart-amplification *aerA* gene.

Step		Temperature / °C	Time / s
Initial Denaturation		98	180
x 25 cycles	Denaturation	98	30
	Annealing	54-68 (product at 58 chosen for ligation)	30
	Extension	72	270
Final Extension		72	120

The resulting PCR product was analysed by agarose gel electrophoresis, the gel band excised and the DNA purified using the GeneJET Gel Extraction Kit according to the manufacturer’s specifications. The purified PCR product (1000 ng, approximately) was digested using KpnI-HF and XbaI-HF restriction enzymes (1.5 µL each), New England Biolabs 10 x CutSmart buffer (5 µL) and dH_2O (up to 50 µL). Empty pMLBAD vector was digested using the same reaction mixture. The reactions were incubated for 37 °C for 3 h. The resulting reactions were analysed by agarose gel electrophoresis, the gel band excised

and the DNA purified using the GeneJET Gel Extraction Kit according to the manufacturer's specifications.

The digested DNA fragment (approximately 120 ng) and linearized vector (approximately 70 ng) were ligated by mixing with T4 DNA ligase (1 µL), T4 DNA ligase buffer (2 µL) and H_2O (up to 20 µL) and incubated at room temperature overnight. The ligation reaction mixture (5 µL) was used to transform an aliquot of chemically-competent *E. coli* TOP10, plated onto LB agar supplemented with trimethoprim (50 µg mL^{-1}) and incubated at 37 °C overnight. Single colonies were picked and used to inoculate 5 mL LB supplemented with trimethoprim (50 µg mL^{-1}) and incubated at 37 °C, 180 rpm overnight. Plasmids were purified using the GeneJET Plasmid Miniprep Kit. Fidelity of the plasmid was confirmed by digestion with KpnI-HF and XbaI-HF as above and positive hits were confirmed by Sanger sequencing.

6.1.4. Protein overproduction, purification and characterisation

6.1.4.1. SDS-PAGE analysis

SDS-PAGE gels were prepared as outlined in table **6.10** and cast between glass plates to a 1 mm thickness. Gels were made with 5 mL of the appropriate percentage resolving gel and 1 mL resolving gel.

The protein samples containing 1 x SDS-PAGE loading dye (15 µL) were loaded into the gel alongside Pageruler Plus molecular weight reference ladder (3 µL). The gel was electrophoresed at 180-200 V for 40-60 min. Gels were stained with Instant Blue protein stain for at least 1 h with rocking (12 rpm) and destained in dH_2O for at least 3 h with rocking (12 rpm).

Table 6.10 Reagents used for preparation of SDS-PAGE gels. Volumes given are per plate of 1.0 mm thickness.

Reagent	Volume for 6% resolving gel / mL	Volume for 10% resolving gel / mL	Volume for 15% resolving gel / mL	Volume for 5% stacking gel / mL
H_2O	2.6	1.9	1.1	0.68
30% acrylamide solution	1.0	1.7	2.5	0.17
1.5 M Tris pH 8.8	1.3	1.3	1.3	-
1.0 M Tris pH 6.8	-	-	-	0.13
10 % SDS	0.05	0.05	0.05	0.01
10 % ammonium persulfate	0.05	0.05	0.05	0.01
N,*N*,*N*',*N*'-tetramethylethylenediamine	0.05	0.05	0.05	0.001

6.1.4.2. Overproduction of Proteins

Chemically competent *E. coli* BL21 (DE3) or *E. coli* C43 (DE3) cells (pET151-Bamb_5915 only) were transformed *via* heat-shock (as described in section 6.1.3.2) with pET vectors containing the appropriate insert. A single colony was used to inoculate a seed culture of 10 mL supplemented with kanamycin (50 µg mL^{-1}) or ampicillin (100 µg mL^{-1}) and incubated at 37 °C, 180 rpm for 3-16 h. 1-4 L sterile LB broth supplemented the appropriate antibiotic was inoculated the appropriate seed culture and incubated in a shaking incubator at 37 °C, 180 rpm until OD_{600} measured approximately 0.6-0.8. Expression was induced by addition of IPTG to a final concentration of 0.5 mM and further incubated at 15 °C, 180 rpm for 16 h.

6.1.4.3. Recombinant protein purification

Cells were harvested by centrifugation (5000 rpm, 15 min, 4 °C) and resuspended in loading buffer (20 mM Tris, 20 mM imidazole, 150 mM NaCl, 10% glycerol, pH 7.5). 0.05 mg mL^{-1} lysozyme from chicken egg white was added and the cells were lysed by sonication (5 cycles, 12 W, 3 s pulses, 2.5 min, 0 °C) or by high-pressure cell disruption. The lysate was clarified by centrifugation (17500 rpm, 45 min, 4 °C) and filtered through a 0.22 µm syringe filter. Proteins were loaded onto a 1 mL or 5 mL HiTrap™ HP affinity column that had been equilibrated with loading buffer. Protein was then eluted by a benchtop preparation with sequential stepwise elution with buffer containing 20 mM (15 CV), 50 mM (5 CV), 100 mM (5 CV), 200 mM (3 CV), 300 mM (3 CV) and 500 mM (5 CV) imidazole. Elution fractions containing protein, as determined by SDS-PAGE gel electrophoresis, were collected, concentrated with a Vivaspin 20 centrifugal concentrator (3, 5, 10, 30 or 50 kDa molecular weight cutoff as appropriate, 4000 rpm, 4 °C) and subsequently buffer exchanged into storage buffer using a PD-10 column. Resulting proteins solutions were further concentrated if necessary, aliquoted, flash frozen and stored at -80 °C until use. Protein concentrations were determined by measured absorbance at 280 nm using a Nanodrop spectrophotometer using calculated extinction coefficients.

For AerB(C-A-PCP-E), AerA(A-KR-PCP) and AerA(A-KR-PCP)S1352A, following Ni-NTA chromatography the protein sample was concentrated to 1 mL. The protein was purified by size-exclusion chromatography using a HiLoad™ 26/600 Superdex™ 200 prep grade column (GE Healthcare) eluting with buffer containing 10 mM HEPES, 150 mM NaCl and 10% glycerol (pH 7.5) over 1.2 column volumes at a flow rate of 1 mL min^{-1}. The protein was introduced *via* a 1 mL loop. Fractions containing protein were determined by SDS-PAGE gel electrophoresis, collected and concentrated with a Vivaspin 20 centrifugal concentrator (50 kDa molecular weight cutoff, 5000 rpm, 4 °C). The protein sample was aliquoted, flash frozen and stored at -80 °C until use.

6.1.4.4. Concentration determination by Nanodrop

A_{280} readings were recorded on the NanoDrop instrument according to the manufacturer's instructions using 2 µL of protein sample in storage buffer. Samples were rigorously mixed to homogeneity before measurements. Measurements were conducted in triplicate and the average taken. For measurements reading over 30 mg mL^{-1}, the sample was diluted with a known dilution factor, the readings retaken and the output scaled by the dilution factor. A_{280} readings were converted to molar concentrations by calculating the concentration in mg mL^{-1} using the Beer-Lambert equation:

$$c\ (M) = \frac{A_{280}}{\varepsilon_{molar}\ (M^{-1}\ cm^{-1}) \times l\ (cm)}$$

Where c = molar concentration of the protein, l = pathlength and ε_{molar} = molar extinction coefficient predicted by the Expasy ProtParam online webserver.[247]

6.1.4.5. Concentration determination by Bradford assay

Pierce™ Coomassie Plus™ Protein assay reagent was used according to the manufacturer's instructions in 1.5 mL cuvettes and the absorbance at 595 nm (A_{595}) was measured on a Varian Cary 50 UV-Visible Spectrophotometer. Bovine Serum Albumin was used as a standard, and the A_{595} measured at 0.2, 0.4, 0.6, 0.8, 1.0, 1.5 and 2.0 mg mL^{-1}. A plot of A_{595} vs concentration in mg mL^{-1} was generated. The A_{595} of the protein of interest at a known dilution factor was measured and the concentration in mg mL^{-1} was calculated from interpolation of the standard curve. Measurements were conducted in triplicate and the average taken.

6.1.4.6. Determination of oligomerisation state by analytical size-exclusion chromatography

Size-exclusion chromatography of AerJ and AerA(PCP) was performed on an AKTA pure FPLC instrument connected to a GE Healthcare Superdex® 200 Increase 10/30 GL column. Phosphate buffer was used as a mobile phase. A protein standard solution was made consisting of 1.5 mg thyroglobulin (Sigma-Aldrich, 669 kDa), 3 µL apoferritin in a 50% glycerol and 0.075 M NaCl solution (Sigma-Aldrich, 443 kDa), 1.5 mg β-amylase (Sigma-Aldrich, 200 kDa), 1.5 mg

bovine serum albumin (VWR International, 66 kDa) and 1.5 mg lysozyme for chicken egg (Sigma-Aldrich, 14.3 kDa) in 250 µL phosphate buffer. AerJ and Aer(PCP) were diluted to 3 mg mL^{-1}, 250 µL, as determined by Nanodrop measurements, with phosphate buffer. A solution of Bluedextrin (Sigma-Aldrich, 1 mg in 250 µL phosphate buffer) was made to identify the void volume of the column. Solutions were centrifuged (10000g, 10 mins, 4 °C), then introduced sequentially to the size-exclusion column *via* a 200 µL loop. Each run was 1.2 CVs and ran at a flow rate of 1 mL min^{-1}. Molecular weights of the analytes were estimated by plotting K_{avg} vs log(molecular weight) for the standards, where $K_{avg} = (v_e-v_0)/(v_T-v_0)$ and v_e = elution volume, v_0 = void volume and v_T = total volume, and interpolating from this plot.

6.1.4.7. Mass spectrometry of protein samples

Purified proteins were diluted to 10-30 µM in HPLC-grade H_2O prior to UHPLC-ESI-Q-TOF-MS analysis on a Bruker MaXis II instrument as described in section 6.1.1. Samples were centrifuged (10000g, 5 mins) prior to analysis. Charge-state deconvolution of mass spectra was performed using the program Bruker DataAnalysis 4.0.

6.1.5. Biophysical Techniques

6.1.5.1. Bio-Layer Interferometry

Bio-Layer Interferometry was conducted in collaboration with Dr. Simone Kosol, who performed streptavidin labelling of the AerA(PCP) and recorded data using a ForteBio Octet instrument.

AerA(PCP) was dissolved in phosphate buffer (50 mM phosphate buffer, pH 7.4, 150 mM NaCl) and biotinylated using EZ-Link™ Sulfo-NHS-LC-Biotinylation cross-linking agent (ThermoFisher Scientific) according to the manufacturer's instructions. After incubation for 30 min at room temperature, non-reacted cross-linking agent was removed on ZEBA spin desalting columns (7 kDa MWCO, ThermoFisher Scientific). Biotinylated AerA was loaded on Streptavidin biosensors (Pall ForteBio) at a concentration of 65 nM. The binding affinities between immobilised biotinylated AerA(PCP) and AerJ, AerJΔbhDD, AerB or AerBΔbhDD were measured on the Octet RED96 system

(Pall ForteBio) in triplicate. The association and dissociation phases were recorded 500 and 3000 seconds, respectively, at concentrations detailed in table **6.11** for each protein. All steps were performed at RT and the sensor data was fitted using GraphPad Prism 9 software. Affinity constants were extracted from steady-state fittings using the equation B = Bmax*X/(X+Kd), where B is the measured response in nm, B_{max} is the maximum response and X corresponds to the analyte protein concentration.

Table 6.11 Concentrations of analyte proteins used in BLI experiments for interaction with immobilised AerA(PCP).

Analyte Protein	Concentrations of protein used for BLI measurements / μM
AerJ	0, 0.73, 1.5, 3.1, 6.3, 12.5, 25.0, 50.0
AerJΔβHDD	0, 0.73, 1.5, 3.1, 6.3, 12.5, 25.0, 50.0
AerB(C)	0, 0.8, 1.7, 2.7, 6.0, 7.8, 16.0, 30.0
AerB(C)ΔβHDD	0, 0.6, 1.0, 2.3, 5.8, 16.5, 32.0, 74.0

6.1.5.2. Carbene Footprinting

Carbene footrprinting experiments were carried out on AerJ, AerA(PCP) and a mixture of both which had been buffer exchanged into 20 mM KPi, 150 mM NaCl, pH 7.4. Reaction mixtures contained 50 μM AerJ and/or 100 μM AerA(PCP) and 10 mM aryl diazirine **91** at a total reaction volume of 20 μL. The mixture was equilibrated at RT for 5 mins. 6 μL aliquots were then taken, placed into crystal clear vials (Fisher Scientific) and flash frozen in liquid N_2. Photolysis was initiated on the frozen samples using the third harmonic of a Nd:YLF laser (Spectra Physics) with repetition frequency 1000 Hz and pulse energy 125 μJ at a wavelength of 349 nm. The samples were irradiated for 10 s. The samples were then sequentially reduced by addition of dithiothreitol (10 mM in 10 mM ammonium bicarbonate), alkylated by addition of iodoacetamide (55 mM in 10 mM ammonium bicarbonate) and digested with trypsin (1:20 protease:protein ratio in 10 mM ammonium bicarbonate) at 37 °C overnight.

The analysis of the tryptic digests was conducted using a Dionex UltiMate 3000 UHPLC system connected to a Zorbax Eclipse Plus C_{18} column (100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis II mass spectrometer. Mobile phases consisted of water (A) and acetonitrile (B), each supplemented with 0.1% formic acid. A gradient of 5% B to 100% B over 30 min was used at a flow rate of 0.2 ml min^{-1}. The mass spectrometer was operated in positive-ion mode with a scan range of 50 – 3,000 *m/z*. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µl at the start of each run.

Fractional modification of peptides was quantitated by extracting the masses of the unlabelled and singly-labelled peptide with a range of ± 0.05 Da. Manual inspection of the mass spectra was carried out to ensure sampling of the correct mass ion. Fractional modification (P) was then calculated from the peak area of unlabelled ($A_{unlabelled}$) and singly-labelled ($A_{labelled}$) peptide using the equation $P = A_{labelled} / (A_{labelled} + A_{unlabelled})$. Differences in the extent of labelling between individual protein and the proteins in complex was considered significant when the P value obtained from a two-tailed unpaired Student's t-test was < 0.05.

6.1.6. Biochemical Assays

6.1.6.1. Preparation of carrier protein-tethered substrates

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 800 µM substrate-*S*-pantetheine (dissolved as a 20 mM stock in DMSO), 100 µM appropriate carrier protein, 10 mM $MgCl_2$, 10 µM CoaE, 3 µM CoaA, 0.5 µM CoaD, 10 µM Sfp and 800 µM ATP at final reaction volumes of 50-800 µL were incubated at room temperature for 3 hours. Loading of pantetheine substrates was confirmed by UHPLC-ESI-Q-TOF-MS analysis prior to downstream reactions.

6.1.6.2. Halogenation of carrier protein-tethered substrates

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 50 µM 4-HPLA-*S*-carrier protein, 2 mM NADH, 200 µM FAD, 10 µM Fre and 10 µM AerJ (or appropriate mutant) at a final reaction volume of 50 µL were incubated at

room temperature for 1 h. The reaction was quenched by addition of 5.5 µL of 10% formic acid solution to a final concentration of 1%. 12.5 µL of the reaction was diluted by addition of 37.5 µL of HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS. Control reactions were carried out by carrying out the reaction above with the volume of AerJ replaced by buffer.

For attempted quenching of the halogenation reaction, the reaction was scaled up to a final volume of 500 µL and incubated at 25 °C. At each time interval, a 45 µL aliquot was taken and added to 5 µL of 10% formic acid. The quenched mixture was stored at 4 °C prior to dilution for UHPLC-ESI-Q-TOF-MS.

6.1.6.3. Attempted condensation of 4-HPLA-*S*-AerA(PCP) with Ile-*S*-NACs by AerB(C)

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 75 µM 4-HPLA-*S*-AerA, 20 µM AerB(C), 1.3 mM L-Ile-*S*-NAC (**62**) or D-*allo*-Ile-*S*-NAC (**63**) (dissolved in buffer) at a final reaction volume of 60 µL were incubated at room temperature for 1 h. Control reactions were performed as above but with the volume of AerB replaced by storage buffer. After incubation, 120 µL MeOH was added to facilitate protein precipitation. The precipitate was pelleted by centrifugation (2 min, 11000 rpm), the supernatant collected and passed through a 3 kDa molecular weight cut-off centrifugal concentrator prior to analysis by LC-MS.

6.1.6.4. Condensation of 4-HPLA-*S*-AerA(PCP) with Ile-*S*-AerB(C-A-PCP-E)

A reaction mixture containing 20 mM Tris (pH 7.5), 100 mM NaCl, 230 µM AerB(C-A-PCP-E), 10 mM $MgCl_2$, 1 mM CoA, 10 µM Sfp at a final volume of 50 µL was incubated at room temperature for 1 h to effect *apo* to *holo* conversion. Following conversion to the *holo* form, 205 µM *holo*-AerB, 500 µM L-isoleucine and 500 µM ATP at a final reaction volume of 45 µL was incubated at room temperature for 4-6 h to effect isoleucine loading by the native adenylation domain. 60 µM Ile-AerB and 60 µM 4-HPLA-*S*-AerA(PCP) at a final volume of 50 µL was then incubated for 1 h at room temperature. The reaction mixture was then diluted with HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS.

For the control reaction, the AerA(PCP) loading reaction was conducting as per the protocol described in section 6.1.6.1 with the volume of AerA(PCP) replaced by storage buffer. The loading reaction was then used at the same volume as in the protocol outlined above.

6.1.6.5. Attempted halogenation-condensation cascade of 4-HPLA-S-AerA(PCP) with AerJ and Ile-*S*-AerB(C-A-PCP-E)

A reaction mixture containing 20 mM Tris (pH 7.5), 100 mM NaCl, 230 µM AerB(C-A-PCP-E), 10 mM $MgCl_2$, 1 mM CoA, 10 µM Sfp at a final volume of 50 µL was incubated at room temperature for 1 h to effect *apo* to *holo* conversion. Following conversion to the *holo* form, 205 µM *holo*-AerB, 500 µM L-isoleucine and 500 µM ATP at a final reaction volume of 45 µL was incubated at room temperature for 4-6 h to effect isoleucine loading by the native adenylation domain. A reaction mixture containing 50 µM Ile-AerB, 50 µM 4-HPLA-*S*-AerA, 50 µM AerJ, 10 µM Fre, 3 mM NADH and 250 µM FAD was incubated for 1 h at room temperature. The reaction mixture was then diluted with HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS.

6.1.6.6. Condensation of 4-HPLA-*S*-AerA(PCP) with L-Ile-AerB(PCP) catalysed by AerB(C-A)

A reaction mixture containing 20 mM Tris (pH 7.5), 100 mM NaCl, 215 µM AerB(PCP), 10 mM $MgCl_2$, 1 mM CoA, 10 µM Sfp at a final volume of 100 µL was incubated at room temperature for 1 h to effect *apo* to *holo* conversion. Following conversion to the *holo* form, 100 µM *holo*-AerB CP, 100 µM AerB(C-A), 500 µM L-isoleucine and 650 µM ATP at a final reaction volume of 150 µL was incubated at room temperature for 3 h. 50 µM 4-HPLA-*S*-AerA, generated as described in section 6.1.6.1, was then introduced to 50 µM Ile-AerB CP at a final reaction volume of 100 µL and incubated for 1 h at room temperature. The reaction mixture was diluted with HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS.

For the control reaction, the AerA(PCP) loading reaction was conducting as per the protocol described in section 6.1.6.1 with the volume of AerA(PCP) replaced by storage buffer. The loading reaction was then used at the same

volume as in the protocol outlined above. Additionally, the entire reaction was conducted with AerA(C-A) replaced by Aer(C-A)H222A.

6.1.6.7. Halogenation and condensation cascade of 4-HPLA-*S*-AerA(PCP) with L-Ile-AerB(PCP), AerB(C-A) and AerJ

L-Ile-AerB(PCP) was prepared as in section 6.1.6.6 A reaction mixture containing 20 mM Tris (pH 7.5), 100 mM NaCl, 25 µM 4-HPLA-AerA, 25 µM Ile-AerB(PCP), 0.5-25 µM AerJ as appropriate, 2 mM NADH, 200 µM FAD and 10 µM Fre was incubated for 1 h at room temperature. The reaction mixture was diluted with HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS.

6.1.6.8. Loading and attempted halogenation of 4-HPLA-Ile-*S*-AerB(PCP)

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 100 µM AerB(PCP), 800 µM 4-HPLA-Ile-*S*-pantetheine (**72**) (one of four diastereomers, dissolved as a 20mM stock in DMSO), 10 µM $MgCl_2$, 10 µM CoaE, 3 µM CoaA, 0.5 µM CoaD, 10 µM Sfp and 800 µM ATP at final reaction volumes of 50 µL were incubated at room temperature for 3 hours. After pantetheine loading had been confirmed by UHPLC-ESI-Q-TOF-MS analysis, reaction mixtures containing 50 µM 4-HPLA-Ile-*S*-AerB(PCP), 10 µM AerJ, 2 mM NADH, 200 µM FAD, and 10 µM Fre at final reaction volumes of 50 µL were incubated for 1 h at room temperature.

6.1.6.9. Timepoint experiments with AerJ and AerJΔβHDD

Halogenations with AerJ or AerJΔβHDD were set up as previously described in section 6.1.6.2 The reactions were incubated at 25 °C. 45 µL aliquot were taken from 0-60 minutes at time intervals of 10 minutes and added to 5 µL 10% formic acid solution in HPLC-grade H_2O to quench the reaction. Quenched reactions were centrifuged and 15 µL of the supernatant was diluted by addition of 35 µL of HPLC-grade H_2O prior to analysis by UHPLC-ESI-Q-TOF-MS.

6.1.6.10. Halogenation quenching and subsequent cleavage of the pantetheine-tethered substrates by methylamine

Halogenations with AerJ or AerJΔβHDD were set up as previously described in section 6.1.6.2. The reactions were incubated at 25 °C. 90 μL aliquots were taken at time intervals and added to 10 μL 70% trichloroacetic acid solution to quench the reaction and precipitate all proteins. Quenched reactions were incubated on ice for 10 mins, then pelleted by centrifugation (10000 g, 10 mins). The pellet was washed with 3 x 500 μL 10% trichloroacetic acid solution and subsequently redissolved in 50 μL 1 M methylamine solution and incubated at 37 °C for 1 h. Excess methylamine was removed *in vacuo* and the resulting solution was lyophilised. The resulting solid was resuspended in 50 μL HPLC-grade H_2O with 0.1 % formic acid. The solution was pelleted by centrifugation (10000 g, 10 mins) prior to analysis by HPLC-ESI-IT-MS.

6.1.6.11. Inhibition of C domain activity by excised βHDDs

A reaction mixture containing 250 μM AerB(PCP), 10 mM $MgCl_2$, 500 μM CoA, 10 μM Sfp at a final volume of 300 μL was incubated at room temperature for 1 h to effect *apo* to *holo* conversion. Following conversion to the *holo* form, 100 μM *holo*-AerB(PCP), 85 μM AerB(C-A), 500 μM L-isoleucine and 2 mM ATP at a final reaction volume of 700 μL was incubated at room temperature for 3 h. 60 μM 4-HPLA-*S*-AerA(PCP), generated as described in section 6.1.6.1, was then introduced to 30 μM L-Ile-*S*-AerB(PCP) and 25 μM AerB(C-A) with AerJ βHDD or AerB βHDD at 0 μM, 125 μM, 250 μM or 500 μM at a final reaction volume of 250 μL and incubated for 2 mins at 25 °C.

After this time, to each reaction was added 45 μL 70% trichloroacetic acid solution to quench and precipitate all proteins. Quenched reactions were incubated on ice for 10 mins, then pelleted by centrifugation (10000 g, 10 mins). The pellet was washed with 3 x 500 μL 10% trichloroacetic acid solution and subsequently redissolved in 250 μL 1 M methylamine solution and incubated at 37 °C for 1 h. Excess methylamine was removed *in vacuo* and the resulting solution was lyophilised. The resulting solid was resuspended in 50 μL HPLC-grade MeOH. The solution was pelleted by centrifugation

(10000 g, 10 mins) and the supernatant filtered through a centrifugal filter fitted with a 0.2 μm PTFE membrane (Thermo Scientific, 10000g, 5 mins) prior to analysis by UHPLC-ESI-IT-MS.

As a control reaction, the reaction with no added βHDD was repeated where the AerA(PCP) loading reaction was conducting as per the protocol described in section 6.1.6.1 with the volume of AerA(PCP) replaced by storage buffer. The loading reaction was then used at the same volume as in the protocol outlined above.

6.1.6.12. Condensation of acetyl-*S*-carrier protein substrates with AHCCA (38) by Bamb_5915

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 125 μM appropriate carrier protein, 1 mM acetyl-CoA, 10 mM $MgCl_2$ and 10 μM Sfp at a final reaction volume of 150 μL were incubated at ambient temperature for 1 h. Then, 110 μM acetyl-*S*-carrier protein was incubated with 20 μM Bamb_5915 and 1 mM AHCCA (**38**) at a final reaction volume of 75 μL for 1 h at ambient temperature. The reactions were then quenched by addition of 75 μL HPLC-grade MeOH to precipitate proteins. The mixture was pelleted by centrifugation (10000g, 10 mins) and the supernatant filtered through a centrifugal filter fitted with a 0.2 μm PTFE membrane (Thermo Scientific, 10000g, 5 mins) prior to analysis by UHPLC-ESI-Q-TOF-MS.

For the control reactions, the condensation was conducted above with the volume of Bamb_5915 replaced by storage buffer. Additionally, the AerA(PCP) loading reaction was conducted as above with the volume of AerA(PCP) replaced by storage buffer. The same volume of loading reaction was then introduced to Bamb_5915 and AHCCA (**38**) and processed as described above.

6.1.6.13. Condensation of 4-HPLA-*S*- and 4-HPPA-*S*-carrier protein substrates with AHCCA (38) or DHCCA (37) by Bamb_5915

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 800 μM 4-HPLA-*S*- (**51**) or 4-HPPA-*S*-pantetheine (**96**) (dissolved as a 20mM stock in

DMSO), 200 μM AerA(PCP) or Bamb_5917(PCP), 10 mM $MgCl_2$, 10 μM CoaE, 3 μM CoaA, 0.5 μM CoaD, 10 μM Sfp and 800 μM ATP at final reaction volumes of 50-800 μL were incubated at room temperature for 3 hours. Loading of pantetheine substrates was confirmed by UHPLC-ESI-Q-TOF-MS analysis.

150 μM 4-HPLA-*S*-PCP or 4-HPPA-*S*-PCP were then incubated with 20 μM Bamb_5915 and 900 μM AHCCA (**38**) or DHCCA (**37**) (as appropriate) at a final reaction volume of 75 μL for 1 h at ambient temperature. The reaction was then quenched by addition of 75 μL HPLC-grade MeOH to precipitate proteins. The mixture was pelleted by centrifugation (10000g, 10 mins) and the supernatant filtered through a centrifugal filter fitted with a 0.2 μm PTFE membrane (Thermo Scientific, 10000g, 5 mins) prior to analysis by UHPLC-ESI-Q-TOF-MS.

For the control reactions, the condensation was conducted above with the volume of Bamb_5915 replaced by storage buffer. Additionally, the AerA(PCP) loading reaction was conducted as above with the volume of AerA(PCP) replaced by storage buffer. The same volume of loading reaction was then introduced to Bamb_5915 and AHCCA (**38**) and processed as described above.

6.1.6.14. Condensation and halogenation of 4-HPPA-*S*-PCP domains with AHCCA by Bamb_5915 and AerJ

Reaction mixtures containing 20 mM Tris (pH 7.5), 100 mM NaCl, 800 μM 4-HPPA-*S*-pantetheine (**96**) (dissolved as a 20mM stock in DMSO), 200 μM AerA(PCP) or Bamb_5917(PCP), 10 mM $MgCl_2$, 10 μM CoaE, 3 μM CoaA, 0.5 μM CoaD, 10 μM Sfp and 800 μM ATP at final reaction volumes of 50-800 μL were incubated at room temperature for 3 hours. Loading of pantetheine substrates was confirmed by UHPLC-ESI-Q-TOF-MS analysis.

150 μM 4-HPPA-*S*-PCP were then incubated with 2 mM NADH, 200 μM FAD, 10 μM Fre, 10 μM AerJ, 20 μM Bamb_5915 and 900 μM AHCCA (**38**) at a final reaction volume of 82.5 μL for 1 h at ambient temperature. The reaction was

then quenched by addition of 82.5 µL HPLC-grade MeOH to precipitate proteins. The mixture was pelleted by centrifugation (10000g, 10 mins) and the supernatant filtered through a centrifugal filter fitted with a 0.2 µm PTFE membrane (Thermo Scientific, 10000g, 5 mins) prior to analysis by UHPLC-ESI-Q-TOF-MS.

For the control reactions, the condensation was conducted above with the volume of Bamb_5915 replaced by storage buffer. Additionally, the AerA(PCP) loading reaction was conducted as above with the volume of AerA(PCP) replaced by storage buffer. The same volume of loading reaction was then introduced to Bamb_5915 and AHCCA (**38**) and processed as described above.

6.1.7. *Burkholderia* Metabolite Production

6.1.7.1. Complementation of *Burkholderia ambifaria* BCC0203 with pMLBAD plasmids

pMLBAD-AerA(A-KR-PCP) was introduced into chemically-competent *E. coli* TOP10 by heat-shock transformation as described in section 6.1.3.2. Transformants were selected on LB agar plates supplemented with trimethoprim (50 µg mL^{-1}) with incubation overnight at 37 °C. Single colonies of transformants were used to inoculate 5 mL LB supplemented with trimethoprim (50 µg mL^{-1}). Simultaneously, a glycerol stock of *E. coli* HB101 w/pRK2013 was used to inoculate 5 mL LB supplemented with kanamycin (50 µg mL^{-1}) The two cultures were incubated 37 °C, 180 rpm overnight. A glycerol stock of *B. ambifaria* AMMD Δ5917 was used to inoculate 5 mL LB and incubated at 30 °C overnight. The two *E. coli* cultures were pelleted (3000 rpm, 4 °C, 15 min), the supernatant discarded and resuspended in 5 mL LB. 100 µL of each of the two *E .coli* cell suspensions and the *Burkholderia* culture were added to a sterile 1.5 mL microcentrifuge tube and gently mixed. A sterile nitrocellulose disk was placed on an LB agar plate. 100 µL of the mixture was pipetted onto the nitrocellulose disk and spread. The plate was incubated at 30 °C overnight. The cellulose disk was placed into a 50 mL falcon tube and cells washed off with 5 mL 0.9% sterile NaCl solution. Transconjugants were

selected by mixing 25 µL cell resuspension with 25 µL LB and plating on LB agar plates supplemented with trimethoprim (200 µg mL^{-1}) and polymyxin (600 U mL^{-1}), with incubation overnight at 37 °C. A single colony of transconjugants were used to inoculate 10 mL LB incubated at 37 °C, 180 rpm. Colony PCR of the culture using OneTaq Hotstart quick-load polymerase mastermix as outlined in section 6.1.7.2.

6.1.7.2. Colony PCR of *Burkholderia* strains

To carry out colony PCR of *Burkholderia ambifaria* BCC0203 to check fidelity of transconjugants, a single colony was used to inoculate 5 mL LB supplemented with trimethoprim (50 µg mL^{-1}) and grown at 30 °C, 180 rpm overnight. The PCR reactions were set up as described in table **6.12** using the forward and reverse primers used for amplification of the *aerA* insert described in section 6.1.3.6. and the thermocycling conditions described in table **6.13**.

The PCR products were analysed by agarose gel electrophoresis and confirmed by comparison to the same PCR using the isolated AerA pMLBAD plasmid as a template (positive control) and the same PCR conducted with an overnight culture of *Burkholderia ambifaria* BCC0203Δ*bamb_5917* (negative control).

Table 6.12 Reaction components used for OneTaq PCR amplification in colony PCR screening of *Burkholderia ambifaria* BCC0203 transconjugants.

PCR component	Volume per 25 µL reaction / µL
OneTaq Hot Start Master Mix x2	12.5
Forward primer (10 µM)	0.5
Reverse primer (10 µM)	0.5
Overnight culture	1
Autoclaved dH_2O	Up to 25

Table 6.13 Cycling conditions used for OneTaq Hotstart-amplification for colony PCR of *Burkholderia* ambifaria BCC0203 transconjugants.

Step		Temperature / °C	Time / s
Initial Denaturation		98	180
x 30 cycles	Denaturation	98	30
	Annealing	62	30
	Extension	72	270
Final Extension		72	120

6.1.7.3. Metabolite profiling of *Burkholderia ambifaria* transconjugants

To establish if the AerA(A-KR-PCP) pMLBAD-complemented *B. ambifaria* BCC0203 Δ*bamb_5917* produced the desired metabolite, single colonies of transconjugants were used to inoculate 5 mL LB supplemented with trimethoprim (50 µg mL^{-1}) and polymyxin (600 U mL^{-1}) and incubated at 37 °C, 180 rpm. The resulting cultures were centrifuged (4000 rpm, 4 °C, 10 min), the supernatant discarded and the cell pellet resuspended in 5 mL 0.9% NaCl. The resulting cell suspension was streaked onto BSM-agar supplemented with trimethoprim (50 µg mL^{-1}) or BSM-agar supplemented with trimethoprim (50 µg mL^{-1}) and filter-sterilised L-arabinose (0.2%, if used) using cotton swabs and grown for 72 h at 30 °C. The cells were scraped off the agar, then the agar was cut into small pieces and extracted using MeCN (15 mL). The extracts were filtered and concentrated *in vacuo*, and further lyophilised if necessary. The resulting extracts were redissolved in HPLC-grade MeOH (100 µL). The supernatants were filtered through a nylon-coated centrifugal filter prior to UHPLC-ESI-Q-TOF-MS analysis.

6.1.8. Computational Methods

6.1.8.1. Multiple sequence alignments

Protein sequences were retrieved from the NCBI or MiBIG database, as appropriate.[239,248] Multiple sequence alignments were generated using the EMBL-EBI webserver using the ClustalOmega algorithm with default parameters.[249] Alignment outputs were visualised and analysed using Jalview version 2.11.[250] Sequence identity was retrieved from the ClustalOmega webserver output.

6.1.8.2. Homology modelling

Homology modelling was conducted using the Phyre2 webserver using default parameters and the output with the highest alignment score assigned by Phyre2 was used (table **6.14**).[215]

For the AerA(PCP) homology model, the Phyre2 output was further energy minimised using the ModRefiner webserver using default parameters.[220]

Table 6.14 Details of homology model output from the Phyre2 webserver.[215]

Modelled protein	Template for homology model output from Phyre2 (PDB accession code)	Sequence coverage	Confidence score from Phyre2 output
AerA(PCP)	TycC Module 5-6 C-PCP didomain[219] (2JGP)	71 %	100 %
AerJ	MalA' C112S/C128S[97] (5WGY)	96 %	99.9%
AerJ βHDD	TubC ^{N}DD[101] (2JUG)	59%	97%
AerB βHDD	TubC ^{N}DD[101] (2JUG)	72%	99.7%

6.1.8.3. Protein structure visualisation and manipulation

Protein structures were visualised using UCSF Chimera 1.14 or UCSF ChimeraX 1.0.[221,251] Coulombic surface potentials were computed using UCSF Chimera's inbuilt tool using default parameters and visualised over a range of -5 to +5 kcal(mol*e)$^{-1}$.

6.1.8.4. Comparison of codon usage

Codon usage of *aerA* (this study) and *bamb_5110-bamb_5933* from the enacyloxin IIa biosynthetic gene cluster from *Burkholderia ambifariaa* AMMD[235] (MiBIG accession code BGC0001094) were counted with the sequence manipulation suite webserver[240] using the standard genetic code DNA codons.

6.2. Synthetic Chemistry Methods

6.2.1. Materials and Equipment

All reported ^{1}H and ^{13}C NMR spectra were acquired by the University of Warwick NMR submission service on a Bruker Avance III HD 500 MHz instrument equipped with a DCH CryoProbe or Bruker Avance III 600 MHz instrument. Chemical shifts (δ) and coupling constants (J) are expressed in ppm and Hz, respectively. Chemical shifts are referenced to the internal standard tetramethylsilane or the residual protonated solvent peak, as appropriate. Coupling constants are reported to the nearest 0.5 Hz. ^{1}H and ^{13}C NMR signals were assigned with the aid of COSY, HSQC and HMBC data. Routine NMR to monitor reaction progress and for NMR of crude samples was recorded by the author on a Bruker Avance 300 MHz or Bruker Avance III HD 400 MHz instrument.

HRMS data were recorded on a Bruker MaXis Plus TOF ESI instrument by the University of Warwick mass spectrometry submission service. LRMS data for monitor reaction progress and for crude samples was conducted by the author on an Agilent 6130B ESI-Quad instrument.

All reagents were purchased from Sigma-Aldrich, Alfa Aesar, Thermo Fisher Scientific, VWR or Acros Organics as appropriate and used with no further purification. Anhydrous solvents were either purchased from Sigma-Aldrich and used as supplied or obtained by distillation over calcium hydride at the University of Warwick and stored over 4 Å molecular sieves. Deuterated solvents for NMR were purchased from Sigma-Aldrich or Cambridge Isotope Laboratories Inc. and used as supplied.

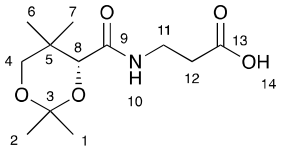
TLC was performed on aluminium plates coated with 0.2 mm silica gel 60 F_{254} purchased from Fisher Scientific. Compound detection on TLC plates was achieved using UV light, potassium permanganate staining or ninhydrin staining. Silica gel flash chromatography was performed using 40-63 particle size silica, 60 Å pore size, purchased from Sigma-Aldrich.

Room temperature refers to ambient temperature (18-25 °C) and 0 °C refers to an ice-water bath.

Anhydrous reactions were conducted in glassware that had been dried in an oven at 200 °C. Glassware was allowed to cool, fitted with rubber septa and evacuated under vacuum. The vessel was then filled with argon.

6.2.2. Synthetic Procedures

6.2.2.1. (*R*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (53)



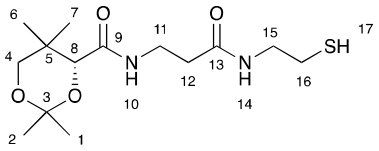
D-pantothenic acid hemicalcium salt (**52**) (5.00 g, 17.3 mmol) and *p*-toluene sulfonic acid (4.79 g, 25.2 mmol) were suspended in acetone. Activated 4 Å molecular sieves were added (10 g) and the reaction stirred at room temperature for 16 h. The resulting slurry was filtered over celite and concentrated *in vacuo*. The resulting crude yellow oil was redissolved in EtOAc (100 mL) and washed with brine (3 x 100 mL). The organic layer was collected, dried over $MgSO_4$ and concentrated *in vacuo* to yield the desired product as a white solid (4.03 g, 89.9%).

HRMS: *m/z* calculated for $C_{12}H_{21}NO_5Na$ $[M+Na]^+$ = 282.1312, found: 282.1309.

^1H NMR (500 MHz, $CDCl_3$) δ: 7.05 (bs, 1H, H-10), 4.12 (s, 1H, H-8), 3.71 (d, 1H, ^2J = 12.0, H-4a), 3.66-3.58 (m, 1H, H-11a), 3.56-3.48 (m, 1H, H-11b), 3.31 (d, 1H, ^2J = 12.0, H-4b), 2.65 (t, 1H, ^3J = 6.5, H-12), 1.49 (s, 3H, H-1/2) 1.45 (s, 3H, H-1/2), 1.06 (s, 3H, H-6/7), 1.00 (s, 3H, H-6/7).

^{13}C NMR (125 MHz, $CDCl_3$) δ: 175.7 (C-13), 170.2 (C-9), 99.3 (C-3), 77.2 (C-8), 71.5 (C-4), 34.1 (C-11), 33.9 (C-12), 33.0 (C-5), 29.4 (C-1/2), 22.0 (C-6/7), 18.9 (C-6/7), 18.7 (C-1/2).

6.2.2.2. (*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (54)



A 250 mL round-bottomed flask was charged with (*R*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (**53**) (3 g, 11.6 mmol) and 1,1'-carbonyldiimidazole (2.72 g, 12.7 mmol). The flask was evacuated and placed under and Ar atmosphere. Anhydrous THF (50 mL) was transferred to the flask *via* cannula. The reaction was stirred at room temperature for 1 h. Cysteamine hydrochloride (**55**) (1.45 g, 12.7 mmol) was dissolved in anhydrous THF (20 mL) under Ar and added to the reaction mixture dropwise over 5 mins. The reaction mixture was stirred at room temperature for 16 h.

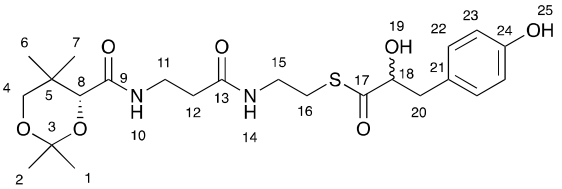
THF was removed in vacuo to yield a crude yellow oil. The oil was redissolved in DCM (100 mL) and washed with sat. NH_4Cl solution (2 x 50 mL) and brine (2 x 50 mL). The organic layer was collected, dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as an off-white amorphous solid. The crude product was purified by silica gel chromatography (100% EtOAc then 5% MeOH in DCM) to yield the desired compound as a white solid (1.70 g, 46.6%).

HRMS: *m/z* calculated for $C_{14}H_{26}N_2O_4SNa$ $[M+Na]^+$ = 341.1505, found: 341.1505.

^1H NMR (500 MHz, $CDCl_3$) δ: 7.01 (bs, 1H, H-10), 6.25 (bs, 1H, H-14), 4.09 (s, 1H, H-8), 3.69 (d, 1H, ^2J = 12.0, H-4a), 3.63-3.50 (m, 2H, H-11), 3.50-3.36 (m, 2H, H-15), 3.28 (d, 1H, ^2J = 12.0, H-4b), 2.69-2.63 (m, 2H, H-16), 2.47 (dt, 2H, ^3J = 6.0, 2.0, H-12), 1.46 (s, 3H, H-1/2), 1.42 (s, 3H, H-1/2), 1.37 (t, 1H, ^3J = 8.5, H-17), 1.04 (s, 3H, H-6/7), 0.97 (s, 3H, H-6/7)

^{13}C NMR (125 MHz, $CDCl_3$) δ: 171.1 (C-13), 170.4 (C-9), 99.2 (C-3), 77.2 (C-8), 71.5 (C-4), 42.5 (C-15), 36.2 (C-12), 34.9 (C-11), 33.0 (C-5), 29.5 (C-1/2), 24.6 (C-16), 22.2 (C-6/7), 19.0 (C-6/7), 18.7 (C-1/2).

6.2.2.3. *S*-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) 2-hydroxy-3-(4-hydroxyphenyl)propanethioate (57)



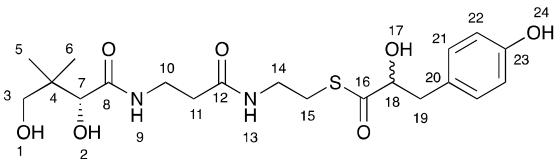
(*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (**54**) (210 mg, 0.66 mmol), PyBOP (286 mg, 0.55 mmol) and DL-p-hydroxyphenyllactic acid (**56**) (100 mg, 0.55 mmol) were dissolved in anhydrous DMF (15 mL) at 0 °C. Diisopropylethylamine (190 µL, 1.10 mmol) was added dropwise at 0 °C. The reaction mixture was slowly warmed to room temperature and stirred for 16 h. The reaction mixture was diluted with EtOAc (50 mL) and washed with sat. NH_4Cl (2 x 50 mL), brine (2 x 50 mL) and 1M LiCl solution (4 x 25 mL). The organic layer was collected, dried over $MgSO_4$ and concentrated in vacuo to yield the crude product as a colourless oil. The crude product was purified by silica gel chromatography with a layer of $CuSO_4$ (5% MeOH in DCM) to yield the desired product as a colourless oil (196 mg, 74.0%).

HRMS: *m/z* calculated for $C_{23}H_{34}N_2O_7SNa$ $[M+Na]^+$ = 505.1979, found: 505.1976.

^{1}H NMR (500 MHz, d_6-acetone) δ: 7.78 (bm, 1H, H-25), 7.09-6.97 (m, 3H, H-10/22), 6.76 (d, 1H, ^3J = 8.5, H-23), 6.73 (d, 1H, ^3J = 7.5, H-), 5.90 (t, 1H, ^3J = 6.0, H-14), 4.35 (bm, 1H, H-18), 4.03 (s, 1H, H-8), 3.62 (d, 1H, ^2J = 11.5, H-4a), 3.49-3.32 (m, 4H, H-11/15a/19), 3.22 (d, 1H, ^2J = 11.5, H-4b), 3.21-3.12 (m, 1H, H-15b), 2.98-2.92 (m, 2H, H-20), 2.89 (t, 2H, , ^3J = 5.5, H-16), 2.35-2.22 (m, 2H, H-12), 1.39 (s, 3H, H-1/2), 1.35 (s, 3H, H-1/2), 0.98 (s, 3H, H-6/7), 0.90 (s, 3H, H-6/7).

^{13}C NMR (125 MHz, d_6-acetone) δ: 200.5 (C-17), 173.6 (C-9), 171.7 (C-13), 155.9 (C-24), 131.0 (C-22), 126.7 (C-21), 115.6 (C-23), 99.2 (C-3), 77.8 (C-18), 77.1 (C-8), 70.9 (C-4), 39.9 (C-20), 38.5 (C-15), 36.1 (C-12), 35.6 (C-11), 33.0 (C-5), 29.4 (C-1/2), 28.3 (C-16), 22.2 (C-6/7), 18.9 (C-6/7), 18.7 (C-1/2).

6.2.2.4. *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) 2-hydroxy-3-(4-hydroxyphenyl)propanethioate (51)



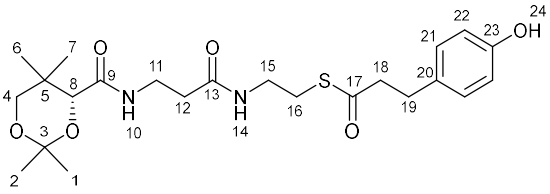
S-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) 2-hydroxy-3-(4-hydroxyphenyl)propanethioate (**57**) (50 mg, 0.10 mmol) was dissolved in 2 mL of acetic acid and 1 mL of H_2O and stirred for 8 h. After full conversion was observed by TLC (10% MeOH in DCM), the solvents were removed *in vacuo* and further dried under high vacuum. The product was then purified by silica gel chromatography (10% MeOH in EtOAc) to yield the desired product as a colourless oil (32 mg, 70.0%).

HRMS: *m/z* calculated for $C_{20}H_{30}N_2O_7SNa$ $[M+Na]^+$ = 465.1666, found: 465.1666.

^{1}H NMR (500 MHz, d_6-acetone) δ: 8.13 (bs, 1H, H-24), 7.52 (bm, 1H, H-9), 7.28 (bm, 1H, H-13), 6.96 (d, 2H, ^3J = 8.5, H-21), 6.61 (d, 2H, ^3J = 8.5, H-22), 5.07 (dd, 1H, ^3J = 6.0, 3.5, H-17), 4.77-4.71 (bm, 1H, H-2), 4.23-4.17 (m, 1H, H-18), 4.16-4.09 (bm, 1H, H-1) 3.81 (s, 1H, H-7), 3.43-3.23 (m, 4H, H-3/10), 3.22-3.14 (m, 2H, H-14), 2.88-2.79 (m, 3H, H-15/19a), 2.63 (dd, 1H, ^3J = 14.0, 8.7, H-19b), 2.29 (t, 2H, ^3J = 6.5, H-11), 0.82 (s, 3H, H-5/6), 0.75 (s, 3H, H-5/6).

^{13}C NMR (125 MHz, d_6-acetone) δ: 204.2 (C-16), 173.4 (C-8), 171.1 (C-12), 156.1 (C-20), 130.6 (C-21), 128.2 (C-23), 114.9 (C-22), 78.6, (C-18), 76.6 (C-7), 69.9 (C-3), 40.0 (C-19), 38.7 (C-19), 35.3 (C-11), 35.1 (C-10), 27.4 (C-15), 21.0 (C-5/6), 19.6 (C-5/6).

6.2.2.5. (*R*)-*S*-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) 3-(4-hydroxyphenyl)propanethioate (95)



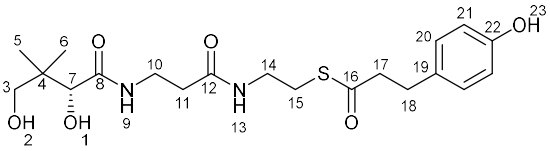
(*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (**54**), 3-(4-hydroxyphenyl)propionic acid (**94**) (0.250 g, 1.50 mmol), HATU (0.572g, 1.50 mmol) were suspended in anhydrous DMF (30 mL) under argon and cooled to 0 °C. DIPEA (0.524 mL, 3.01 mmol) was added dropwise at 0 °C and the resulting mixture was allowed to warm to RT and stirred for 18 h at RT. The solution was diluted with EtOAc (150 mL), washed with H_2O (100 mL), brine (2 x 100 mL), 1M LiCl (2 x 100 mL) and dried over $MgSO_4$. The solvent was removed *in vacuo* to give crude colourless oil which was purified by silica gel chromatography (EtOAc/Pet ether 4:1) to give the desired product (0.228 g, 33%) as a colourless oil.

HRMS: *m/z* calculated for $C_{23}H_{34}N_2O_6SNa$ $[M+Na]^+$ = 489.2035, found: 489.2034.

^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.39 (s, 1H, H-24), 7.09 (t, 1H, ^3J = 7.0, H-10), 7.07 (d, 2H, ^3J = 8.5, H-21), 6.85 (d, 2H, ^3J = 8.5, H-22), 5.44 (t, 1H, ^3J = 5.0, H-14), 4.14 (s, 1H, H-8), 3.72 (d, 1H, ^3J =12.0, H-4a), 3.51 (q, 2H, ^3J = 7.0, H-11), 3.38-3.27 (m, 2H, H-15), 3.32 (d, 1H, ^3J = 12.0, H-4b), 2.96 (t, 2H, ^3J = 7.0, H-19), 2.95 (t, 2H ^3J = 5.5, H-16), 2.87 (t, 2H, ^3J = 7.0, H-18), 2.30 (t, 2H, ^3J = 7.0, 12), 1.49 (s, 3H, H-1/2), 1.46 (s, 3H, H-1/2), 1.11 (s, 3H, H-20/21), 1.01 (s, 3H, H-20/21).

^{13}C NMR (125 MHz, $CDCl_3$) δ: 199.5 (C-17), 171.2 (C-9), 170.5 (C-13), 155.3 (C-23), 130.7 (C-20), 129.9 (C-21), 115.6 (C-22), 99.2 (C-3), 77.2 (C-8), 71.4 (C-4), 45.2 (C-18), 38.5 (C-15), 36.2 (C-12), 35.7 (C-11), 33.1 (C-5), 31.2 (C-19), 29.5 (C-1/2), 29.0 (C-14), 22.2 (C-6/7), 18.9 (C-6/7), 18.7 (C-1/2).

6.2.2.6. (*R*)-*S*-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) 3-(4-hydroxyphenyl)propanethioate (96)



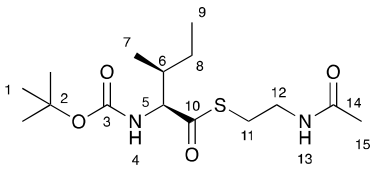
(*R*)-*S*-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) 3-(4-hydroxyphenyl)propanethioate (**95**) (94 mg, 0.20 mmol) was dissolved in 2 mL of acetic acid and 1 mL of H_2O and stirred for 20 h. After full conversion was observed by TLC (5% MeOH in DCM), the solvents were removed *in vacuo* and further dried under high vacuum. The product was then purified by silica gel chromatography (2-5% MeOH in EtOAc) to yield the desired product as a colourless oil (49 mg, 57%).

HRMS: *m/z* calculated for $C_{20}H_{30}N_2O_6SNa$ $[M+Na]^+$ = 449.1722, found: 449.1718.

^{1}H NMR (500 MHz, d_6-acetone) δ: 8.23 (s, 1H, H-23), 7.63 (t, 1H, H-9), 7.39 (t, 1H, ^3J = 6.0, H-13), 7.07 (d, 2H, ^3J = 8.5, H-20), 6.77 (d, 2H, ^3J = 8.5, H-21), 4.82 (d, 1H, ^3J = 4.0, H-1), 4.25-4.17 (m, 1H, H-2), 3.95 (d, 1H, ^3J = H-7), 3.57-3.50 (m, 1H, H10a), 3.49-3.44 (m, 1H, H10b), 3.43-3.40 (m, 2H, H-3), 3.34 (q, 1H, ^3J = 6.5, H-14), 3.00 (t, 2H, ^3J = 7.0, H-15), 2.88-2.83 (m, 4H, H-17/18), 2.41 (t, 2H, ^3J =6.5, H-11), 0.97 (s, 3H, H-5/6), 0.89 (s, 3H, H-5/6).

^{13}C NMR (125 MHz, d_6-acetone) δ: 197.6 (C-16), 173.3 (C-8), 171.0 (C-12), 155.9 (C-22), 130.9 (C-19), 129.3 (C-20), 115.2 (C-21), 76.6 (C-7), 69.9 (C-3), 45.5 (C-17), 39.3 (C-4), 38.7 (C-14), 35.2 (C-11), 35.0 (C-10), 30.3 (C-18), 28.1 (C-15), 21.0 (C-5/6), 19.7 (C-5/6).

6.2.2.7. *S*-(2-acetamidoethyl) (2*S*,3*S*)-2-((*tert*-butoxycarbonyl)amino)-3-methylpentanethioate (66)



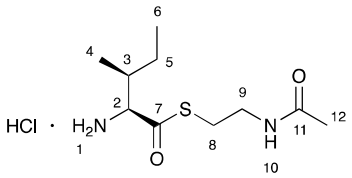
N-Boc-L-isoleucine hemihydrate (**65**) (200 mg, 0.83 mmol) and K_2CO_3 (58 mg, 0.42 mmol) were suspended in THF (10 mL). A solution of 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (160 mg, 0.83 mmol) and 1-hydroxybenzotriazole hydrate (113 mg, 0.83 mmol) in THF (5 mL) was then added dropwise *via* syringe at 0 °C. The reaction mixture was stirred for 15 mins. *N*-acetyl cysteamine (**64**) (100 mg, 0.83 mmol) in THF (2 mL) was then added dropwise *via* syringe at 0 °C. The reaction was allowed to warm to room temperature and stirred for 16 h. THF was then removed *in vacuo* and the resulting crude solid partitioned between EtOAc (20 mL) and H_2O (20 mL). The aqueous layer was extracted with EtOAc (3 x 20 mL) and the combined organics were washed with brine (50 mL), sat. NH_4Cl solution (50 mL) and sat. $NaHCO_3$ solution (50 mL). The organic layer was collected, dried over $MgSO_4$ and concentrated *in vacuo* to give a crude colourless oil. The crude product was purified by silica gel chromatography (100% EtOAc) to yield the desired product as a colourless oil (131 mg, 47.5%).

HRMS: *m/z* calculated for $C_{15}H_{28}N_2O_4SNa$ $[M+Na]^+$ = 355.1662, found: 355.1665.

^1H NMR (500 MHz, d_4-methanol) δ: 4.15-4.08 (m, 1H, H-5), 3.35-3.28 (m, 2H, H-11), 3.06-2.95 (m, 2H, H-12), 1.99-1.92 (m, 1H, H-6), 1.93 (s, 3H, H-15), 1.48 (s, 9H, H-1), 1.48-1.43 (m, 1H, H-8a), 1.27-1.16 (m, 1H, H-8b), 0.95 (d, 3H, ^3J = 6.8, H-7), 0.90 (t, 3H, ^3J = 7.3, H-9).

^{13}C NMR (125 MHz, d_4-methanol) δ: 201.5 (C-10), 172.0 (C-14), 156.9 (C-3), 65.4 (C-5), 38.7 (C-11), 36.8 (C-6), 27.5 (C-12), 27.4 (C-1), 21.2 (C-15), 14.8 (C-7), 10.4 (C-9).

6.2.2.8. *S*-(2-acetamidoethyl) (2*S*,3*S*)-2-amino-3-methylpentanethioate hydrochloride salt (62)



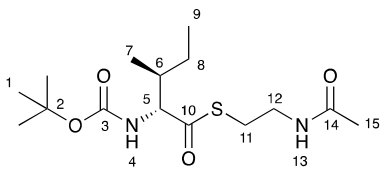
S-(2-acetamidoethyl) (2*S*,3*S*)-2-((*tert*-butoxycarbonyl)amino)-3-methylpentanethioate (**66**) (65 mg, 0.19 mmol) was dissolved in 1,4-dioxane (900 µL) and 4 M HCl in 1,4-dioxane (100 µL) was added. The reaction was stirred at room temperature for 16 h. A further portion of 4 M HCl in 1,4-dioxane (200 µL) was added and the reaction stirred at room temperature for 48 h. Solvents were removed *in vacuo* to yield the desired compound as a white solid (50 mg, 98%).

HRMS: *m/z* calculated for $C_{10}H_{21}N_2O_2S$ $[M+H]^+$ = 233.1318, found: 233.1317.

^{1}H NMR (500 MHz, d_4-methanol) δ: 4.21 (d, 1H, ^3J = 4.5, H-2), 3.42 (t, 2H, ^3J = 6.5, H-9), 3.26-3.14 (m, 2H, H-8), 2.11-2.02 (m, 1H, H-3), 1.95 (s, 3H, H-12), 1.62-1.52 (m, 1H, H-5a), 1.38-1.28 (m, 1H, H-5b), 1.09 (d, 3H, ^3J = 7.0, H-4), 1.03 (t, 3H, ^3J = 7.5, H-6).

^{13}C NMR (125 MHz, d_4-methanol) δ: 195.5 (C-7), 172.1 (C-11), 63.5 (C-2), 38.7 (C-9), 36.9 (C-3), 28.4 (C-8), 24.4 (C-5), 21.1 (C-12), 13.7 (C-4), 10.6 (C-6).

6.2.2.9. *S*-(2-acetamidoethyl) (2R,3S)-2-((tert-butoxycarbonyl)amino)-3-methylpentanethioate (69)



D-allo-isoleucine (**67**) (250 mg, 1.90 mmol), di-tert-butyl decarbonate (458 mg, 2.09 mmol) and sodium bicarbonate (240 mg, 2.85 mmol) were dissolved in acetone (5 mL) and water (5 mL). The reaction mixture was stirred at room temperature for 16 h. Acetone was removed *in vacuo* and the remaining aqueous phase was acidified to pH 4 by careful addition of 1 M HCl solution.

The aqueous phase extracted with EtOAc (3 x 10 mL). The organic layers were collected, washed with brine (10 mL), dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude *N*-Boc-protected amino acid (**68**).

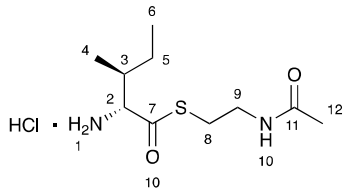
The *N*-Boc-protected amino acid (**68**) (200 mg), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (160 mg, 0.83 mmol), 1-hydroxybenzotriazole hydrate (113 mg, 0.83 mmol) and K_2CO_3 (58 mg, 0.42 mmol) were suspended in THF (10 mL). A solution of *N*-acetyl cysteamine (**64**) in THF (2 mL) was then added dropwise *via* syringe at 0 °C. The reaction mixture was allowed to warm to RT and stirred for 16 h. The solvents were removed *in vacuo* to yield a crude white solid. The crude mixture was purified by silica gel chromatography (100% EtOAc) to yield the desired product as a colourless oil (92 mg, 15% over 2 steps).

HRMS: *m/z* calculated for $C_{15}H_{28}N_2O_4SNa$ $[M+Na]^+$ = 355.1662, found: 355.1665.

^{1}H NMR (500 MHz, $CDCl_3$) δ: 5.99-5.80 (m, 1H, H-13) ,4.96 (d, 1H, ^3J = 9.0, H-4), 4.43 (dd, 1H, ^3J = 9.0 and 3.5, H-5), 3.51-3.38 (m, 2H, H-11), 3.06 (t, 2H, ^3J = 6.5, H-12), 2.12-2.04 (m, 1H, H-6), 1.98 (s, 3H, H-15), 1.48 (s, 9H, H-1), 1.46-1.38 (m, 1H, H-8a), 1.36-1.24 (m, 1H, H-8b), 0.96 (t, 3H, ^3J = 7.5 Hz, H-9), 0.85 (d, 3H, ^3J = 7.0 Hz, H-7).

^{13}C NMR (125 MHz, d_4-methanol) δ: 202.5 (C-10), 170.3 (C-14) 155.7 (C-3), 80.4 (C-2), 63.7 (C-5), 39.4 (C-11), 37.3 (C-6), 28.4 (C-12), 28.3 (C-1), 26.5 (C-8), 23.2 (C-15), 14.0 (C-7), 11.6 (C-9).

6.2.2.10. *S*-(2-acetamidoethyl) (2R,3S)-2-amino-3-methylpentanethioate hydrochloride salt (63)



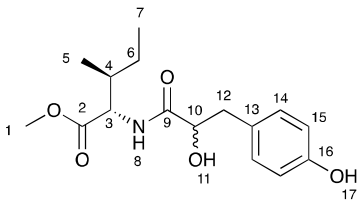
S-(2-acetamidoethyl) (2R,3S)-2-((tert-butoxycarbonyl)amino)-3-methylpentanethioate (**69**) (92 mg, 0.28 mmol) was dissolved in 1,4-dioxane (1 mL) and 4 M HCl in 1,4-dioxane was added (200 µL). The reaction was stirred at room temperature for 48 h. The solvents were removed in vacuo to yield the desired product as a white solid (32 mg, 45%).

HRMS: *m/z* calculated for $C_{10}H_{21}N_2O_2S$ $[M+H]^+$ = 233.1318, found: 233.1320

^{1}H NMR (500 MHz, d_4-methanol) δ: 4.26 (d, 1H, ^3J = 4.0, H-2), 3.42 (t, 2H, ^3J = 6.5, H-9), 3.26-3.13 (m, 2H, H-8), 2.21-2.10 (m, 1H, H-3), 1.96 (s, 3H, H-12), 1.59 (dquint, 1H, ^2J = 14.0, ^3J = 7.5, H-5a), 1.44-1.33 (m, 1H, H-5b), 1.05 (t, 3H, ^3J = 7.5, H-6), 1.01 (d, 3H, ^3J = 7.0, H-4).

^{13}C NMR (125 MHz, d_4-methanol) δ: 196.1 (C-7), 172.2 (C-11), 63.0 (C-2), 38.2 (C-9), 37.0 (C-3), 28.4 (C-8), 25.4 (C-), 21.1 (C-12), 12.3 (C-4), 10.6 (C-6).

6.2.2.11. Methyl (2-hydroxy-3-(4-hydroxyphenyl)propanoyl)-L-isoleucinate (74)



L-Isoleucine methyl ester hydrochloride (**73**) (149 mg, 0.82 mmol), DL-p-hydroxyphenyllactic acid (**56**) (100 mg, 0.55 mmol) and PyBOP (343 mg, 0.66 mmol) were dissolved in anhydrous DMF (10 mL) at 0 °C, under an argon atmosphere. Diisopropylethylamine (238 µL, 1.37 mmol) was added dropwise

at 0 °C. The reaction was allowed to warm to room temperature and stirred for 16 h. DMF was removed *in vacuo* by formation of an azeotrope with *n*-heptane. The resultant crude mixture was dissolved in 20% isopropanol in chloroform (30 mL), washed with H_2O (2 x 10 mL), 1 M HCl (2 x 10 mL) and brine (2 x 10 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as a colourless oil. The crude product was purified by silica gel chromatography (50% EtOAc in petroleum ether) to yield the desired product (**74**) as two separated diastereomers.

Diastereomer 1:

White solid, 38 mg (20.8 %).

HRMS: *m/z* calculated for $C_{16}H_{23}NNaO_5$ $[M+Na]^+$ = 332.1468, found: 332.1465.

^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.07 (d, 2H, ^3J = 8.5, H-14), 6.78 (d, 2H, ^3J = 8.5, H-15), 4.58 (dd, 1H, ^3J = 9.0, 5.5, H-3), 4.27 (dd, 1H, ^3J = 8.5, 4.0, H-10), 3.75 (s, 3H, H-1), 3.13 (dd, 1H, ^2J = 14.0, ^3J = 4.5, H-12a), 2.84 (dd, ^2J = 14.0 , ^3J = 8.5, H-12b), 1.96-1.83 (m, 1H, H-4), 1.49-1.39 (m, 1H, H-6a), 1.22-1.12 (m, 1H, H-6b), 0.96-0.86 (m, 6H, H-6/H-7).

^{13}C NMR (125 MHz, $CDCl_3$) δ: 173.1 (C-9), 173.0 (C-2), 155.2 (C-16), 130.6 (C-14), 128.0 (C-16), 115.8 (C-15), 72.9 (C-10), 56.2 (C-3), 52.3 (C-1), 39.9 (C-12), 37.8 (C-4), 25.1 (C-6), 15.5 (C-5), 11.5 (C-7).

Diastereomer 2:

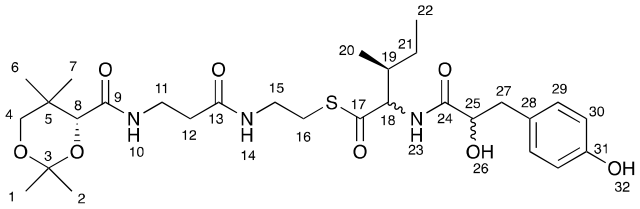
Colourless oil, 43 mg (23.5 %).

HRMS: *m/z* calculated for $C_{16}H_{23}NNaO_5$ $[M+Na]^+$ = 332.1468, found: 332.1471.

^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.17 (d, 1H, ^3J = , H-8), 7.07 (d, 2H, ^3J = 8.5, H-14), 6.77 (d, 2H, ^3J = 8.5, H-15), 4.53 (dd, 1H, ^3J = 9.0, 5.0, H-3), 4.27 (dd, 1H, ^3J = 8.0, 4.0, H-10), 3.74 (s, 3H, H-1), 3.12 (dd, 1H, ^2J = 14.0, ^3J = 4.0, H-12a), 2.84 (dd, ^2J = 14.0 , ^3J = 8.0, H-12b), 1.90-1.80 (m, 1H, H-4), 1.43-1.32 (m, 1H, H-6a), 1.17-1.06 (m, 1H, H-6b), 0.90 (t, 3H, ^3J = 7.5), 0.84 (d, 3H, ^3J = 7.0).

^{13}C NMR (125 MHz, $CDCl_3$) δ: 173.5 (C-9), 172.4 (C-2), 155.2 (C-16), 131.0 (C-14), 128.2 (C-16), 115.6 (C-15), 73.0 (C-10), 56.6 (C-3), 52.3 (C-1), 39.9 (C-12), 37.7 (C-4), 25.1 (C-6), 15.5 (C-5), 11.5 (C-7).

6.2.2.12. *S*-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) (3*S*)-2-(2-hydroxy-3-(4-hydroxyphenyl)propanamido)-3-methylpentanethioate (76)



Diastereomer 1 of methyl (2-hydroxy-3-(4-hydroxyphenyl)propanoyl)-*L*-isoleucinate (**74**) (38 mg, 0.12 mmol) was dissolved in 2 mL of a 2:1 THF:H_2O mixture and lithium hydroxide (10 mg, 0.42 mmol) was added. The mixtures were stirred at room temperature for 4 h. After completion of the reaction was observed by TLC (5% MeOH in DCM), the reaction was acidified to pH 3 by careful addition of 1 M HCl, diluted with H_2O (10 mL) and extracted with EtOAc (4 x 10 mL). The organic layers were dried over $MgSO_4$ and concentrated in vacuo to yield the crude carboxylic acid (**75**) as a colourless oil. The crude products were used in the next step with no further purification.

The crude carboxylic acid (**75**) was dissolved in anhydrous DMF (5 mL). The solution was cooled to 0 °C and (*R*)-*N*-(3-((2-mercaptoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (**54**) (60 mg, 0.19 mmol), PyBOP (90 mg, 0.17 mmol) and diisopropylethylamine (30µL, 0.17 mmol) were added in one charge. The reaction mixture was allowed to warm to room temperature and stirred for 16 h. DMF was removed *in vacuo* by formation of an azeotrope with *n*-heptane. The crude product was purified by silica gel chromatography (5% MeOH in DCM) and diastereomers resulting from racemisation of the amino acid α-carbon were separated by HPLC with a gradient of 50-100% MeOH in H_2O over 30 mins to yield diastereomers 1 and 3 of the product **76**.

The process was repeated for diastereomer 2 of the starting material (**74**) (43 mg, 0.14 mmol) using the same conditions. The resulting diastereomers were separated by HPLC with a gradient of 50-100% MeOH in H_2O over 30 mins to give diastereomers 2 and 4 of the product **76.**

Diastereomer 1:

White solid, 10.3 mg (13.9% over 2 steps)

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 618.2820, found: 618.2821

^{1}H NMR (500 MHz, d_6-acetone) δ: 7.51-7.41 (m, 2H, H-10/H-23), 7.40-7.33 (m, 1H, H-14), 7.11 (d, 2H, ^3J = 8.5, H-29), 6.75 (d, 2H, ^3J = 8.5, H-30), 4.50 (dd, 1H, ^3J = 9.0, 6.0, H-18), 4.31 (dd, 1H, ^3J = 8.0, 4.0, H-25), 4.09 (s, 1H, H-8), 3.72 (d, 1H, ^2J = 11.5, H-4a), 3.55-3.46 (m, 1H, H-11a), 3.45-3.32 (m, 3H, H-11b/H-15), 3.24 (d, 1H, ^2J = 11.5, H-4b), 3.07-2.95 (m, 3H, H-16/H-27a), 2.82 (dd, 1H, ^2J = 14.0, ^3J = 8.0, H-27b), 2.41 (t, 2H, ^3J = 6.5, H-12), 2.04-1.96 (m, 1H, H-19), 1.53-1.44 (m, 1H, H-21a), 1.43 (s, 3H, H-1/2), 1.40 (s, 3H, H-1/2), 1.21-1.10 (m, 1H, H-21b), 1.01 (s, 3H, H-6/7), 0.97 (s, 3H, H-6/7), 0.93 (d, 3H, ^3J = 7.0, H-20), 0.89 (t, 3H, ^3J = 7.5, H-22).

^{13}C NMR (125 MHz, d_6-acetone) δ: 199.5 (C-17), 173.3 (C-24), 171.1 (C-13), 169.2 (C-9), 156.0 (C-31), 130.7 (C-29), 128.6 (C-28), 114.9 (C-30), 98.6 (C-3), 76.9 (C-8), 73.0 (C-25), 70.9 (C-4), 63.1 (C-18), 39.8 (C-27), 38.6 (C-15), 37.2 (C-19), 35.3 (C-12), 34.7 (C-11), 32.6 (C-5), 29.2 (C-1/2), 28.0 (C-16), 24.4 (C-21), 21.5 (C-6/7), 18.5 (C-6/7), 18.2 (C-1/2), 15.2 (C-20), 10.8 (C-22).

Diastereomer 2:

White solid, 5.6 mg (7.5% over 2 steps).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 618.2820, found: 618.2813.

^{1}H NMR (500 MHz, d_6-acetone) δ: 8.13 (s, 1H, H-32), 7.46-7.29 (m, 3H, H-3/H-10/H-14), 7.11 (d, 2H, ^3J = 8.5, H-29), 6.75 (d, 2H, ^3J = 8.5, H-30), 5.02 (d, 1H, ^3J = 6.0, H-26), 4.65 (dd, 1H, ^3J = 9.5, 4.5, H-18), 4.39-4.33 (m, 1H, H-25), 4.07 (s, 1H, H-8), 3.72 (d, 1H, ^2J = 12.0, H-4a), 3.55-3.44 (m, 1H, H-11a), 3.44-3.30 (m, 3H, H-11b/H-15), 3.24 (d, 1H, ^2J = 12.0), 3.09-2.96 (m, 3H, H-16/H-27a),

2.87-2.78 (m, 1H, H-27b), 2.38 (t, 2H, ^3J = 6.5, H-12), 2.06-2.00 (m, 1H, H-19), 1.43 (s, 3H, H-1/2), 1.39 (s, 3H, H-1/2), 1.24-1.14 (m, 1H, H-21a), 1.13-1.03 (m, 1H, H-21b), 1.00 (s, 3H, H-6/7), 0.96 (s, 3H, H-6/7), 0.88 (t, 3H, ^3J = 7.5, H-22), 0.79 (d, 3H, ^3J = 7.0, H-20).

^{13}C NMR (125 MHz, d_6-acetone) δ: 200.2 (C-17), 173.2 (C-24), 170.9 (C-13), 169.1 (C-9), 156.0 (C-31), 130.8 (C-29), 128.5 (C-28), 114.8 (C-30), 98.6 (C-3), 76.9 (C-8), 72.9 (C-25), 70.9 (C-4), 61.2 (C-18), 39.7 (C-27), 38.5 (C-15), 37.3 (C-19), 35.3 (C-12), 34.7 (C-11), 32.6 (C-5), 29.2 (C-1/2), 28.2 (C-16), 26.0 (C-21), 21.5 (C-6/7), 18.4 (C-6/7), 18.1 (C-1/2), 13.5 (C-20), 11.1 (C-22).

Diastereomer 3:

White solid, 8.1 mg (9.3% over 2 steps).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 618.2820, found: 618.2814.

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.73-7.62 (m, 1H, H-10), 7.10 (d, 2H, ^3J = 8.5, H-29), 6.71 (d, 2H, ^3J = 8.5, H-30), 4.62 (d, 1H, ^3J = 4.5, H-18), 4.29 (d, 1H, ^3J = 7.5, 4.0, H-25), 4.13 (s, 1H, H-8), 3.75 (d, 1H, ^3J = 11.5, H-4a), 3.56-3.41 (m, 2H, H-11), 3.38-3.31 (m, 2H, H-15), 3.28 (d, 1H, ^3J = 11.5, H-4b), 3.05-2.97 (m, 3H, H-16/H-27a), 2.82 (dd, 1H, ^2J = 13.0, ^3J = 8.0, H-27b), 2.42 (t, 2H, ^3J = 7.0, H-12), 2.13-2.03 (m, 1H, H-19), 1.46 (s, 3H, H-1/2), 1.45 (s, 3H, H-1/2), 1.37-1.26 (m, 1H, H-21a), 1.25-1.16 (m, 1H, H-21b), 1.01 (s, 3H, H-6/7), 0.99 (s, 3H, H-6/7), 0.93 (t, 3H, ^3J = 7.5, H-22), 0.86 (d, 3H, ^3J = 7.0, H-20).

^{13}C NMR (125 MHz, d_4-methanol) δ: 200.0 (C-17), 175.3 (C-24), 172.5 (C-13), 170.8 (C-9), 155.9 (C-31), 130.3 (C-29), 128.1 (C-28), 114.7 (C-30), 99.0 (C-3), 77.0 (C-8), 72.8 (C-25), 70.9 (C-4), 61.6 (C-18), 39.6 (C-27), 38.5 (C-15), 37.1 (C-19), 34.9 (C-11/12), 34.8 (C-11/12), 32.5 (C-5), 28.3 (C-1/2), 27.7 (C-16), 26.0 (C-21), 21.0 (C-6/7), 18.0 (C-6/7), 17.6 (C-1/2), 13.2 (C-20), 10.5 (C-22)

Diastereomer 4:

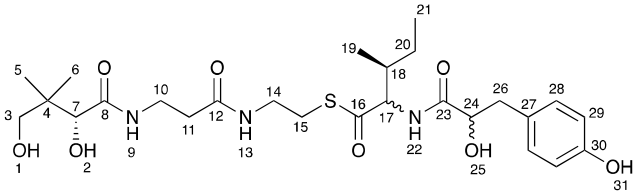
Colourless oil, 21 mg (24.2% over 2 steps).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 618.2820, found: 618.281.

^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.17 (t, 1H, ^3J = 6.5, H-10), 7.11 (d, 2H, ^3J = 8.5, H-29), 6.78 (d, 2H, ^3J = 8.5, H-30), 6.64 (m, 1H, H-13), 4.47 (dd, 1H, ^3J = 8.5, 6.0, H-18), 4.35 (dd, 1H, ^3J = 7.5, 4.0, H-25), 4.09 (s, 1H, H-8), 3.72-3.61 (m, 2H, H-4a/H-15a), 3.50 (dt, 2H, ^3J = 6.5, H-11), 3.27 (d, 1H, ^2J = 11.5, H-4b), 3.25-3.14 (m, 2H, H-16a/15b), 3.10 (dd, 1H, ^2J = 14.0, ^3J = 4.0, H-27a), 2.94-2.79 (m, 2H, H-27b/H16b), 2.47-2.31 (m, 2H, H-12), 1.92-1.82 (m, 2H, H-19), 1.46 (s, 3H, H-1/2), 1.40 (s, 3H, H-1/2), 1.40-1.31 (m, 1H, H-21a), 1.10-1.00 (m, 1H, H-21b), 1.00 (s, 3H, H-6/7), 0.97 (s, 3H, H-6/7), 0.91-0.84 (m, 6H, H-20/22).

^{13}C NMR (125 MHz, $CDCl_3$) δ: 199.5 (C-17), 174.6 (C-24), 171.6 (C-13), 171.0 (C-9), 155.2 (C-31), 130.7 (C-29), 128.4 (C-28), 115.4 (C-30), 99.2 (C-3), 77.1 (C-8), 73.2 (C-25), 71.3 (C-4), 63.6 (C-18), 39.8 (C-27), 39.2 (C-15), 37.1 (C-19), 35.6 (C-12), 35.1 (C-11), 33.0 (C-5), 29.5 (C-1/2), 28.5 (C-16), 24.6 (C-21), 22.2 (C-6/7), 18.9 (C-6/7), 18.7 (C-1/2), 15.7 (C-20), 11.4 (C-22).

6.2.2.13. *S*-(2-(3-((*R*)-2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl) (3*S*)-2-(2-hydroxy-3-(4-hydroxyphenyl)propanamido)-3-methylpentanethioate (72)



S-(2-(3-((*R*)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido)ethyl) (3*S*)-2-(2-hydroxy-3-(4-hydroxyphenyl)propanamido)-3-methylpentanethioate (**76**) diastereomer 1 was dissolved in 2 mL of a 2:1:1 H_2O:Acetic acid:THF mixture and stirred for 72 h. Once all of the starting material had been consumed, as observed by TLC (10% MeOH in DCM), the acetic acid and THF were removed *in vacuo*. H_2O was then removed by lyophilisation to yield the desired product as colourless amorphous solid.

The reaction was repeated for diasteromers 2-4 of **76.** Diastereomer 2 was stirred for 48 and diastereomers 3 and 4 were stirred for 24 h.

Diastereomer 1:

Colourless oil, 9.2 mg (96 %).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 578.2507, found: 578.2508.

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.09 (d, 2H, ^3J = 8.5, H-28), 6.71 (d, 2H, ^3J = 8.5, H-29), 4.43 (d, 1H, ^3J = 6.0, H-17), 4.28 (dd, 1H, ^3J = 8.0, 4.5, H-24), 3.55-3.43 (m, 3H, H-3a/H-11), 3.41 (d, 1H, ^2J = 11.0, H-3b), 3.36-3.31 (m, 2H, H-14), 3.02 (t, 2H, ^3J = 7.0, H-15), 3.00 (dd, 1H, ^2J = 14.0, ^3J = 4.5, H-26a), 2.82 (dd, 1H, ^2J = 14.0, ^3J = 8.0, H-26b), 2.43 (t, 2H, ^3J = 7.0, H-10), 2.01-1.95 (m, 1H, H-18), 1.47- 1.38 (m, 1H, H-20a), 1.18-1.07 (m, 1H, H-20b), 0.94 (s, 6H, C-5/6), 0.92 (d, 3H, ^3J = 7.0, H-19), 0.90 (t, 3H, ^3J = 7.5, H-21).

^{13}C NMR (125 MHz, d_4-methanol) δ: 199.3 (C-16), 175.2 (C-23), 174.7 (C-8), 172.5 (C-12), 155.7 (C-30), 130.3 (C-28), 128.1 (C-27), 114.6 (C-29), 75.9 (C-7), 72.8 (C-24), 69.0 (C-3), 63.3 (C-17), 39.6 (C-4), 39.0 (C-26), 38.5 (C-14), 36.9 (C-18), 35.0 (C-10/11), 34.9 (C-10/11), 27.7 (C-15), 24.3 (C-20), 19.9 (C-5/6), 19.5 (C-5/6), 14.7 (C-19), 10.2 (C-21).

Diastereomer 2:

Colourless oil, 4.2 mg (80 %).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 578.2507, found: 578.2500.

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.09 (d, 2H, ^3J = 8.5, H-28), 6.70 (d, 2H, ^3J = 8.5, H-29), 4.57 (d, 1H, ^3J = 4.0, H-17), 4.35 (dd, 1H, ^3J = 6.5, 4.0, H-24), 3.91 (s, 1H, H-7), 3.55-3.43 (m, 3H, H-3a/H-11), 3.40 (d, 1H, ^2J = 11.0, H-3b), 3.34-3.31 (m, 2H, H-14), 3.06-2.99 (m, 3H, H-15/H-26a), 2.85 (dd, 1H, ^2J = 14.5, ^3J = 7.0, H-26b), 2.41 (t, 2H, ^3J = 6.5, H-10), 2.06-1.98 (m, 1H, H-18), 1.13-0.97 (m, 2H, H-20), 0.94 (s, 6H, H-5/H-6), 0.87 (t, 3H, ^3J = 7.5, H-21), 0.77 (d, ^3J = 7.0, H-19).

^{13}C NMR (125 MHz, d_4-methanol) δ: 200.0 (C-16), 175.0 (C-8/23), 174.7 (C-8/23), 172.5 (C-12), 155.8 (C-30), 130.5 (C-28), 127.8 (C-27), 114.5 (C-29), 75.9 (C-7), 72.4 (C-24), 68.9 (C-3), 61.4 (C-17), 39.2 (C-4/26), 39.0 (C-4/26), 38.5 (C-14), 37.1 (C-18), 35.0 (C-10/11), 34.9 (C-10/11), 27.7 (C-15), 25.8 (C-20), 19.9 (C-5/6), 19.5 (C-5/6), 12.9 (C-19), 10.6 (C-21).

Diastereomer 3:

Colourless oil, 7.4 mg (98 %).

HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 578.2507, found: 578.2508.

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.10 (d, 2H, ^3J = 8.0, H-28), 6.72 (d, 2H, ^3J = 8.0, H-29), 4.62 (d, 1H, ^3J = 4.5, H-17), 4.29 (dd, 1H, ^3J = 8.0, 4.5, H-24), 3.90 (s, 1H, H-7), 3.56-3.43 (m, 3H, H-3a/H-11), 3.41 (d, 1H, ^2J = 11.0, H-3b), 3.38-3.31 (m, 2H, H-14), 3.05-2.97 (m, 3H, H-15, H-26a), 2.83 (dd, 1H, ^2J = 14.0, ^3J = 8.0, H-26b), 2.43 (t, 2H, ^3J = 7.0, H-10), 2.12-2.03 (m, 1H, H-18), 1.37-1.26 (m, 1H, H-20a), 1.25-1.16 (m, 1H, H-20b), 0.94 (s, 6H, H-5/6), 0.92 (t, 3H, ^3J = 7.5, H-21), 0.86 (d, 3H, ^3J = 7.0, H-19).

^{13}C NMR (125 MHz, d_4-methanol) δ: 199.9 (C-16), 175.3 (C-23), 174.7 (C-8), 172.5 (C-12), 155.7 (C-30), 130.3 (C-28), 128.1 (C-27), 114.7 (C-29), 75.9 (C-7), 72.8 (C-24), 69.0 (C-3), 61.6 (C-17), 39.6 (C-26), 39.0 (C-4), 38.5 (C-14), 37.1 (C-18), 35.0 (C-10/11), 34.9 (C-10/11), 27.7 (C-15), 26.0 (C-20), 19.9 (C-5/6), 19.5 (C-5/6), 13.1 (C-19), 10.5 (C-21).

Diastereomer 4:

Colourless oil, 14.1mg (67 %)

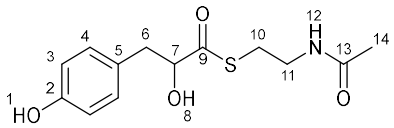
HRMS: *m/z* calculated for $C_{26}H_{41}N_3O_8S$ $[M+Na]^+$ = 578.2507, found: 578.2504.

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.08 (d, 2H, ^3J = 8.5, H-28), 7.00 (d, 2H ^3J = 8.5, H-29), 4.40 (d, 1H, ^3J = 6.0, H-17), 4.32 (dd, 1H, ^3J = 6.5, 4.0, H-24), 3.91 (s, 1H, H-7), 3.55-3.43 (m, 3H, H-3a/H-11), 3.41 (d, 1H, ^2J = 11.0, H-3b), 3.35-3.30 (m, 2H, H-14), 3.05-2.98 (m, 3H, H-15/H-26a), 2.84 (dd, 1H, ^2J = 14.0, ^3J = 6.5, H-26b), 2.41 (t, 2H, ^3J = 6.5, H-10), 1.93-1.83 (m, 1H, H-18),

1.37-1.25 (m, 1H, H-20a), 1.06-0.95 (m, 1H, H-20b), 0.94 (s, 6H, H5/6), 0.86 (t, 3H, ^3J = 7.5, H-21), 0.80 (d, 3H, ^3J = 7.0, H-19).

^{13}C NMR (125 MHz, d_4-methanol) δ: 199.4 (C-16), 175.0 (C-23), 174.7 (C-8), 172.5 (C-12), 155.8 (C-30), 130.5 (C-28), 127.8 (C-27), 114.6 (C-29), 75.9 (C-7), 72.4 (C-24), 69.0 (C-3), 63.1 (C-17), 39.3 (C-4), 39.0 (C-26), 38.5 (C-14), 37.0 (C-18), 35.0 (C-10), 34.9 (C-11), 27.7 (C-15), 24.2 (C-20), 20.0 (C-5/6), 19.6 (C-5/6), 14.6 (C-19), 10.3 (C-21).

6.2.2.14. *S*-(2-acetamidoethyl) 2-hydroxy-3-(4-hydroxyphenyl)propanethioate (86)



DL-4-hydroxyphenyllactic acid (**56**) (60 mg, 0.33 mmol), potassium carbonate (137 mg, 0.99 mmol) and PyBOP (189 mg, 0.36 mmol) were dissolved in anhydrous DMF (5 mL) and cooled to 0 °C. *N*-acetyl cysteamine (**55**) (43 mg, 0.36 mmol) was dissolved in anhydrous DMF (1mL) and added to the reaction dropwise over 5 minutes. The reaction was warmed to RT and stirred for 1 h. The reaction was then diluted with EtOAc (50 mL) and washed with sat. NH_4Cl solution (50 mL), sat. $NaHCO_3$ solution (50 mL) and brine (50 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as a colourless oil. The crude product was then purified by two rounds of silica gel chromatography (5% MeOH in DCM then 100% EtOAc) to yield the desired product as a colourless oil (21 mg, 22%).

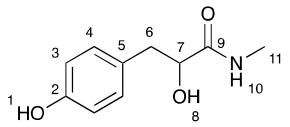
HRMS: *m/z* calculated for $C_{13}H_{17}NO_4SNa$ $[M+Na]^+$ = 306.0776, found: 306.0771

^{1}H NMR (500 MHz, d_6-acetone) δ: 8.17 (s, 1H, H-1), 7.19 (bs, 1H, H-12), 7.10 (d, 2H, ^3J = 8.0, H-4), 6.76 (d, 2H, ^3J = 8.0, H-3), 5.09 (d, 1H, ^3J = 6.5, H-8), 4.34 (ddd, 1H, ^3J = 9.0, 6.5, 4.0), 3.35-3.24 (m, 2H, H-11), 2.99 (dd, 1H, ^2J =

14.0, ^3J = 4.0, H-6a), 2.94 (t, 2H, ^3J = 7.0, H-10), 2.76(dd, 1H, ^2J = 14.0, ^3J = 9.0, H-6b), 1.87 (s, 3H, H-14).

^{13}C NMR (125 MHz, d_6-acetone) δ: 204.1 (C-9), 169.1 (C-13), 156.0 (C-2), 130.6 (C-5), 128.2 (C-4), 114.9 (C-3), 78.5 (C-7), 40.0 (C-6), 38.6 (C-11), 27.4 (C-10), 21.9 (C-14).

6.2.2.15. 2-hydroxy-3-(4-hydroxyphenyl)-*N*-methylpropanamide (110)



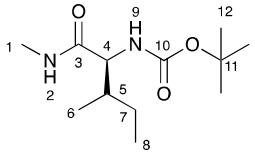
DL-4-hydroxyphenyllactic acid (**56**) (200 mg, 1.10 mmol), 1-hydroxybenzotriazole (145 mg, 1.10 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (315 mg, 1.65 mmol) and methylamine hydrochloride (370 mg , 5.50 mmol) were dissolved in anhydrous DMF (10 mL) under an argon atmosphere and cooled to 0 °C. *N,N*-Diisopropylethylamine (1.15 mL, 6.58 mmol) was added dropwise and the reaction stirred for 16 h with gradual warming to RT. The reaction mixture was diluted by addition of sat. NH_4Cl solution (20 mL), extracted with 10 % isopropanol in chloroform (4 x 20 mL). The organic layer was collected and washed with 1 M HCl solution (40 mL), sat. LiCl solution (40 mL) and brine (40 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product. The crude product was purified by silica gel chromatography (10% DCM in MeOH) to yield the desired product as a white solid (38 mg, 18%).

HRMS: *m/z* calculated for $C_{10}H_{13}NO_3Na$ $[M+Na]^+$ = 218.0788, found: 218.0786.

^{1}H NMR (500 MHz, d_4-methanol) δ: 6.95 (d, 2H, ^3J = 8.5, H-4), 6.58 (d, 2H, ^3J = 8.5, H-3), 4.05 (dd, 1H, ^3J = 7.5 and 4.0, H-7), 2.87 (dd, 1H, ^2J =14.0 and ^3J = 4.0, H-6a), 2.62 (dd, 1H, ^2J = 14.0 and ^3J = 7.5, H-6b), 2.60 (s, 3H, H-11).

^{13}C NMR (125 MHz, d_4-methanol) δ: 175.8 (C-9), 155.6 (C-2), 130.2 (C-4), 128.4 (C-5), 114.8 (C-3), 72.8 (C-7), 39.6 (C-6), 29.4 (C-), 24.5 (C-11).

6.2.2.16. *tert*-butyl ((2*S*,3*S*)-3-methyl-1-(methylamino)-1-oxopentan-2-yl)carbamate (89)



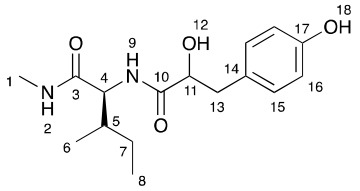
To a stirred solution of L-isoleucine (1 g, 7.62 mmol) and Di-*tert*-butyl dicarbonate (1.74 g, 8.00 mmol) in anhydrous THF (50 mL) under an argon atmosphere at 0 °C was added triethylamine (1.2 mL, 8.38 mmol) dropwise. The reaction was stirred for 48 h with gradual warming to RT. The solvent was removed *in vacuo* and the resulting clear oil redissolved in 1 M HCl solution (50 mL) and EtOAc (50 mL). The phases were separated and the aqueous layer extracted with further EtOAc (2 x 50 mL). The organic layers were combined and washed with brine (2 x 50 mL), dried over $MgSO_4$ and dried *in vacuo* to yield crude intermediate (*tert*-butoxycarbonyl)-L-isoleucine **65** as a clear oil (1.76 g, 92%) which was used with no further purification.

Crude (*tert*-butoxycarbonyl)-L-isoleucine **65** (500 mg, 2.16 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (620 mg, 3.24 mmol), methylamine hydrochloride (175 mg , 2.59 mmol) and 4-Dimethylaminopyridine (264 mg, 2.15 mmol) were dissolved in anhydrous DCM (10 mL) under an argon atmosphere and cooled to 0 °C. *N,N*-Diisopropylethylamine (1.15 mL, 6.58 mmol) was added dropwise and the reaction stirred for 16 h with gradual warming to RT. The reaction mixture was diluted with DCM (20 mL) and washed with 1 M HCl solution (20 mL) and brine (20 mL). The organic layer was dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as a white solid. The crude product was purified by silica gel chromatography (50% EtOAc in petroleum ether 60-80 °C) to yield the desired product (**89**) as a white solid (144 mg, 28%).

HRMS: *m/z* calculated for $C_{12}H_{24}N_2O_3Na$ $[M+Na]^+$ = 267.1679, found: 267.1676.

^{1}H NMR (300 MHz, $CDCl_3$) δ: 6.21-5.97 (m, 1H, H-2), 5.22-4.91 (m, 1H, H-9), 3.92 (dd, 1H, ^3J = 9.0 and 6.5, H-4), 2.83 (d, 3H, ^3J = 5.0, H-1), 2.00-1.78 (m, 1H, H-5), 1.59-1.46 (m, 1H, H-7a), 1.45 (s, 9H, H-12), 1.20-1.03 (m, 1H, H-7b), 0.97-0.87 (m, 6H, H-6/H-8).

6.2.2.17. (2*S*,3*S*)-2-(2-hydroxy-3-(4-hydroxyphenyl)propanamido)-*N*,3-dimethylpentanamide (90)



Tert-butyl ((2*S*,3*S*)-3-methyl-1-(methylamino)-1-oxopentan-2-yl)carbamate (**89**) (144 mg, 0.59 mmol) was dissolved in a solution 2 M HCl in diethyl ether (2.5 mL) under and argon atmosphere and stirred at RT for 16 h. The solvent was removed in vacuo to yield the crude intermediate (2*S*,3*S*)-2-amino-*N*,3-dimethylpentanamide hydrochloride as a white solid (84 mg).

(2*S*,3*S*)-2-amino-*N*,3-dimethylpentanamide hydrochloride (40mg, 0.27 mmol), DL-4-hydroxyphenyllactic acid (**56**) (50 mg, 0.27 mmol) and PyBOP (172 mg, 0.329 mmol) were dissolved in anhydrous DMF (5 mL) under an argon atmosphere and cooled to 0 °C. The reaction was stirred for 16 h with gradual warming to RT. The reaction mixture was diluted by addition of EtOAc (30 mL) and partitioned with saturated LiCl solution (30 mL). The layers were separated and the aqueous layer extracted with further EtOAc (3 x 30 mL). The organic layers were combined, washed with 10 % citric acid solution (30 mL) and brine (60 mL), dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as a brown oil. The crude product was purified by silica gel chromatography (10% MeOH in DCM) and the two resultant diastereomers separated by HPLC with a gradient of 50-100% MeOH in H_2O with 0.1% formic acid over 30 mins to yield the desired products as colourless amorphous solids.

Diastereomer 1:

Colourless oil, 18 mg (5.8 %).

HRMS: *m/z* calculated for $C_{16}H_{24}N_2O_4Na$ $[M+Na]^+$ = 331.1628, found: 331.1629.

^{1}H NMR (600 MHz, d_4-methanol) δ: 6.94 (d, 2H, ^3J = 8.5, H-16), 6.57 (d, 2H, ^3J = 8.5, H-15), 4.14 (dd, 1H, ^3J = 7.0 and 4.5, H-11), 4.02 (d, 1H, ^3J = 7.5, H-4), 2.84 (dd, 1H, ^2J = 14.0, ^3J = 4.5, H-13a), 2.69 (dd, 1H, ^2J = 14.0, ^3J = 7.0, H-13b), 2.59 (s, 3H, H-1), 1.70-1.61 (m, 1H, H-5), 1.32 (dqd, 1H, ^2J = 15.0, ^3J = 7.5 and 3.5, H-7a), 0.95 (dquintd, 1H, ^2J = 15.0, ^3J = 7.5 and 1.5, H-7b), 0.79-0.73 (m, 6H, H-6/H-8).

^{13}C NMR (150 MHz, d_4-methanol) δ: 176.0 (C-10), 173.7 (C-3). 157.1 (C-17), 131.7 (C-16), 129.3 (C-14), 116.0 (C-15), 74.0 (C-11), 58.6 (C-4), 40.8 (C-13), 38.5 (C-5), 26.2 (C-1), 25.9 (C-7), 15.8 (C-6), 11.4 (C-8).

Diastereomer 2:

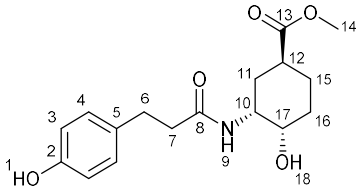
Colourless oil, 14 mg (4.5 %).

HRMS: *m/z* calculated for $C_{16}H_{24}N_2O_4Na$ $[M+Na]^+$ = 331.1628, found: 331.1632.

^{1}H NMR (600 MHz, d_4-methanol) δ: 6.97 (d, 2H, ^3J = 8.5, H-16), 6.5 (d, 2H, ^3J = 8.5, H-15), 4.14 (dd, 1H, ^3J = 6.5 and 4.0, H-11), 4.00 (d, 1H, ^3J = 7.5, H-4), 2.87 (dd, 1H, ^2J = 14.0, ^3J = 4.0, H-13a), 2.70 (dd, 1H, ^2J = 14.0, ^3J = 6.5, H-13b), 2.60 (s, 3H, H-1), 1.58-1.50 (m, 1H, H-5), 1.17 (dqd, 1H, ^2J = 15.0, ^3J = 7.5 and 3.5, H-7a), 0.86-0.77 (m, 1H, H-7b), 0.73 (t, 3H, ^3J = 7.5, H-8), 0.67 (d, 3H, ^3J = 7.0. H-6).

^{13}C NMR (150 MHz, d_4-methanol) δ: 175.8 (C-10), 173.9 (C-3). 157.2 (C-17), 131.9 (C-16), 129.3 (C-14), 115.9 (C-15), 73.9 (C-11), 58.5 (C-4), 40.7 (C-13), 38.6 (C-5), 26.1 (C-1), 25.8 (C-7), 15.7 (C-6), 11.5 (C-8).

6.2.2.18. Methyl (1S,3R,4S)-4-hydroxy-3-(3-(4-hydroxyphenyl)propanamido)-cyclo-hexane-1-carboxylate (101)



(1S,3R,4S)-3-amino-4-hydroxycyclohexane-1-carboxylic acid (**38**) (100 mg, 0.63 mmol) was dissolved in methanol (1 mL) at RT. 4 M HCl in MeOH solution was added dropwise at RT to a final concentration of 2 M then stirred overnight at room temperature. The reaction mixture was concentrated *in vacuo* to yield the crude product (**100**) as a clear gum (110 mg, quantitative yield) and used directly in the next step with no further purification.

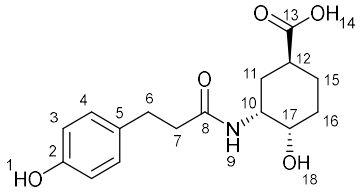
The crude product (**100**) was dissolved in anhydrous DMF (5 mL) and added dropwise to a stirred solution of 3-(4-hydroxyphenyl)propionic acid (**94**) (104 mg, 0.63 mmol), PyBOP (350 mg, 0.68 mmol) and *N*,*N*-diisopropylethylamine (220 µL, 1.24 mmol) at 0 °C. The reaction mixture was allowed to slowly warm to RT and stirred for 16 h. The reaction mixture was diluted with EtOAc (50 mL) and washed with 1 M LiCl solution (4 x 25 mL), 1 M HCl solution (25 mL) and brine (25 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo* to yield the crude product as a light brown oil. The crude product was purified by silica gel chromatography (10% MeOH in DCM) to yield the desired product as a colourless oil (14 mg, 6.9% over 2 steps).

HRMS: *m/z* calculated for $C_{17}H_{23}NO_5Na$ $[M+Na]^+$ = 344.1474, found: 344.1480

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.05 (d, 2H, ^3J = 8.5, H-4), 6.70 (d, 2H, ^3J = 8.5, H-3), 4.10 (dt, 1H, ^3J = 7.0 and 3.5, H-10), 3.76 (dt, 1H, ^3J = 8.0 and 3.5, H-17), 3.69 (s, 3H, H-14), 2.84 (t, 2H, ^3J = 7.5, H-6), 2.56-2.44 (m, 3H, H-7/H-12), 1.99-1.84(m, 2H, H-11a/H-15a), 1.77-1.54 (m, 4H, H-11b/H-16/H-15b).

^{13}C NMR (125 MHz, d_4-methanol) δ: 175.7 (C-13), 174.2 (C-8), 154.4 (C-2), 131.5 (C-5), 129.0 (C-4), 114.8 (C-3), 68.0 (C-17), 50.8 (C-14), 48.9 (C-10), 37.8 (C-7), 37.4(C-12), 30.8 (C-6), 29.3 (C-11), 27.9 (C-16), 23.8 (C-15).

6.2.2.19. (1S,3R,4S)-4-hydroxy-3-(3-(4-hydroxyphenyl)propanamido)cyclohexane-1-carboxylic acid (97)



Methyl (1S,3R,4S)-4-hydroxy-3-(3-(4-hydroxyphenyl)propanamido)cyclohexane-1-carboxylate (**101**) (14 mg, 0.04 mmol) was dissolved in THF (2 mL) and H_2O (1 mL) and lithium hydroxide (2.7 mg, 0.11 mmol) was added in one charge. The reaction mixture was stirred for 6 h at RT. THF was removed *in vacuo* and remaining H_2O was removed by lyophilisation to yield the crude product. The crude product was purified by HPLC with a gradient of 5-100% MeOH in H_2O and appropriate fractions concentrated to yield the desired compound as a colourless oil (5.2 mg, 17 %).

HRMS: *m/z* calculated for $C_{16}H_{21}NO_5Na$ $[M+Na]^+$ = 330.1317, found: 330.1321

^{1}H NMR (500 MHz, d_4-methanol) δ: 7.06 (d, 2H, ^3J = 8.5, H-4), 6.71 (d, 2H ^3J = 8.5, H-3), 4.27-4.21 (m, 1H, H-10), 3.74 (dt, 1H, ^3J = 9.5 and 4.0, H-17), 2.84 (t, 2H, ^3J = 7.5, H-6), 2.53 (t, 2H, ^3J = 7.5, H-7), 2.32-2.23 (m, 1H, H-12), 1.97-1.86 (m, 2H, H-11a/16a), 1.76-1.67 (m, 2H, H-11b/H-15a), 1.62-1.48 (m, 2H, H-15b/16b).

^{13}C NMR (125 MHz, d_4-methanol) δ: 182.5 (C-13), 174.7 (C-10), 155.3 (C-2), 131.8 (C-5), 128.9 (C-4), 114.8 (C-3), 69.7 (C-17), 49.7 (C-10), 40.3 (C-12), 37.9 (C-7), 31.9 (C-11), 30.9 (C-6), 28.3 (C-15), 26.7 (C-16).

References

1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.
2 E. Patridge, P. Gareiss, M. S. Kinch and D. Hoyer, *Drug Discov. Today*, 2016, **21**, 204–207.
3 A. Schatz, E. Bugle and S. A. Waksman, *Proc. Soc. Exp. Biol. Med.*, 1942, **55**, 66-69.
4 A. Schatz and S. A. Waksman, *Exp. Biol. Med.*, 1944, **57**, 244–248.
5 M. J. Weinstein, G. M. Luedemann, E. M. Oden, G. H. Wagman, J. P. Rosselet, J. A. Marquez, C. T. Coniglio, W. Charney, H. L. Herzog and J. Black, *J. Med. Chem.*, 1963, **6**, 463–464.
6 D. P. Levine, 2006, **42**, 5–12.
7 C. L. Ventola, *Pharm. Ther.*, 2015, **40**, 277–283.
8 A. H. Fairlamb, N. A. R. Gow, K. R. Matthews and A. P. Waters, *Nat. Microbiol.*, 2016, **1**, 1–14.
9 D. S. Perlin, R. Rautemaa-Richardson and A. Alastruey-Izquierdo, *Lancet Infect. Dis.*, 2017, **17**, e383–e392.
10 P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
11 R. Subramani and W. Aalbersberg, *Microbiol. Res.*, 2012, **167**, 571–580.
12 S. Agrawal, D. Acharya, A. Adholeya, C. J. Barrow and S. K. Deshmukh, *Front. Pharmacol.*, 2017, **8**, 1–26.
13 Y. Tian, Y. L. Li and F. C. Zhao, *Mar. Drugs*, 2017, **15**, 28.
14 L.-A. Giddings and D. J. Newman, Springer, Cham, Switzerland, 2014.
15 B. A. P. Wilson, C. C. Thornburg, C. J. Henrich, T. Grkovic and B. R. O'Keefe, *Nat. Prod. Rep.*, 2020, **37**, 893–918.
16 C. T. Walsh, *Acc. Chem. Res.*, 2008, **41**, 4–10.
17 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923-936.
18 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493-505.
19 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.
20 M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O.

Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, 362–367.

21 T. Izoré, Y. T. Candace Ho, J. A. Kaczmarski, A. Gavriilidou, K. H. Chow, D. L. Steer, R. J. A. Goode, R. B. Schittenhelm, J. Tailhades, M. Tosin, G. L. Challis, E. H. Krenske, N. Ziemert, C. J. Jackson and M. J. Cryle, *Nat. Commun.*, 2021, **12**, 1–14.

22 T. Stachelhaus, H. D. Mootz, V. Bergendah and M. A. Marahiel, *J. Biol. Chem.*, 1998, **273**, 22773–22781.

23 T. A. Keating, D. E. Ehmann, R. M. Kohli, C. G. Marshall, J. W. Trauger and C. T. Walsh, *ChemBioChem*, 2001, **2**, 99–107.

24 L. Du and L. Lou, *Nat. Prod. Rep.*, 2010, **27**, 255–278.

25 F. Kopp and M. A. Marahiel, *Nat. Prod. Rep.*, 2007, **24**, 735–749.

26 M. Mullowney, R. A. McClure, M. T. Robey, N. L. Kelleher and R. J. Thomson, *Nat. Prod. Rep.*, 2018, **35**, 847–878.

27 K. Zaleta-Rivera, C. Xu, F. Yu, R. A. E. Butchko, R. H. Proctor, M. E. Hidalgo-Lara, A. Raza, P. H. Dussault and L. Du, *Biochemistry*, 2006, **45**, 2561–2569.

28 S. Lin, S. G. Van Lanen and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4183–4188.

29 J. Masschelein, P. K. Sydor, C. Hobson, R. Howe, C. Jones, D. M. Roberts, Z. Ling Yap, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *Nat. Chem.*, 2019, **11**, 906–912.

30 T. Duerfahrt, K. Eppelmann, R. Müller and M. A. Marahiel, *Chem. Biol.*, 2004, **11**, 261–271.

31 D. A. Miller and C. T. Walsh, *Biochemistry*, 2001, **40**, 5313–5321.

32 D. A. Miller, C. T. Walsh and L. Luo, *J. Am. Chem. Soc.*, 2001, **123**, 8434–8435.

33 T. Stachelhaus and C. T. Walsh, *Biochemistry*, 2000, **39**, 5775–5787.

34 H. M. Patel and C. T. Walsh, *Biochemistry*, 2001, **40**, 9023–9031.

35 G. Schoenafinger, N. Schracke, U. Linne and M. A. Marahiel, 2006, **128**, 7406–7407.

36 L. Li, W. Deng, J. Song, W. Ding, Q. F. Zhao, C. Peng, W. W. Song, G. L. Tang and W. Liu, *J. Bacteriol.*, 2008, **190**, 251–263.

37 T. L. Schneider, B. Shen and C. T. Walsh, *Biochemistry*, 2003, **42**, 9722–9730.

38 D. A. Miller, L. Luo, N. Hillson, T. A. Keating and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 333–344.

39 P. C. Schmartz, K. Zerbe, K. Abou-Hadeed and J. A. Robinson, *Org. Biomol. Chem.*, 2014, **12**, 5574–5577.

40 P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13843–13848.

41 P. J. Solenberg, P. Matsushima, D. R. Stack, S. C. Wilkie, R. C. Thompson and R. H. Baltz, *Chem. Biol.*, 1997, **4**, 195–202.

42 A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.

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

44 J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1263–1270.

45 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.

46 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.

47 C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr and D. E. Cane, *Annu. Rev. Biochem.*, 2007, **76**, 195–221.

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

49 R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. Richard Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79.

50 C. R. Valenzano, Y. O. You, A. Garg, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2010, **132**, 14697–14699.

51 S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson and L. Katz, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7119–7123.

52 M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, 2016, **11**, 3319–3327.

53 M. A. Skiba, M. M. Bivins, J. R. Schultz, S. M. Bernard, W. D. Fiers, Q. Dan, S. Kulkarni, P. Wipf, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, 2018, **13**, 3221–3228.

54 D. T. Wagner, Z. Zhang, R. A. Meoded, A. J. Cepeda, A. T. Keatinge-clay, U. States and U. States, *ACS Chem. Biol.*, 2018, **13**, 975–983.

55 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.

56 Y. Q. Cheng, G. L. Tang and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3149–3154.

57 S. B. Bumpus, N. A. Magarvey, N. L. Kelleher, C. T. Walsh and C. T. Calderone, *J. Am. Chem. Soc.*, 2008, **130**, 11614–11616.

58 G. W. Gribble, *Acc. Chem. Res.*, 1998, **31**, 141–152.

59 G. W. Gribble, *J. Chem. Educ.*, 2004, **81**, 1441–1449.

60 C. M. Harris, R. Kannan, H. Kopecka and T. M. Harris, *J. Am. Chem. Soc.*, 1985, **107**, 6652–6658.

61 C. A. Bunders, M. J. Minvielle, R. J. Worthington, M. Ortiz, J. Cavanagh and C. Melander, *J. Am. Chem. Soc.*, 2011, **133**, 20160–20163.

62 E. R. Pereira, L. Belin, M. Sancelme, M. Prudhomme, M. Ollier, M. Rapp, D. Sevère, J. F. Riou, D. Fabbro and T. Meyer, *J. Med. Chem.*, 1996, **39**, 4471–4477.

63 K. H. Van Pée and E. P. Patallo, *Appl. Microbiol. Biotechnol.*, 2006, **70**, 631–641.

64 C. S. Neumann, D. G. Fujimori and C. T. Walsh, *Chem. Biol.*, 2008, **15**, 99–109.

65 D. O'Hagan and H. Deng, *Chem. Rev.*, 2015, **115**, 634–649.

66 D. S. Gkotsi, H. Ludewig, S. V. Sharma, J. A. Connolly, J. Dhaliwal, Y. Wang, W. P. Unsworth, R. J. K. Taylor, M. M. W. McLachlan, S. Shanahan, J. H. Naismith and R. J. M. Goss, *Nat. Chem.*, 2019, **11**, 1091–1097.

67 J. Büchler, A. Papadopoulou and R. Buller, *Catalysts*, 2019, **9**, 1030–1050.

68 J. Latham, E. Brandenburger, S. A. Shepherd, B. R. K. Menon and J. Micklefield, *Chem. Rev.*, 2018, **118**, 232–269.

69 A. E. Fraley and D. H. Sherman, *Bioorganic Med. Chem. Lett.*, 2018, **28**, 1992–1999.

70 A. J. Mitchell, Q. Zhu, A. O. Maggiolo, N. R. Ananth, M. L. Hillwig, X. Liu and A. K. Boal, *Nat. Chem. Biol.*, 2016, **12**, 636–640.

71 M. L. Matthews, C. S. Neumann, L. A. Miles, T. L. Grove, S. J. Booker,

C. Krebs, C. T. Walsh and J. M. Bollinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17723–17728.

72 A. J. Mitchell, N. P. Dunham, J. A. Bergman, B. Wang, Q. Zhu, W. C. Chang, X. Liu and A. K. Boal, *Biochemistry*, 2017, **56**, 441–444.

73 F. H. Vaillancourt, J. Yin and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10111–10116.

74 D. P. Galonić, F. H. Vaillancourt and C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 3900–3901.

75 F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O'Connor and C. T. Walsh, *Nature*, 2005, **436**, 1191–1194.

76 M. Ueki, D. P. Galonić, F. H. Vaillancourt, S. Garneau-Tsodikova, E. Yeh, D. A. Vosburg, F. C. Schroeder, H. Osada and C. T. Walsh, *Chem. Biol.*, 2006, **13**, 1183–1191.

77 W. Jiang, J. R. Heemstra, R. R. Forseth, C. S. Neumann, S. Manaviazar, F. C. Schroeder, K. J. Hale and C. T. Walsh, *Biochemistry*, 2011, **50**, 6063–6072.

78 A. Timmins, N. J. Fowler, J. Warwicker, G. D. Straganz and S. P. de Visser, *Front. Chem.*, 2018, **6**, 513.

79 L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Hkansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, 2009, **459**, 731–735.

80 M. L. Hillwig and X. Liu, *Nat. Chem. Biol.*, 2014, **10**, 921–923.

81 K. H. Van Pée, *Biotechnol. Adv.*, 1990, **8**, 185–205.

82 W. Hemrika, R. Renirie, H. L. Dekker, P. Barnett and R. Wever, *Proc. Natl. Acad. Sci.*, 1997, **94**, 2145–2149.

83 R. Renirie, W. Hemrika and R. Wever, *J. Biol. Chem.*, 2000, **275**, 11650–11657.

84 S. Macedo-Ribeiro, W. Hemrika, R. Renirie, R. Wever and A. Messerschmidt, *J. Biol. Inorg. Chem.*, 1999, **4**, 209–219.

85 W. Hemrika, R. Renirie, S. Macedo-Ribeiro, A. Messerschmidt and R. Wever, *J. Biol. Chem.*, 1999, **274**, 23820–23827.

86 P. Bernhardt, T. Okino, J. M. Winter, A. Miyanaga and B. S. Moore, *J. Am. Chem. Soc.*, 2011, **133**, 4268–4270.

87 S. Diethelm, R. Teufel, L. Kaysser and B. S. Moore, *Angew. Chem.*,

2014, **53**, 11023–11026.

88 S. M. Mantovani and B. S. Moore, *J. Am. Chem. Soc.*, 2013, **135**, 18032–18035.

89 K. S. Ryan, *PLoS One*, 2011, **6**, e23694.

90 B. R. K. Menon, E. Brandenburger, H. H. Sharif, U. Klemstein, S. A. Shepherd, M. F. Greaney and J. Micklefield, *Angew. Chem.*, 2017, **56**, 11841–11845.

91 C. Dong, S. Flecks, S. Unversucht, C. Haupt, K. H. Van Pée and J. H. Naismith, *Science*, 2005, **309**, 2216–2219.

92 K. Podzelinska, R. Latimer, A. Bhattacharya, L. C. Vining, D. L. Zechel and Z. Jia, *J. Mol. Biol.*, 2010, **397**, 316–331.

93 M. C. Andorfer and J. C. Lewis, *Annu. Rev. Biochem.*, 2018, **87**, 159–185.

94 S. Flecks, E. P. Patallo, X. Zhu, A. J. Ernyei, G. Seifert, A. Schneider, C. Dong, J. H. Naismith and K. H. Van Pée, *Angew. Chem.*, 2008, **47**, 9533–9536.

95 C. Schnepel, H. Minges, M. Frese and N. Sewald, *Angew. Chem.*, 2016, **55**, 14159–14163.

96 E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan and C. T. Walsh, *Biochemistry*, 2007, **46**, 1284–1292.

97 A. E. Fraley, M. Garcia-Borràs, A. Tripathi, D. Khare, E. V. Mercado-marin, H. Tran, Q. Dan, G. P. Webb, K. R. Watts, P. Crews, R. Sarpong, R. M. Williams, J. L. Smith, K. N. Houk and D. H. Sherman, *J. Am. Chem. Soc.*, 2017, **139**, 12060–12068.

98 N. Wu, S. Y. Tsuji, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2001, **123**, 6465–6474.

99 R. S. Gokhale and C. Khosla, *Curr. Opin. Chem. Biol.*, 2000, **4**, 22–27.

100 M. Hahn and T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15585–15590.

101 C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nat. Chem. Biol.*, 2008, **4**, 75–81.

102 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340–1351.

103 R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu and E. Al., *Chem. Biol.*, 2003, **10**, 723–731.

104 R. S. Gokhale, S. Y. Tsuji, D. E. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.

105 T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, D. H. Sherman and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.

106 F. Risser, S. Collin, R. Dos Santos-Morais, A. Gruez, B. Chagot and K. J. Weissman, *J. Struct. Biol.*, 2020, **212**, 107581.

107 K. J. Weissman, *ChemBioChem*, 2006, **7**, 485–494.

108 K. J. Weissman, *ChemBioChem*, 2006, **7**, 1334–1342.

109 M. Klaus, M. P. Ostrowski, J. Austerjost, T. Robbins, B. Lowry, D. E. Cane and C. Khosla, *J. Biol. Chem.*, 2016, **291**, 16404–16415.

110 M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, e186.

111 G. Yadav, R. S. Gokhale and D. Mohanty, *PLoS Comput. Biol.*, 2009, **5**, e1000351.

112 T. Li, A. Tripathi, F. Yu, D. H. Sherman and A. Rao, *Bioinformatics*, 2020, **36**, 942–944.

113 W. B. Porterfield, N. Poenateetai and W. Zhang, *iScience*, 2020, **23**, 100938.

114 J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot and K. J. Weissman, *J. Am. Chem. Soc.*, 2016, **138**, 4155–4167.

115 J. Davison, J. Dorival, H. Rabeharindranto, H. Mazon, B. Chagot, A. Gruez and K. J. Weissman, *Chem. Sci.*, 2014, **5**, 3081–3095.

116 J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2016, **11**, 2466–2474.

117 L. Song, M. Jenner, J. Masschelein, C. Jones, M. J. Bull, S. R. Harris, R. C. Hartkoorn, A. Vocat, I. Romero-Canelon, P. Coupland, G. Webster, M. Dunn, R. Weiser, C. Paisey, S. T. Cole, J. Parkhill, E. Mahenthiralingam and G. L. Challis, *J. Am. Chem. Soc.*, 2017, **139**, 7974–7981.

118 M. Jenner, S. Kosol, D. Griffiths, P. Prasongpholchai, L. Manzi, A. S. Barrow, J. E. Moses, N. J. Oldham, J. R. Lewandowski and G. L. Challis, *Nat. Chem. Biol.*, 2018, **14**, 270–275.

119 H. D. Mootz and M. A. Marahiel, *J. Bacteriol.*, 1997, **179**, 6843–6850.

120 A. Tanovic, S. A. Samel, L. O. Essen and M. A. Marahiel, *Science*, 2008, **321**, 659–663.

121 C. Chiocchini, U. Linne and T. Stachelhaus, *Chem. Biol.*, 2006, **13**, 899–908.

122 H. Liu, L. Gao, J. Han, Z. Ma, Z. Lu, C. Dai, C. Zhang and X. Bie, *Front. Microbiol.*, 2016, **7**, 1801.

123 S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de los Santos, L. Manzi, P. K. Sydor, D. Rea, S. Zhou, V. Fülöp, N. J. Oldham, S.-C. Tsai, G. L. Challis and J. R. Lewandowski, *Nat. Chem.*, 2019, **11**, 913–923.

124 A. Sandmann, F. Sasse and R. Müller, *Chem. Biol.*, 2004, **11**, 1071–1079.

125 Y. Chai, D. Pistorius, A. Ullrich, K. J. Weissman, U. Kazmaier and R. Müller, *Chem. Biol.*, 2010, **17**, 296–309.

126 C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode and J. Wöhnert, *Nat. Commun.*, 2018, **9**, 4366.

127 D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh and C. L. Drennan, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 12432–12437.

128 S. A. Sieber, U. Linne, N. J. Hillson, E. Roche, C. T. Walsh and M. A. Marahiel, *Chem. Biol.*, 2002, **9**, 997–1008.

129 B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle, S. Beyer and R. Müller, *J. Biol. Chem.*, 1999, **274**, 37391–37399.

130 Y. Li, K. J. Weissman and R. Müller, *ChemBioChem*, 2010, **11**, 1069–1075.

131 F. Gross, M. W. Ring, O. Perlova, J. Fu, S. Schneider, K. Gerth, S. Kuhlmann, A. F. Stewart, Y. Zhang and R. Müller, *Chem. Biol.*, 2006, **13**, 1253–1264.

132 X. Cai, L. Zhao and H. B. Bode, *Org. Lett.*, 2019, **21**, 2116–2120.

133 M. Hahn and T. Stachelhaus, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 275–280.

134 Y. Inahashi, S. Zhou, M. J. M. J. Bibb, L. Song, M. M. Al-Bassam, M. J. M. J. Bibb and G. L. Challis, *Chem. Sci.*, 2017, **8**, 2823–2831.

135 J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode and J. Wöhnert, *ACS Chem. Biol.*, 2020, **15**, 982–989.

136 S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas and H. B. Bode, *Org. Biomol. Chem.*, 2011, **9**, 3130–3132.

137 K. Ersmark, J. R. Del Valle and S. Hanessian, *Angew. Chem.*, 2008, **47**, 1202–1223.

138 M. Murakami, K. Ishida, T. Okino, Y. Okita, H. Matsuda and K. Yamaguchi, *Tetrahedron Lett.*, 1995, **36**, 2785–2788.

139 K. Ishida, M. Welker, G. Christiansen, S. Cadel-Six, C. Bouchier, E. Dittmann, C. Hertweck and N. T. De Marsac, *Appl. Environ. Microbiol.*, 2009, **75**, 2017–2026.

140 H. Yamaguchi, S. Suzuki, T. Sano, Y. Tanabe, N. Nakajima and M. Kawachi, *Genome Announc.*, 2016, **4**, e01187-16.

141 A. Parthasarathy, P. J. Cross, R. C. J. Dobson, L. E. Adams, M. A. Savka and A. O. Hudson, *Front. Mol. Biosci.*, 2018, **5**, 1–30.

142 K. Ishida, G. Christiansen, W. Y. Yoshida, R. Kurmayer, M. Welker, N. Valls, J. Bonjoch, C. Hertweck, T. Börner, T. Hemscheidt and E. Dittmann, *Chem. Biol.*, 2007, **14**, 565–576.

143 X. Qiu, W. Zhu, W. Wang, H. Jin, P. Zhu, R. Zhuang and X. Yan, *J. Struct. Biol.*, 2019, **205**, 44–52.

144 S. Mahlstedt, E. N. Fielding, B. S. Moore and C. T. Walsh, *Biochemistry*, 2010, **49**, 9021–9023.

145 X. Qiu, Y. Wei, W. Zhu, J. Fu, X. Duan, H. Jin, P. Zhu, C. Zhou and X. Yan, *J. Struct. Biol.*, 2020, **209**, 107415.

146 S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *J. Mol. Biol.*, 1990, **215**, 403–410.

147 O. M. Pérez-Carrascal, Y. Terrat, A. Giani, N. Fortin, C. W. Greer, N. Tromas and B. J. Shapiro, *ISME J.*, 2019, **13**, 2887–2900.

148 A. L. Davidson and J. Chen, *Annu. Rev. Biochem.*, 2004, **73**, 241–268.

149 V. Agarwal, Z. D. Miles, J. M. Winter, A. S. Eustáquio, A. A. El Gamal

and B. S. Moore, *Chem. Rev.*, 2017, **117**, 5619–5674.

150 T. Kittilä, C. Kittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K. H. Van Pee, R. D. Süssmuth, W. Wohlleben, M. J. Cryle and E. Stegmann, *Chem. Sci.*, 2017, **8**, 5992–6004.

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

152 J. B. Patteson, Z. D. Dunn and B. Li, *Angew. Chem.*, 2018, **57**, 6780–6785.

153 B. Pang, B. Pang, B. Pang, Y. Chen, Y. Chen, F. Gan, F. Gan, F. Gan, C. Yan, L. Jin, J. W. Gin, J. W. Gin, C. J. Petzold, C. J. Petzold, J. D. Keasling, J. D. Keasling, J. D. Keasling, J. D. Keasling, J. D. Keasling and J. D. Keasling, *J. Am. Chem. Soc.*, 2020, **142**, 10931–10935.

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

155 E. Frérot, J. Coste, A. Pantaloni, M. N. Dufour and P. Jouin, *Tetrahedron*, 1991, **47**, 259–270.

156 N. Valls, M. López-Canet, M. Vallribera and J. Bonjoch, *Chem. - A Eur. J.*, 2001, **7**, 3446–3460.

157 T. Munde, S. Brand, W. Hidalgo, R. K. Maddula, A. Svatoš and B. Schneider, *Phytochemistry*, 2013, **91**, 165–176.

158 S. Cadel-Six, C. Dauga, A. M. Castets, R. Rippka, C. Bouchier, N. Tandeau De Marsac and M. Welker, *Mol. Biol. Evol.*, 2008, **25**, 2031–2041.

159 M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T. L. Madden, *Nucleic Acids Res.*, 2008, **36**, 5–9.

160 A. Marchler-Bauer, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. Deweese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, S. He, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, C. A. Liebert, C. Liu, F. Lu, S. Lu, G. H. Marchler, M. Mullokandov, J. S. Song, A. Tasneem, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang and S. H. Bryant, *Nucleic Acids Res.*, 2009, **37**, 205–210.

161 F. W. Studier and B. A. Moffatt, *J. Mol. Biol.*, 1986, **189**, 113–130.

162 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.

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

164 J. L. Meier, A. C. Mercer, H. Rivera and M. D. Burkart, *J. Am. Chem. Soc.*, 2006, **128**, 12174–12184.

165 A. L. Mandel, J. J. La Clair and M. D. Burkart, *Org. Lett.*, 2004, **6**, 4801–4803.

166 M. Tosin, D. Spiteller and J. B. Spencer, 2009, 1714–1723.

167 L. E. N. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585–1595.

168 P. K. Mishra and D. G. Drueckhammer, *Chem. Rev.*, 2000, **100**, 3283–3309.

169 E. Eichhorn, J. R. van der Ploeg and T. Leisinger, *J. Biol. Chem.*, 1999, **274**, 26639–26646.

170 S. A. Shepherd, C. Karthikeyan, J. Latham, A. W. Struck, M. L. Thompson, B. R. K. K. Menon, M. Q. Styles, C. Levy, D. Leys and J. Micklefield, *Chem. Sci.*, 2015, **6**, 3454–3460.

171 G. Spyrou, E. Haggard-Ljungquist, M. Krook, H. Jornvall, E. Nilsson and P. Reichard, *J. Bacteriol.*, 1991, **173**, 3673–3679.

172 M. M. E. Huijbers, S. Montersino, A. H. Westphal, D. Tischler and W. J. H. Van Berkel, *Arch. Biochem. Biophys.*, 2014, **544**, 2–17.

173 V. Massey, *J. Biol. Chem.*, 1994, **269**, 22459–22462.

174 G. Eberlein and T. C. Bruice, *J. Am. Chem. Soc.*, 1983, **105**, 6685–6697.

175 R. Hertzberger, J. Arents, H. L. Dekker, R. D. Pridmore, C. Gysler, M. Kleerebezem and M. J. T. de Mattos, *Appl. Environ. Microbiol.*, 2014, **80**, 2229–2239.

176 G. Kim, S. J. Weiss and R. L. Levine, *Biochim. Biophys. Acta - Gen. Subj.*, 2014, **1840**, 901–905.

177 K. G. Reddie and K. S. Carroll, *Curr. Opin. Chem. Biol.*, 2008, **12**, 746–754.

178 D. J. C. Pappin, P. Hojrup and A. J. Bleasby, *Curr. Biol.*, 1993, **3**, 327-332.

179 F. W. Krainer and A. Glieder, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 1611–1625.

180 S. Mori, A. H. Pang, N. Thamban Chandrika, S. Garneau-Tsodikova and O. V. Tsodikov, *Nat. Commun.*, 2019, **10**, 2053.

181 J. Zeng, A. K. Lytle, D. Gage, S. J. Johnson and J. Zhan, *Bioorganic Med. Chem. Lett.*, 2013, **23**, 1001–1003.

182 M. P. Weiner, G. L. Costa, W. Schoettlin, J. Cline, E. Mathur and J. C. Bauer, *Gene*, 1994, **151**, 119–123.

183 K. Bloudoff and T. M. Schmeing, *Biochim. Biophys. Acta - Proteins Proteomics*, 2017, **1865**, 1587–1604.

184 J. Franke and C. Hertweck, *Cell Chem. Biol.*, 2016, **23**, 1179–1192.

185 D. E. Ehmann, J. W. Trauger, T. Stachelhaus and C. T. Walsh, *Chem. Biol.*, 2000, **7**, 765–772.

186 B. O. Bachmann and J. Ravel, in *Methods in Enzymology*, 2009, pp. 181–217.

187 L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and Christopher T. Walsh, *Biochemistry*, 2002, **41**, 9184–9196.

188 R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-vegas, G. A. Salazar, J. Tate and A. Bateman, 2016, **44**, 279–285.

189 L. Dumon-Seignovert, G. Cariot and L. Vuillard, *Protein Expr. Purif.*, 2004, **37**, 203–206.

190 M. M. Bradford, *Anal. Biocehmistry*, 1976, **72**, 248–254.

191 K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C. P. Niesert and H. B. Bode, *Nat. Chem.*, 2018, **10**, 275–281.

192 B. R. Miller, J. A. Sundlov, E. J. Drake, T. A. Makin and A. M. Gulick, *Proteins Struct. Funct. Bioinforma.*, 2014, **82**, 2691-2702.

193 D. T. Jones, *J. Mol. Biol.*, 1999, **292**, 195–202.

194 D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson and D. T. Jones, *Nucleic Acids Res.*, 2013, **41**, 349–357.

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

196 B. Boll, T. Taubitz and L. Heide, *J. Biol. Chem.*, 2011, **286**, 36281-

36290.

197 D. A. Herbst, B. Boll, G. Zocher, T. Stehle and L. Heide, *J. Biol. Chem.*, 2013, **288**, 1991–2003.

198 W. Zhang, J. R. Heemstra, C. T. Walsh and H. J. Imker, *Biochemistry*, 2010, **49**, 9946–9947.

199 B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich and A. M. Gulick, *J. Biol. Chem.*, 2016, **291**, 22559–22571.

200 V. De Crécy-Lagard, P. Marlière and W. Saurin, *Comptes rendus l'Académie des Sci.*, 1995, **318**, 927–936.

201 M. A. Marahiel, T. Stachelhaus and H. D. Mootz, *Chem. Rev.*, 1997, **97**, 2651–2673.

202 Z. J. Anderson, C. Hobson, R. Needley, L. Song, M. S. Perryman, P. Kerby and D. J. Fox, *Org. Biomol. Chem.*, 2017, **15**, 9372–9378.

203 T. R. Hoye, C. S. Jeffrey and F. Shao, *Nat. Protoc.*, 2007, **2**, 2451–2458.

204 H. D. Flack and G. Bernardinelli, *Chirality*, 2008, **20**, 681–690.

205 A. Tooming-Klunderud, T. Rohrlack, K. Shalchian-Tabrizi, T. Kristensen and K. S. Jakobsen, *Microbiology*, 2007, **153**, 1382–1393.

206 K. Fujii, K. Sivonen, T. Nakano and K. I. Harada, *Tetrahedron*, 2002, **58**, 6863–6871.

207 E. Von Elert, L. Oberer, P. Merkel, T. Huhn and J. F. Blom, *J. Nat. Prod.*, 2005, **68**, 1324–1327.

208 M. E. Silva-Stenico, J. Rigonato, M. G. Leal, M. G. M. V. Vaz, A. P. D. Andreote and M. F. Fiore, *Curr. Med. Chem.*, 2012, **19**, 5205–5213.

209 P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13843–13848.

210 V. Agarwal, A. A. El Gamal, K. Yamanaka, D. Poth, R. D. Kersten, M. Schorn, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2014, **10**, 640–647.

211 S. Buedenbender, S. Rachid, R. Müller and G. E. Schulz, *J. Mol. Biol.*, 2009, **385**, 520–530.

212 S. Lin, S. G. Van Lanen and B. Shen, *J. Am. Chem. Soc.*, 2007, **129**, 12432–12438.

213 E. Cilia, R. Pancsa, P. Tompa, T. Lenaerts and W. F. Vranken, *Nat.*

Commun., 2013, **4**, 2741.

214 A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, 725–738.

215 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protoc.*, 2015, **10**, 845–858.

216 A. Sultana and J. E. Lee, *Curr. Protoc. Protein Sci.*, 2015, **79**, 19.25.1-19.25.26.

217 L. Manzi, A. S. Barrow, D. Scott, R. Layfield, T. G. Wright, J. E. Moses and N. J. Oldham, *Nat. Commun.*, 2016, **7**, 1-9.

218 L. Manzi, A. S. Barrow, J. T. S. Hopper, R. Kaminska, C. Kleanthous, C. V. Robinson, J. E. Moses and N. J. Oldham, *Angew. Chem.*, 2017, **56**, 14873–14877.

219 S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel and L. O. Essen, *Structure*, 2007, **15**, 781–792.

220 D. Xu and Y. Zhang, *Biophysj*, 2011, **101**, 2525–2534.

221 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

222 P. Giansanti, L. Tsiatsiani, T. Y. Low and A. J. R. Heck, *Nat. Protoc.*, 2016, **11**, 993–1006.

223 A. Biffis, P. Centomo, A. Del Zotto and M. Zecca, *Chem. Rev.*, 2018, **118**, 2249-2295.

224 C. Hobson, PhD Thesis, University of Warwick, 2019.

225 *Br. Pat.*, GB2018051058, 2018.

226 S. Costa, A. Almeida, A. Castro and L. Domingues, *Front. Microbiol.*, 2014, **5**, 1-20.

227 M. D. Lefebre and M. A. Valvano, *Appl. Environ. Microbiol.*, 2002, **68**, 5956–5964.

228 L. M. Guzman, D. S. Weiss and J. Beckwith, *J. Bacteriol.*, 1995, **177**, 4121–4130.

229 P. Huovinen, *Antimicrob. Agents Chemother.*, 1987, **31**, 1451–1456.

230 G. Ditta, S. Stanfield, D. Corbin and D. R. Helinski, *Proc. Natl. Acad. Sci. U. S. A.*, 1980, **77**, 7347–7351.

231 D. H. Figurski and D. R. Helinski, *Proc. Nati. Acad. Sc*, 1979, **76**, 1648–1652.

232 S. A. Loutet and M. A. Valvano, *Front. Microbiol.*, 2011, **2**, 1-8.

233 W. A. Hareland, R. L. Crawford, P. J. Chapman and S. Dagley, *J. Bacteriol.*, 1975, **121**, 272-285.

234 L. A. O'Sullivan, A. J. Weightman, T. H. Jones, A. M. Marchbank, J. M. Tiedje and E. Mahenthiralingam, *Environ. Microbiol.*, 2007, **9**, 1017-1034.

235 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.

236 A. J. Mullins, J. A. H. Murray, M. J. Bull, M. Jenner, C. Jones, G. Webster, A. E. Green, D. R. Neill, T. R. Connor, J. Parkhill, G. L. Challis and E. Mahenthiralingam, *Nat. Microbiol.*, 2019, **4**, 996-1005.

237 C. Gustafsson, S. Govindarajan and J. Minshull, *Trends Biotechnol.*, 2004, **22**, 346-353.

238 J. B. Plotkin and G. Kudla, *Nat. Publ. Gr.*, 2011, **12**, 32-42.

239 S. A. Kautsar, K. Blin, S. Shaw, J. C. Navarro-Muñoz, B. R. Terlouw, J. J. J. Van Der Hooft, J. A. Van Santen, V. Tracanna, H. G. Suarez Duran, V. Pascal Andreu, N. Selem-Mojica, M. Alanjary, S. L. Robinson, G. Lund, S. C. Epstein, A. C. Sisto, L. K. Charkoudian, J. Collemare, R. G. Linington, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2020, **48**, 454-458.

240 P. Stothard, *Biotechniques*, 2000, **28**, 1102-1104.

241 M. C. Andorfer, H. J. Park, J. Vergara-Coll and J. C. Lewis, *Chem. Sci.*, 2016, **7**, 3720-3729.

242 S. Kotha, K. Lahiri and D. Kashinath, *Tetrahedron*, 2002, **58**, 9633-9695.

243 M. S. Driver and J. F. Hartwig, *J. Am. Chem. Soc.*, 1996, **118**, 7217-7218.

244 A. Köllhofer, T. Pullmann and H. Plenio, *Angew. Chem.*, 2003, **42**, 1056-1058.

245 W. A. Herrmann., C. Brossmer, K. Ofele, C.-P. Reisinger, T. Priermeier, M. Beller and H. Fischer, 1995, **34**, 1844-1848.

246 D. H. Figurski and D. R. Helinski, *Proc. Natl. Acad. Sci. U. S. A.*, 1979, **76**, 1648-1652.

247 E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acids Res.*, 2003, **31**, 3784–3788.

248 R. Agarwala, T. Barrett, J. Beck, D. A. Benson, C. Bollin, E. Bolton, D. Bourexis, J. R. Brister, S. H. Bryant, K. Canese, M. Cavanaugh, C. Charowhas, K. Clark, I. Dondoshansky, M. Feolo, L. Fitzpatrick, K. Funk, L. Y. Geer, V. Gorelenkov, A. Graeff, W. Hlavina, B. Holmes, M. Johnson, B. Kattman, V. Khotomlianski, A. Kimchi, M. Kimelman, M. Kimura, P. Kitts, W. Klimke, A. Kotliarov, S. Krasnov, A. Kuznetsov, M. J. Landrum, D. Landsman, S. Lathrop, J. M. Lee, C. Leubsdorf, Z. Lu, T. L. Madden, A. Marchler-Bauer, A. Malheiro, P. Meric, I. Karsch-Mizrachi, A. Mnev, T. Murphy, R. Orris, J. Ostell, C. O'Sullivan, V. Palanigobu, A. R. Panchenko, L. Phan, B. Pierov, K. D. Pruitt, K. Rodarmer, E. W. Sayers, V. Schneider, C. L. Schoch, G. D. Schuler, S. T. Sherry, K. Siyan, A. Soboleva, V. Soussov, G. Starchenko, T. A. Tatusova, F. Thibaud-Nissen, K. Todorov, B. W. Trawick, D. Vakatov, M. Ward, E. Yaschenko, A. Zasypkin and K. Zbicz, *Nucleic Acids Res.*, 2018, **46**, D8–D13.

249 F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Res.*, 2019, **47**, W636–W641.

250 A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, 1189–1191.

251 E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin, *Protein Sci.*, 2021, **30**, 70–82.

Appendices

Appendix 1: Amino acid sequences of novel protein constructs used in this project

Highlighted in red are the His_8 affinity tags with TEV or Thrombin protease sites.

>AerA(PCP) (Calculated pI: 5.95; ε_{280}: 9970 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSELIAPRTEAERQVAKIWQKVLNLSQIGIHDNFFELGGH
SLLASQVISRLRDVFSVELSLHSLLEYPTVASLTQTIEVLNVANNSQSVSKSGKVM
TTASENYEEGEL

>AerJ (Calculated pI: 6.30; ε_{280}: 114625 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSMKTVEFLSDLNHLGVTIWMEGDKLRYRSPQGVMTPDLL
EQLKEHKEELIVLLREQADNFSSETDYDVAICGGGLAGLTLGRQLKLKQPNLSVVV
LDKMARPLPEAGFKVGESTVEVGAFYLANTLQLTDYFDHQHLPKLGLRYFFKPQET
EFHKRPELGLSEFHAPKSYQIDRGKLENDLRQFNIEAGIDLKENCSVKDIEFAEGL
QQQHKIIYTQGSGANQKTHCIKSRWVVDAMGRRRFIQKKLGLAKPNHNNYSAVWFR
VEGRFDVSNFVPASEEKWHRRVPNNNRYYSTNHLCGEGYWVWLIPLSTGYTSIGIV
ARQDIHPLKNYHNYELAFQWLRENEPVLAAHLEGKSPEDFRKMPKYSYSSKQVFSF
NRWACVGEAGLFPDPFYSPGSDSIGFGNSLTTQMIELDLKGQLTPERVDDANHFYL
TYHDGLTFNIQNSYNCMGNGIVMATKMIWDTLAGWTFGCLMMFNSIFLDPELKMKV
QQINAEFFPLSYRIQQLFRDWANQSLGRVSFEFIDYLAIPFVNELRTRNLQSNKTE
AEIIEAYRASLKLLEEFAQVIFQIALEDTSPELLTKINEHPWLNAWAISLDASKWE
TDGLFSPKNEPRNLHSIKKQYLEAINR

Note: Highlighted in blue is the catalytic lysine mutated to generate the K156A mutant.

>AerJΔβHDD (Calculated pI: 6.48; ε_{280}: 107635 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSDNFSSETDYDVAICGGGLAGLTLGRQLKLKQPNLSVVV
LDKMARPLPEAGFKVGESTVEVGAFYLANTLQLTDYFDHQHLPKLGLRYFFKPQET
EFHKRPELGLSEFHAPKSYQIDRGKLENDLRQFNIEAGIDLKENCSVKDIEFAEGL
QQQHKIIYTQGSGANQKTHCIKSRWVVDAMGRRRFIQKKLGLAKPNHNNYSAVWFR
VEGRFDVSNFVPASEEKWHRRVPNNNRYYSTNHLCGEGYWVWLIPLSTGYTSIGIV
ARQDIHPLKNYHNYELAFQWLRENEPVLAAHLEGKSPEDFRKMPKYSYSSKQVFSF
NRWACVGEAGLFPDPFYSPGSDSIGFGNSLTTQMIELDLKGQLTPERVDDANHFYL

TYHDGLTFNIQNSYNCMGNGIVMATKMIWDTLAGWTFGCLMMFNSIFLDPELKMKV
QQINAEFFPLSYRIQQLFRDWANQSLGRVSFEFIDYLAIPFVNELRTRNLQSNKTE
AEIIEAYRASLKLLEEFAQVIFQIALEDTSPELLTKINEHPWLNAWAISLDASKWE
TDGLFSPKNEPRNLHSIKKQYLEAINR

>AerJ βHDD (Calculated pI: 6.16; ε_{280}: 8480 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSMKTVEFLSDLNHLGVTIWMEGDKLRYRSPQGVMTPDLL
EQLKEHKEELIVLLREQA

>AerB(C) (Calculated pI: 5.37; ε_{280}: 62340 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSMKVAEFLSYLNSLDIKLWLEEEKLKYQAPQGAMTPEIK
QEIGTRKPEILTFLRSATTPSKPLESVINSVARTEDLPLSFAQQRMWFLYQMDQQN
SAYNEALTIRLTGRLNIDILEQTINAIIQRHESLRTTFPVVEGKPIQKIAPSLKIK
LLVVNLKDIPQEQIDKQIIEELQKPFDLTQSPLLRCTLFDLGYENYILVNVFHHII
IDGWSKGILFKELSKFYQAFLSNSTVNLPELPIQYADFAVWQRQWLQGEILENQLN
YWKKQLTGAPPLLELPTDKPRPATPNFRGHSISFQIDSELTENLKLLSQKSGVTLF
MTLLAALNTLLFRYSGQDDILIGTPTANRNRQEIEPLIGFFVNTLVLRNSLEGNPT
FSGLLQQARNVVLEAYANQDVPFEQVVDGLEIERSLSYNPLFQVMFAMQNAPLNAL
ELPDLKAQYLAVENQRIKFDLSLVLEEIETEKGAYLEGFWEYDSDLLTPERINRMV
GHFQTLLKGIVANPQQTIGELPLLTES

>AerB(C)ΔβHDD (Calculated pI: 5.31; ε_{280}: 53860 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSTTPSKPLESVINSVARTEDLPLSFAQQRMWFLYQMDQQ
NSAYNEALTIRLTGRLNIDILEQTINAIIQRHESLRTTFPVVEGKPIQKIAPSLKI
KLLVVNLKDIPQEQIDKQIIEELQKPFDLTQSPLLRCTLFDLGYENYILVNVFHHI
IIDGWSKGILFKELSKFYQAFLSNSTVNLPELPIQYADFAVWQRQWLQGEILENQL
NYWKKQLTGAPPLLELPTDKPRPATPNFRGHSISFQIDSELTENLKLLSQKSGVTL
FMTLLAALNTLLFRYSGQDDILIGTPTANRNRQEIEPLIGFFVNTLVLRNSLEGNP
TFSGLLQQARNVVLEAYANQDVPFEQVVDGLEIERSLSYNPLFQVMFAMQNAPLNA
LELPDLKAQYLAVENQRIKFDLSLVLEEIETEKGAYLEGFWEYDSDLLTPERINRM
VGHFQTLLKGIVANPQQTIGELPLLTES

>AerB βHDD (Calculated pI: 6.70; ε_{280}: 9970 M^{-1} cm^{-1})

MHHHHHHHHGENLYFQGSMKVAEFLSYLNSLDIKLWLEEEKLKYQAPQGAMTPEIK
QEIGTRKPEILTFLRSA

>AerB(C-A-PCP-E) (Calculated pI: 5.44; ε_{280}: 201960 M^{-1} cm^{-1})

MKHHHHHHHHGGLVPRGSHGSMKVAEFLSYLNSLDIKLWLEEEKLKYQAPQGAMTP
EIKQEIGTRKPEILTFLRSATTPSKPLESVINSVARTEDLPLSFAQQRMWFLYQMD
QQNSAYNEALTIRLTGRLNIDILEQTINAIIQRHESLRTTFPVVEGKPIQKIAPSL
KIKLLVVNLKDIPQEQIDKQIIEELQKPFDLTQSPLLRCTLFDLGYENYILVNVFH
HIIIDGWSKGILFKELSKFYQAFLSNSTVNLPELPIQYADFAVWQRQWLQGEILEN
QLNYWKKQLTGAPPLLELPTDKPRPATPNFRGHSISFQIDSELTENLKLLSQKSGV
TLFMTLLAALNTLLFRYSGQDDILIGTPTANRNRQEIEPLIGFFVNTLVLRNSLEG
NPTFSGLLQQARNVVLEAYANQDVPFEQVVDGLEIERSLSYNPLFQVMFAMQNAPL
NALELPDLKAQYLAVENQRIKFDLSLVLEEIETEKGAYLEGFWEYDSDLLTPERIN
RMVGHFQTLLKGIVANPQQTIGELPLLTESEKQQLLVEWNQTQTSYPDHYCIHQLF
EEQVVKTPDAIAVIDGEQSLTYEQLNQKANQIARYLQKLGVKPDQLIGICVERLPL
MIIGFLGILKAGGAYVPLDTSYPEERLESIINDAEIKILLTQQHLIKKLPTTVNQL
ICLEQDEDPIKQQSKQNPLSHVNAENLAYVIYTSGSTGKPKGVAIPHRGVTRLVCN
TNYIYLKPTDRMAQVSNPSFDAATFEIWGALIHGAQLIGVDKDRVLSPEKFVHFLQ
KYQITILFLTTALFNQLVQTIPHAFQSLRYLLFGGEAVDPQWVKESLNKGTAQNLL
HVYGPTESTTFTSWYLIKNIPSEATTIPIGRPLANTEIYILDPYLQPVPIGVKGEL
HIGGDGLARCYLNRPDLTEQKFIPNPFSQDSSARLYKTGDIVRYLADGNIEFIGRI
DHQVKIRGFRIELGEIEAVLFQHPQVKDGIVIAREDLLGIKRLYAYVVPKDKQLTQ
PELRIFLQEKLPNYMIPAFLIFLDAFTLNQNGKIDRSVLPRPEIDVNELENYLLPS
TEKETILAQIWQEVLGQNQISINENFFELGGDSIIAIQIVAKANQVGLQITPKQLF
SHQTIAQLATVAETTAITEIDQGLVTGEVPLTPIQKWFFEQNWPERHHFNQSILLE
VPNNLQPDLLKQTISKLLYHHDALRLRFVQKGEQWQQNHSDDCNNFAFEKVDLSHL
SYHEQLTKMAEISEVQQRVLNLEEGPLMAVVFFALGESGKMLIVIHHLAVDGISWR
IILEDFVTIYQQLETQKPLQLPPKTSSFKTWAEELQNYAKTPEFYAQFKYWLNRDF
SSVSPLPVDVQGDAQSNIVAHAKTVSFTLTEEQTRLLLQEVPQAYNTQINDILLTA
LVQAFGRWTGNYKLLLDMEGHGRENVIESVNLSRTVGWFTSIFPVFLTLENLHHPG
ECLKSIKEQLRQIPNRGFDYGIGYYLSSDLTIQSPLKNYPKAQVSFNYLGQFTSHQ
IGEIGWKLSQESTGSIHSPLGQRSHLIAIHGIVVDGQLDMEWQYSENFHHQTTIKN
LAAAYRDSLESLINHCLSVEGGYTPSDFPDADLNQAELDELLSELDF

>AerB(C-A) (Calculated pI: 5.74; ε_{280}: 116005 M^{-1} cm^{-1})

MKHHHHHHHHGGLVPRGSHGSMKVAEFLSYLNSLDIKLWLEEEKLKYQAPQGAMTP
EIKQEIGTRKPEILTFLRSATTPSKPLESVINSVARTEDLPLSFAQQRMWFLYQMD

QQNSAYNEALTIRLTGRLNIDILEQTINAIIQRHESLRTTFPVVEGKPIQKIAPSL
KIKLLVVNLKDIPQEQIDKQIIEELQKPFDLTQSPLLRCTLFDLGYENYILVNVFH
HIIIDGWSKGILFKELSKFYQAFLSNSTVNLPELPIQYADFAVWQRQWLQGEILEN
QLNYWKKQLTGAPPLLELPTDKPRPATPNFRGHSISFQIDSELTENLKLLSQKSGV
TLFMTLLAALNTLLFRYSGQDDILIGTPTANRNRQEIEPLIGFFVNTLVLRNSLEG
NPTFSGLLQQARNVVLEAYANQDVPFEQVVDGLEIERSLSYNPLFQVMFAMQNAPL
NALELPDLKAQYLAVENQRIKFDLSLVLEEIETEKGAYLEGFWEYDSDLLTPERIN
RMVGHFQTLLKGIVANPQQTIGELPLLTESEKQQLLVEWNQTQTSYPDHYCIHQLF
EEQVVKTPDAIAVIDGEQSLTYEQLNQKANQIARYLQKLGVKPDQLIGICVERLPL
MIIGFLGILKAGGAYVPLDTSYPEERLESIINDAEIKILLTQQHLIKKLPTTVNQL
ICLEQDEDPIKQQSKQNPLSHVNAENLAYVIYTSGSTGKPKGVAIPHRGVTRLVCN
TNYIYLKPTDRMAQVSNPSFDAATFEIWGALIHGAQLIGVDKDRVLSPEKFVHFLQ
KYQITILFLTTALFNQLVQTIPHAFQSLRYLLFGGEAVDPQWVKESLNKGTAQNLL
HVYGPTESTTFTSWYLIKNIPSEATTIPIGRPLANTEIYILDPYLQPVPIGVKGEL
HIGGDGLARCYLNRPDLTEQKFIPNPFSQDSSARLYKTGDIVRYLADGNIEFIGRI
DHQVKIRGFRIELGEIEAVLFQHPQVKDGIVIAREDLLGIKRLYAYVVPKDKQLTQ
PELRIFLQEKLPNYMIPAFLIFLDAFTLNQNGKIDRSVLPRPEIDVNELE

Note: Highlighted in blue is the catalytic histidine mutated to generate the H222A mutant.

>AerB(PCP) (Calculated pI: 5.86; ε_{280}: 6990 M^{-1} cm^{-1})

MKHHHHHHHHGGLVPRGSHGSNYLLPSTEKETILAQIWQEVLGQNQISINENFFEL
GGDSIIAIQIVAKANQVGLQITPKQLFSHQTIAQLATVAETTAITEIDQGLLSELD
F

>AerA(A-KR-PCP) (Calculated pI: 5.92; ε_{280}: 192950 M^{-1} cm^{-1})

MKHHHHHHHHGENLYFQGSMRTEHKNPKLNCQKLAEVFLSNPLIEDCYFMIREGEL
VAYIVSSGAWNSEQLSSDLQSQLPDNLLPNIYVQISSLPLTDSGDVDETALKNIEI
IDDHQVQETEEKLGSLSDIDQVAVVMTSKKNKILPLHLSELLPLKIEKKNDHQKSE
TPIHLVIHTSNTQAGISHGQLLEESELQTKTLAEILQKAAENYSNKGIIYIQSNGS
EVFQSYQQLWEEAQQIQAGLQKQGLQAQDKVILQLSENYDIISAFWGCLLGGFIPV
IISVPPSYKDFNNEIHKISQVWQLLDEPTIITNQSRQQEIKHLKQWLPNQTLRLSF
IEELKTYSPHQTHISQPDDIAFFNLTSGSTGIPKCISLTHKNVISRARGTNIICEH
QNDDIILNWLPFDHIGSISDWNIRCVELGCQMVYVQTEYILGRPLNWLDLIAQYRI
THSWAPNFAYNLINEALKKEPDQNWNLDCVKFFLAAGEAVSGQAVGEFINTLHLQY

NLKKTAMRPAFGMAEMGSGITYYQPTEQQPLLFHTVDKSSLNSSLKRVHPEHPNGA
TFTDLGLPIPGISIRIVNADNSLLPEETIGHLQVKGHAVSPGYYKNPEANQDAFLK
DGWFKTGDLAFISNGHLVITGRSKETIIINGVNYYSHEIETVVETIEEVQASYTAA
CAVYDPRNSTDQLALFFSAETFDHQHLAELLKKIRRKVINSFGVNPEYLIPLNKTE
IPKTSIGKIQRSQLTKRFGNGEFNSILKEIDILLENDQTIPDWFYQKVWKPRSPVN
LKPELSKYCSLIFLDCLGLGASLAEKIEGQNLPCITVLAAEEFSQISQNCYTLRPG
TANHYQQLMASLSERKIILGNIIHLGTYQDYQGEISTIEQLEKAQEKGTYDLLFLV
QALAKIHNFNSKIQLLFVSSYSQFVIETDEIAYEKSPVLGLIKTLAQELPWLNTRH
VDLPLDQTEINVSYLLQELSVLSKEREIAYRNGKRLIAGLEKVNLLQHTQQELPFK
SGGIYLITGGLGGIGRQIAQYLLKNYQARLLLIGKTPLPEKHLWSEYLKTEDQLSL
KIKSLQALENLGGEVIYQAVDVANFPQVKQAVEQVKKQWQGELNGVIHLAGIYKDC
LLLDETQEGLSTILHSKVIGTWVLHQLLKESQGIFISFSSLASFFGGAALGSYAAA
NSFLESLNSYQKSKNLFPSYCYSWTTWQETGISQGYQMQYITRSQGYYDMTVRQGL
DSFLTSLYHNQRQLIIGLDGSNSKINRFTSRSEGCQKLTAYYTRKSTVQSVNLPNH
VSLKDRLGTSYQCPLVLRQSLPIKDNGDIDKRQLLKELQGKENNELIAPRTEAERQ
VAKIWQKVLNLSQIGIHDNFFELGGH**S**LLASQVISRLRDVFSVELSLHSLLEYPTV
ASLTQTIEVLNVANNSQSVSKSGKVMTTASENYEEGEL

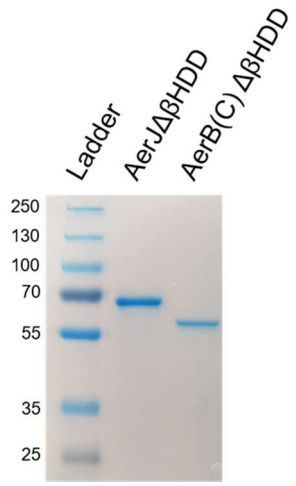
Note: Highlighted in blue is the conserved serine mutated to generate the S1352A mutant.

>Sfp (Calculated pI: 6.23; ε_{280}: 29130 M^{-1} cm^{-1})

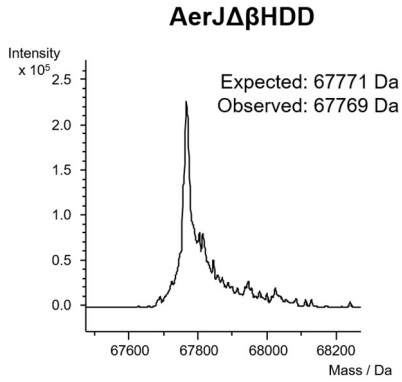
MGSSHHHHHHSSGLVPRGSHMKIYGIYMDRPLSQEENERFMSFISPEKQEKCRRFY
HKEDAHRTLLGDVLVRSVISRQYQLDKADIRFSAQEYGKPCIPDLPDAHFNISHSG
RWVICAFDSHPIGIDIEKMKPISLEIAKRFFSKTEYSDLLAKNKDEQTDYFYHLWS
MKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELPDSHTPCYIKTYEVDPGYKMAV
CAAHPDFPEDITMLSYEALL

Appendix 2: Characterisation of AerJΔβHDD and AerB(C)ΔβHDD constructs

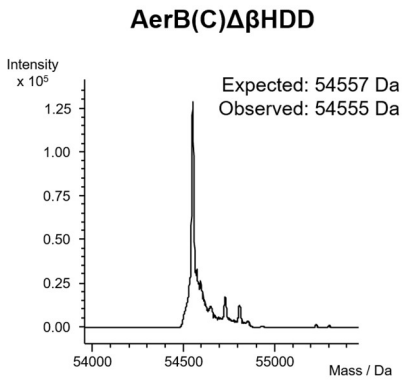
SDS-PAGE analysis:



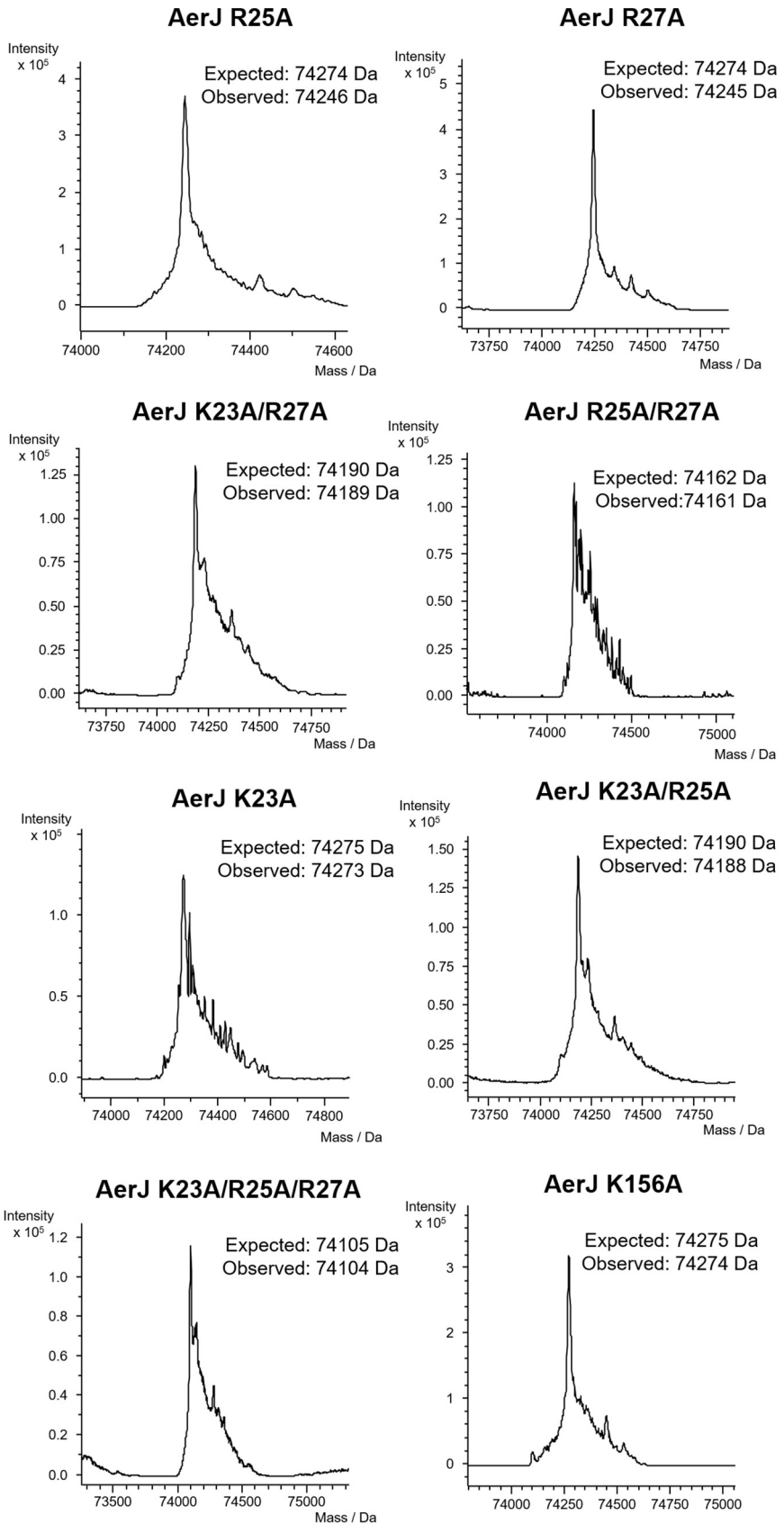
UHPLC-ESI-Q-TOF MS analysis of AerJΔβHDD:



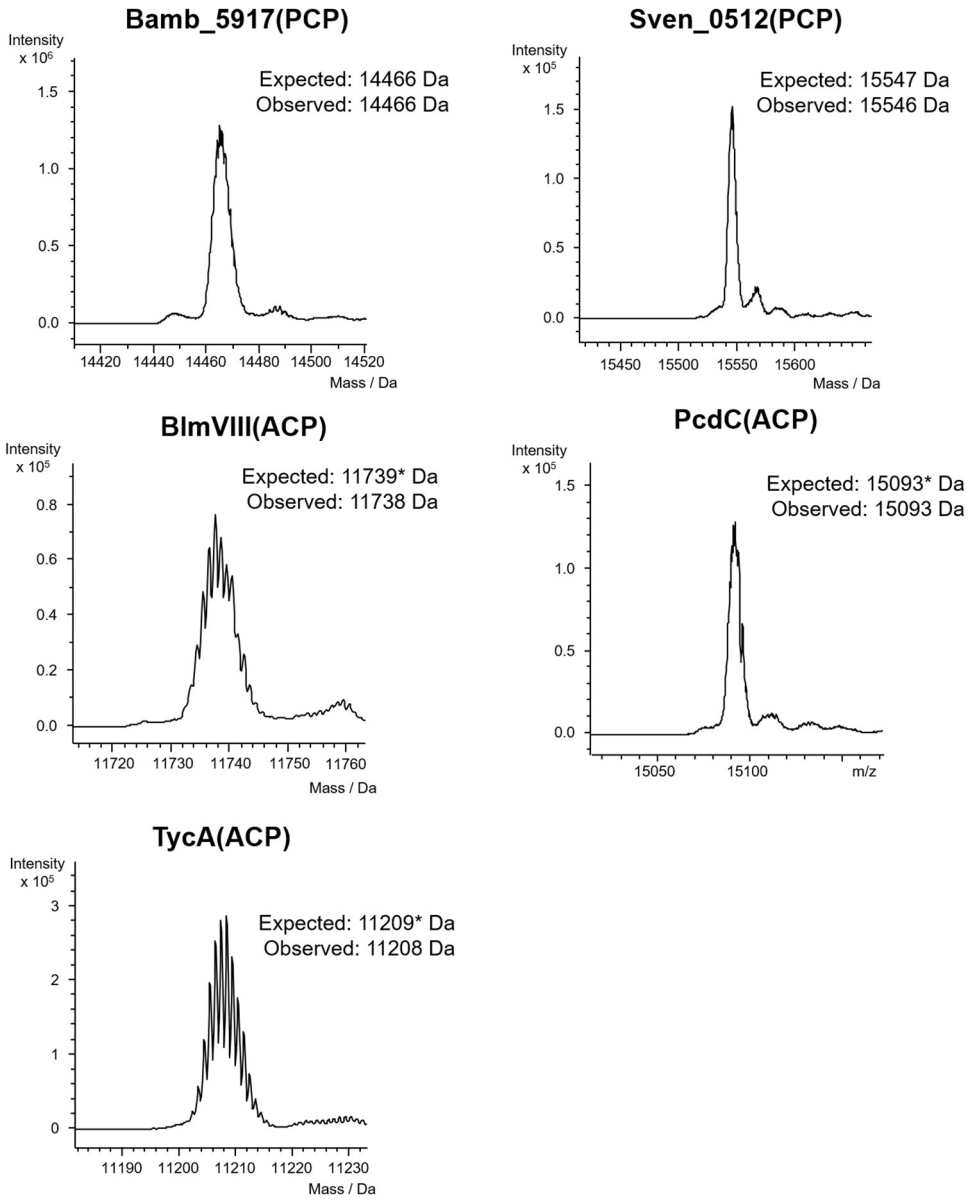
UHPLC-ESI-Q-TOF MS analysis of AerB(C)ΔβHDD:



Appendix 3: Charge-state deconvoluted mass spectra of AerJ mutants from UHPLC-ESI-Q-TOF-MS analysis

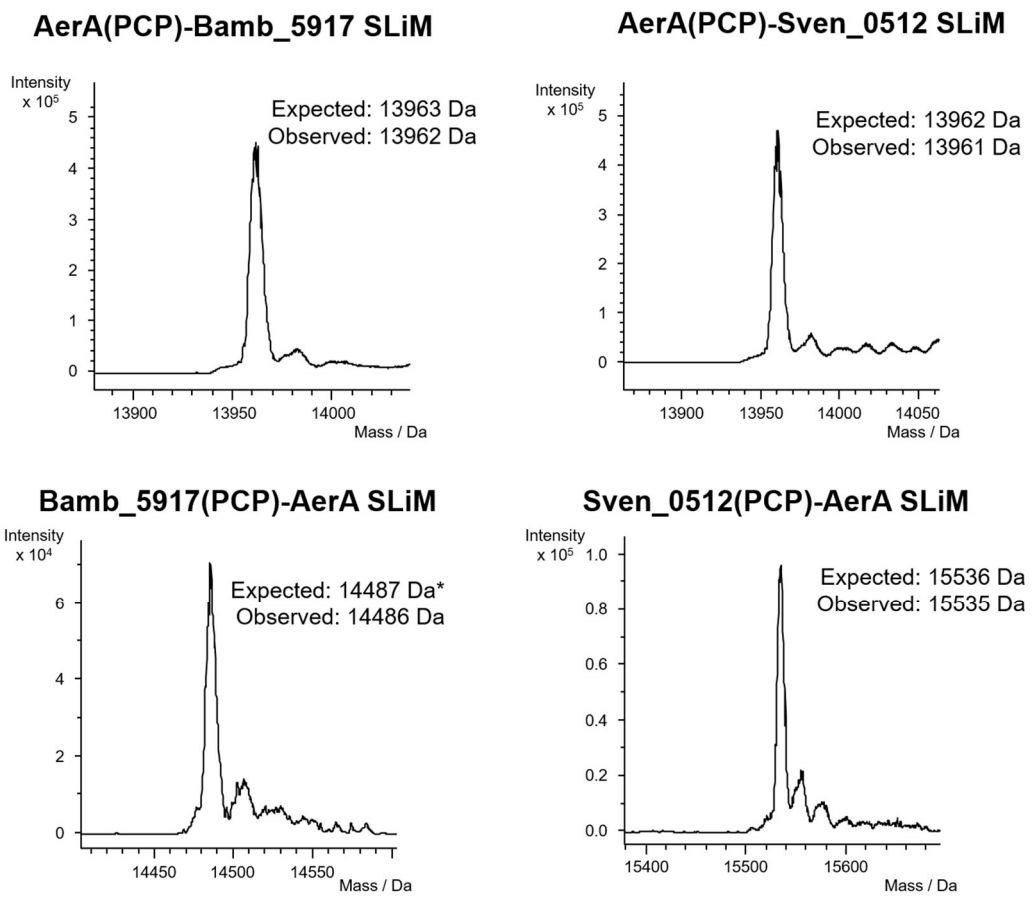


Appendix 4: Charge-state deconvoluted mass spectra of heterologous carrier proteins from UHPLC-ESI-Q-TOF-MS analysis



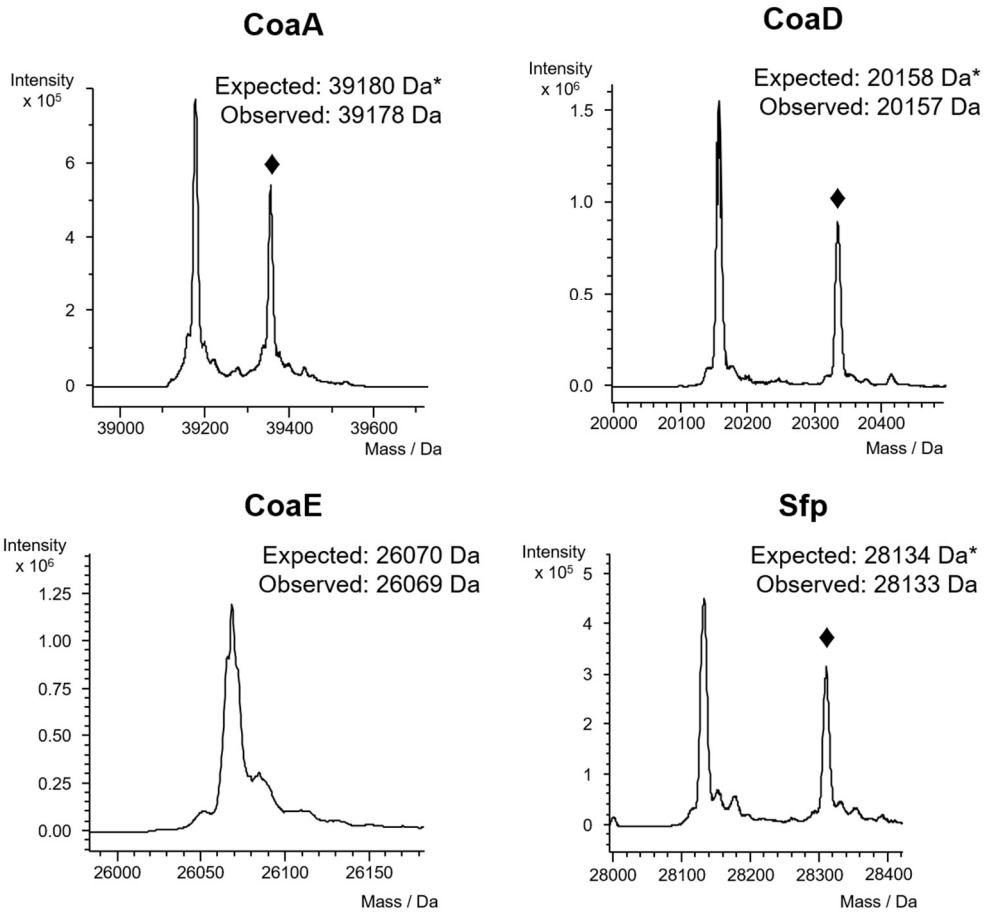
Note: Asterisks (*) indicate the expected mass of the protein was calculated with the N-terminal methionine removed.

Appendix 5: Charge-state deconvoluted mass spectra of SLiM-swapped mutants from UHPLC-ESI-Q-TOF-MS analysis



Note: Asterisks (*) indicate the expected mass of the protein was calculated with the N-terminal methionine removed.

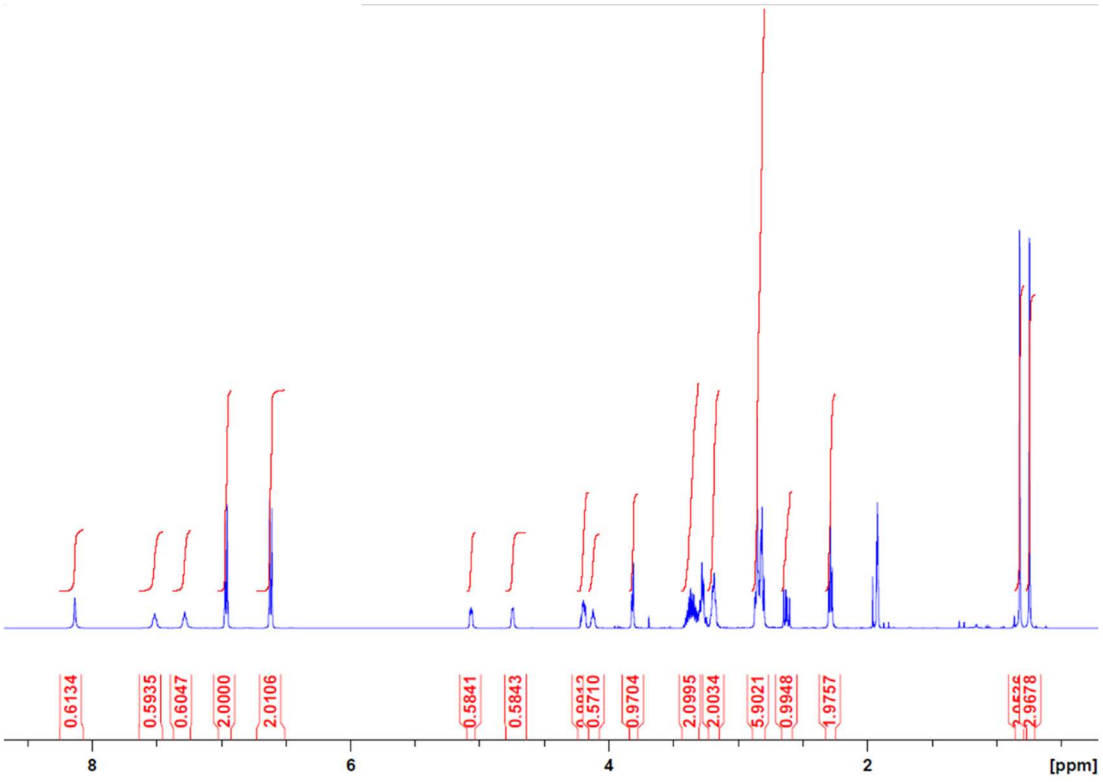
Appendix 6: Charge-state deconvoluted mass spectra of loading enzymes from UHPLC-ESI-Q-TOF-MS analysis



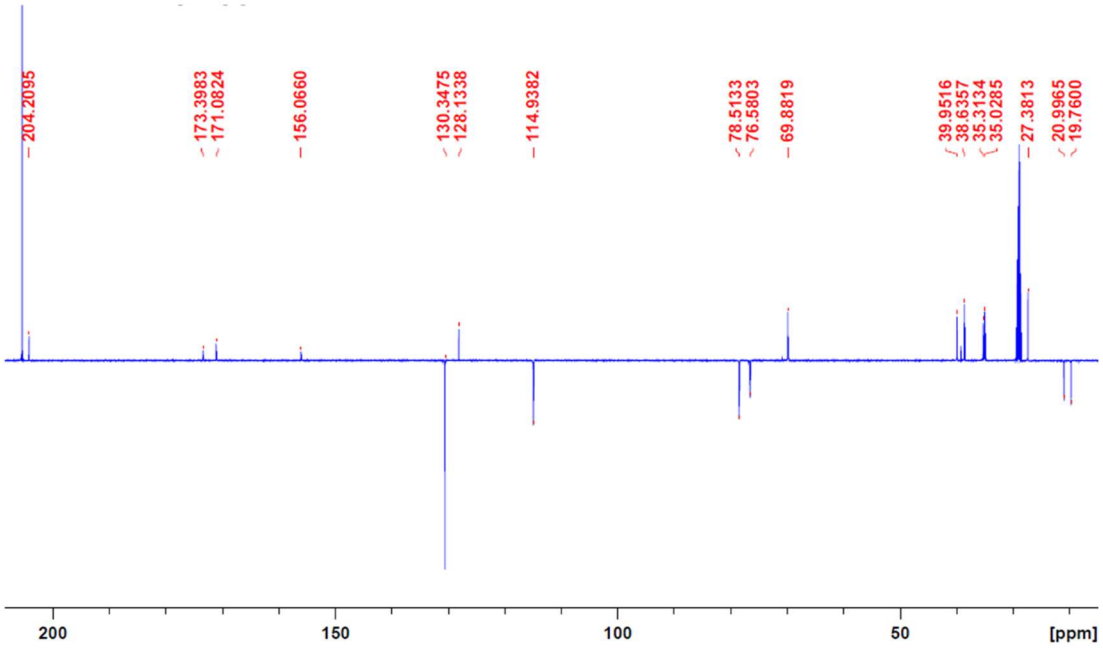
Note: Asterisks (*) indicate the expected mass of the protein was calculated with the N-terminal methionine removed. Diamonds (♦) indicate peaks corresponding to protein gluconylated at the N-terminus following methionine cleavage (+178 Da).

Appendix 7: ^{1}H and ^{13}C NMR spectra of 4-HPLA-*S*-pantetheine (51)

^{1}H



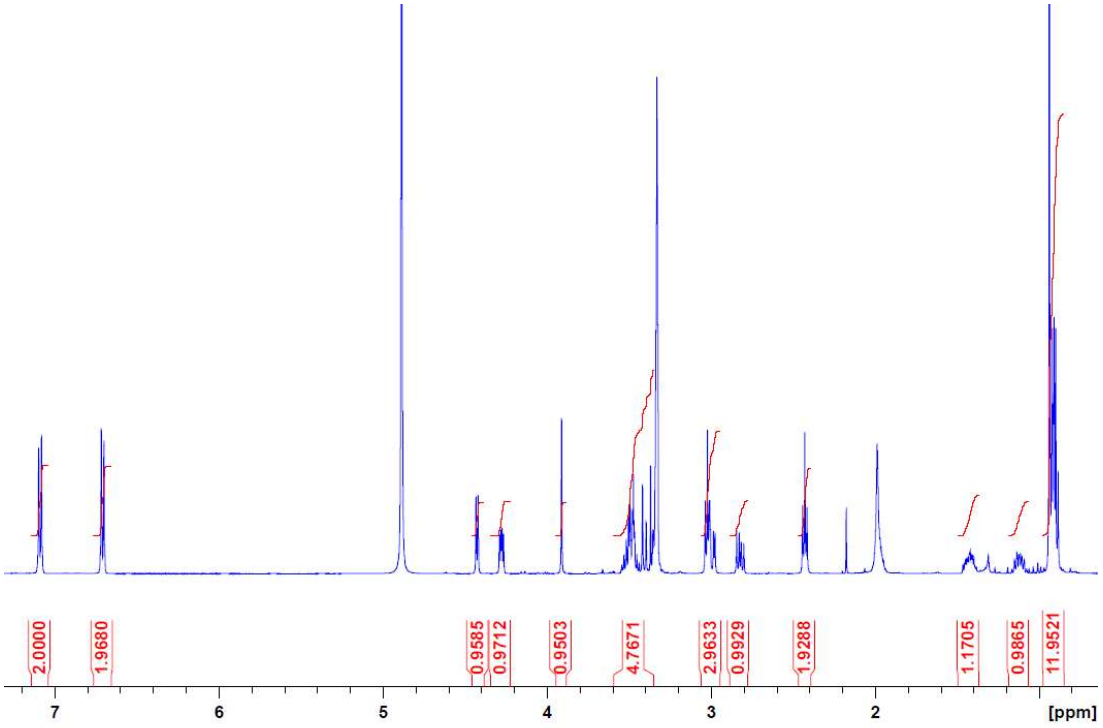
^{13}C



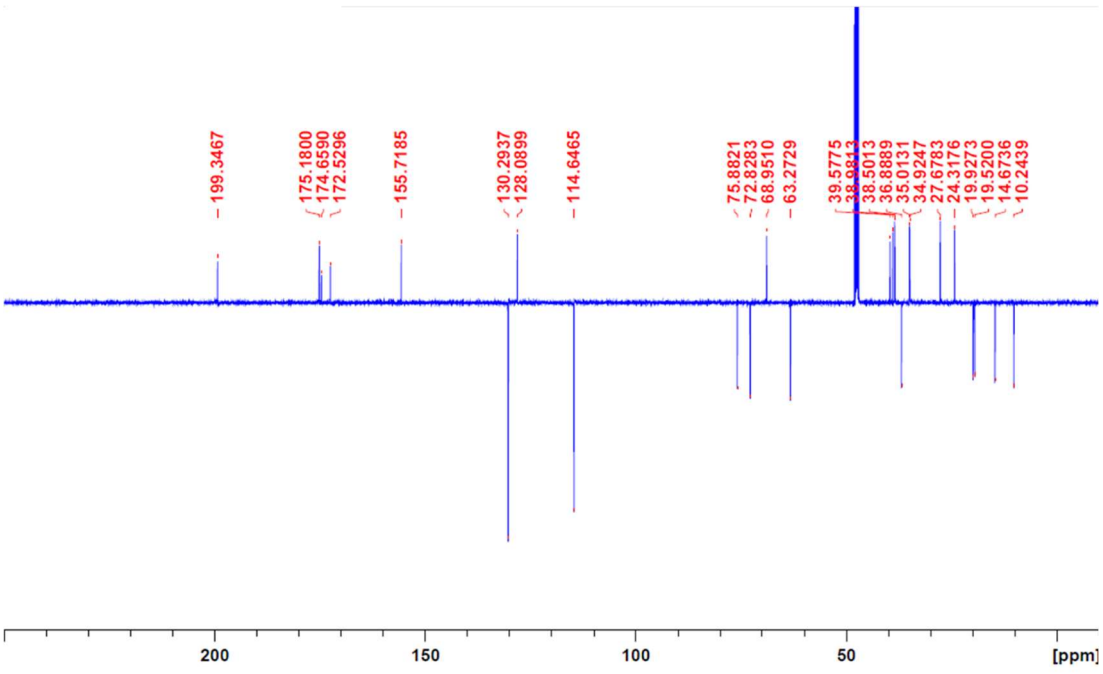
Appendix 8: ^{1}H and ^{13}C NMR spectra of 4-HPLA-Ile-*S*-pantetheine (72) diastereomers

72, diastereomer 1:

^{1}H

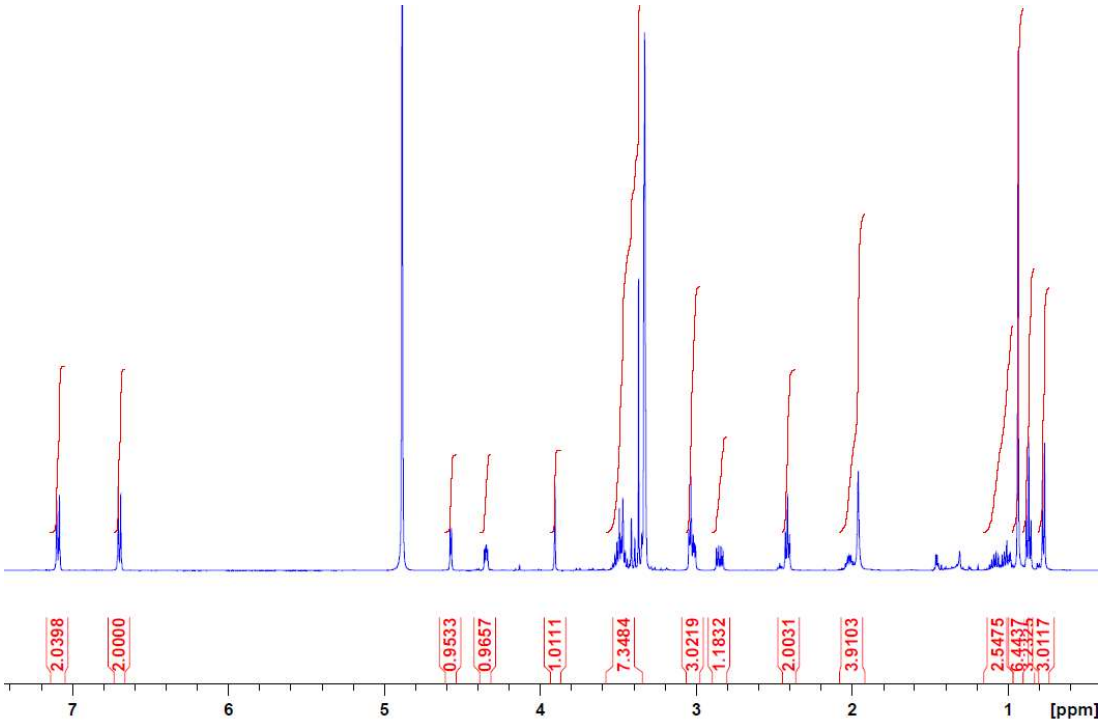


^{13}C

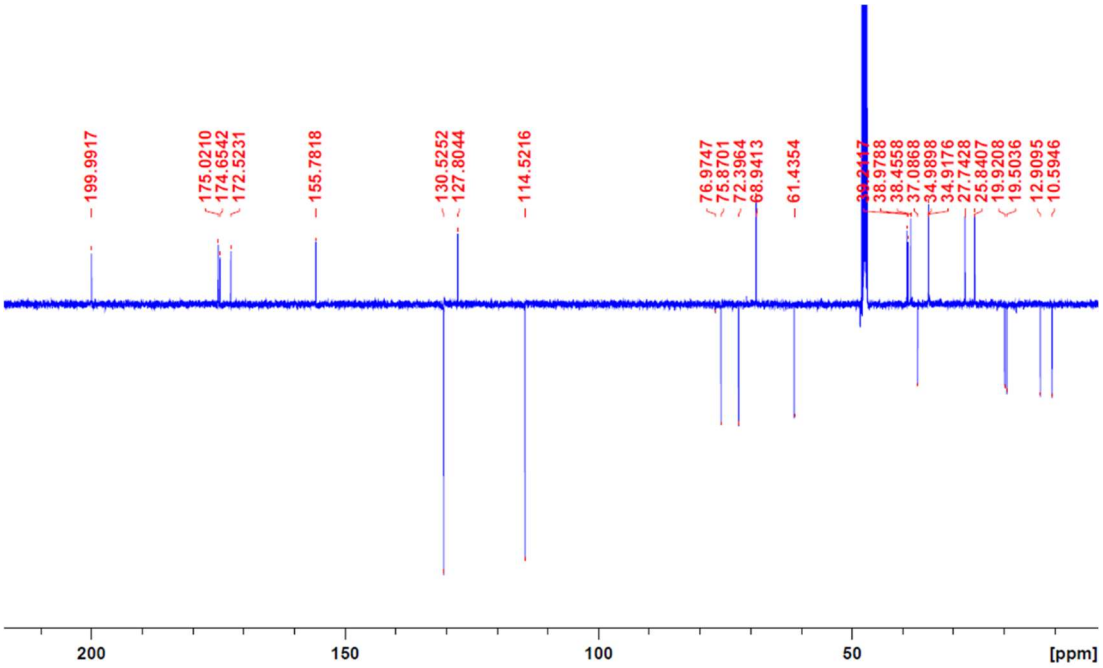


72, diastereomer 2:

^{1}H

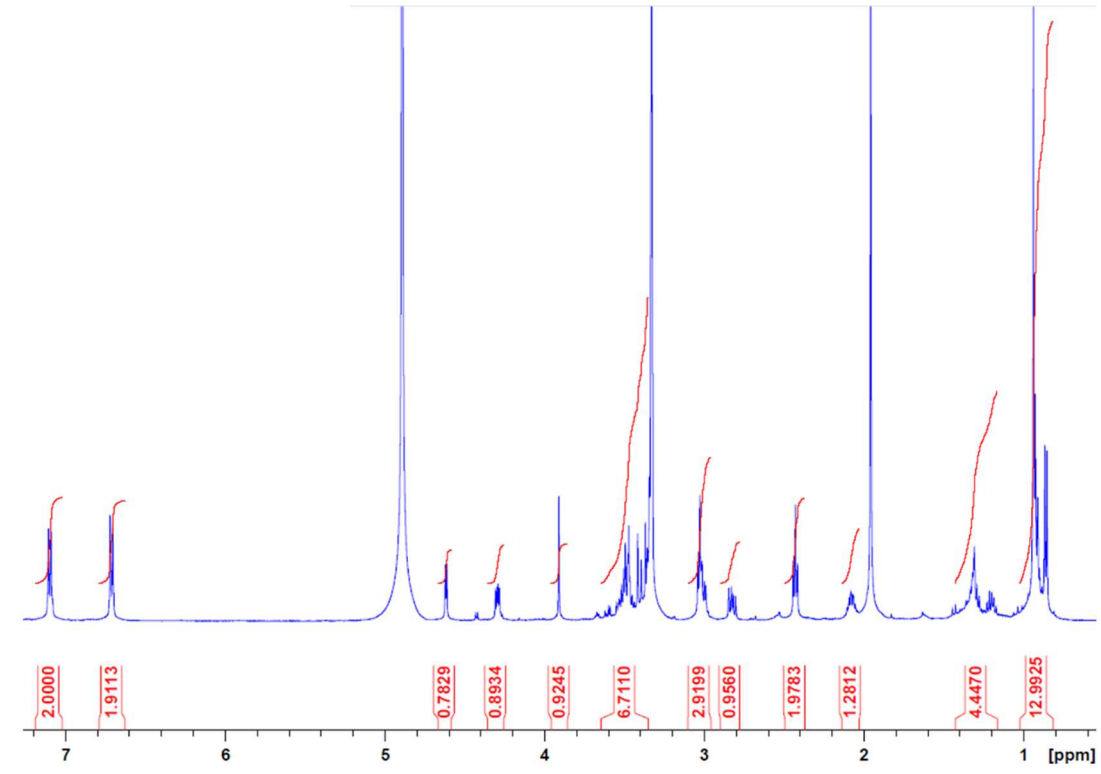


^{13}C

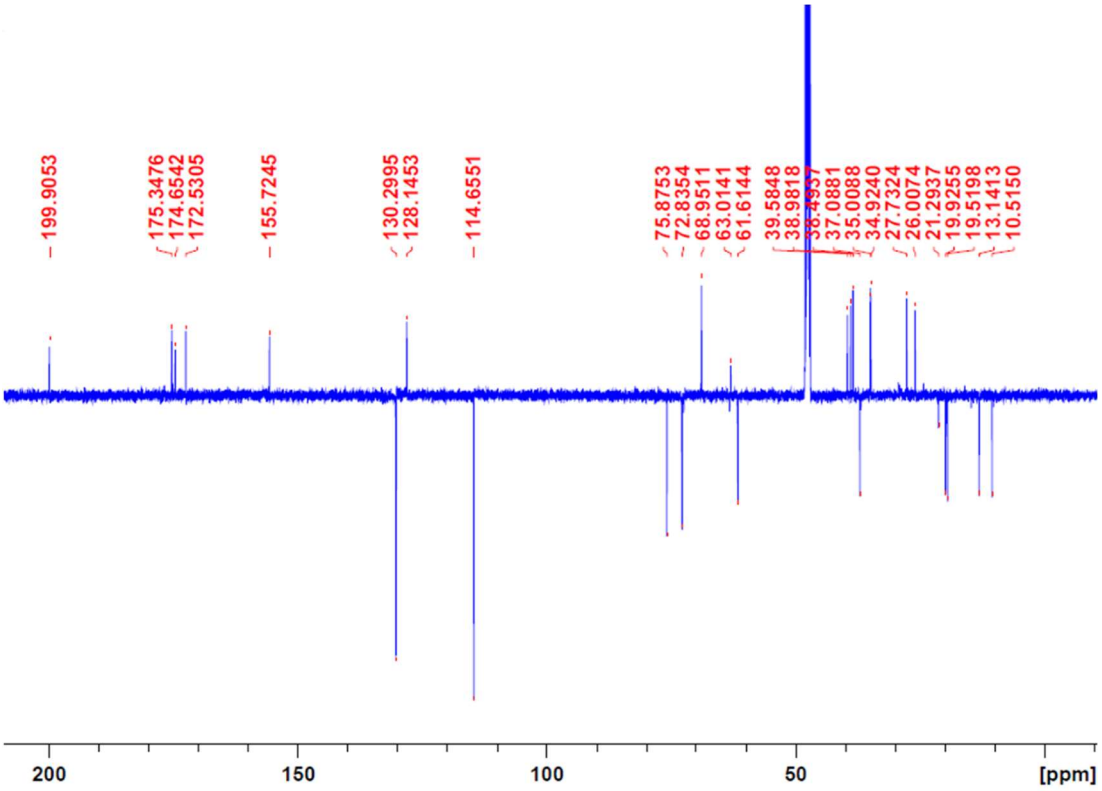


72, diastereomer 3:

^{1}H

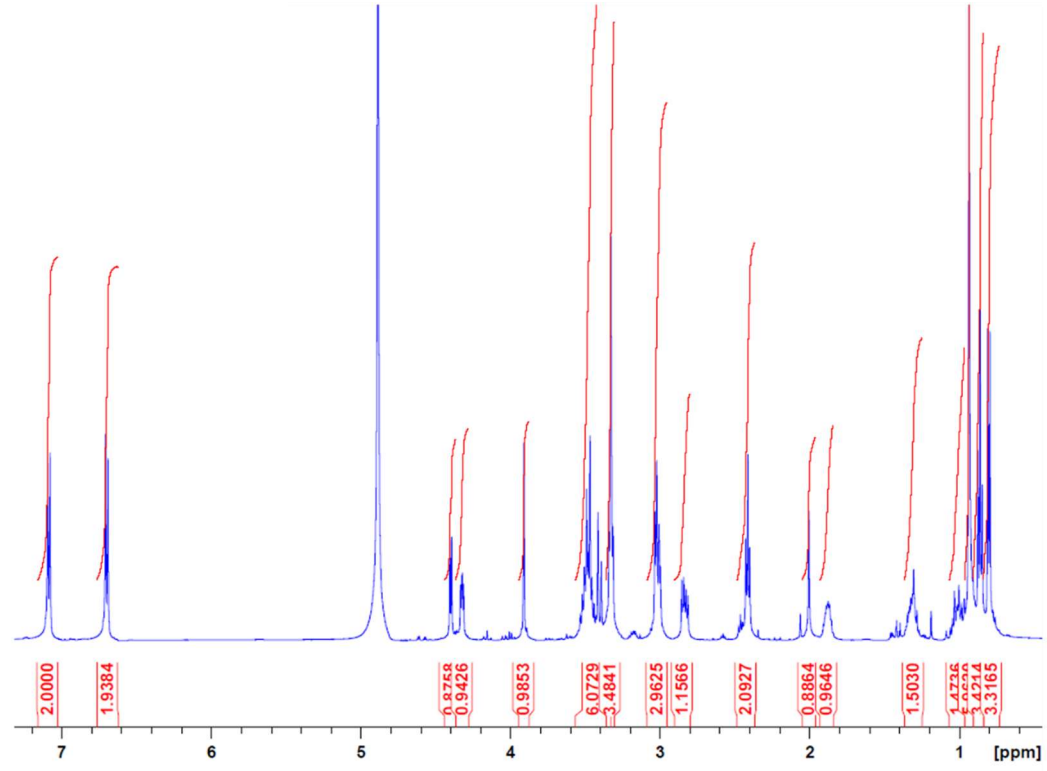


^{13}C

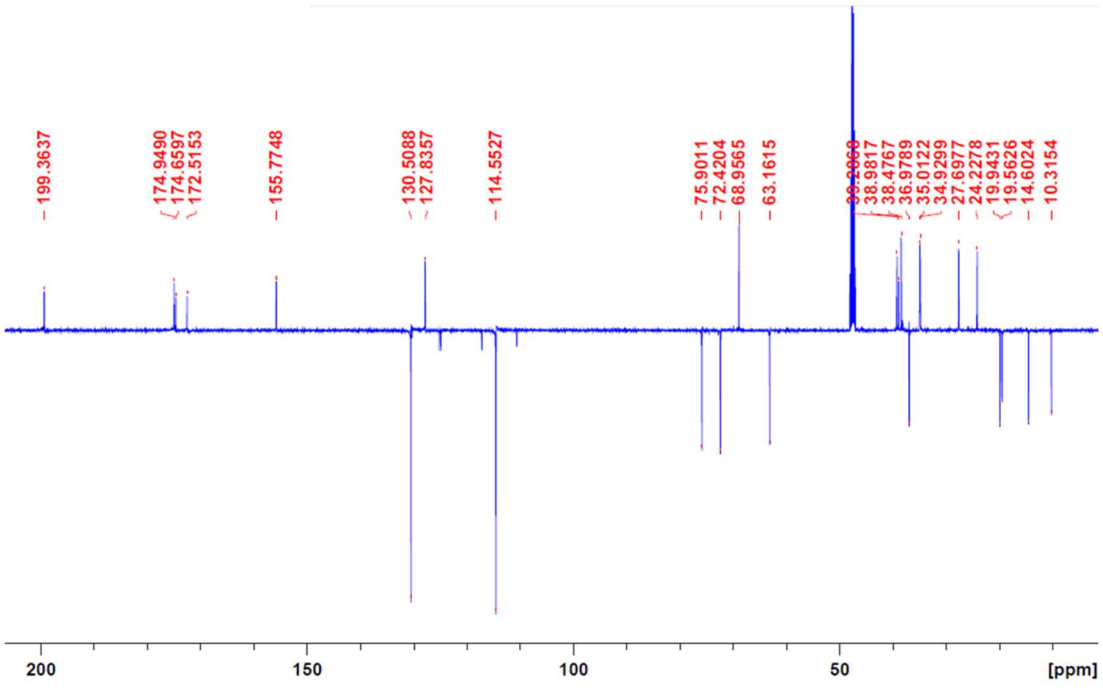


72, diastereomer 4:

^{1}H



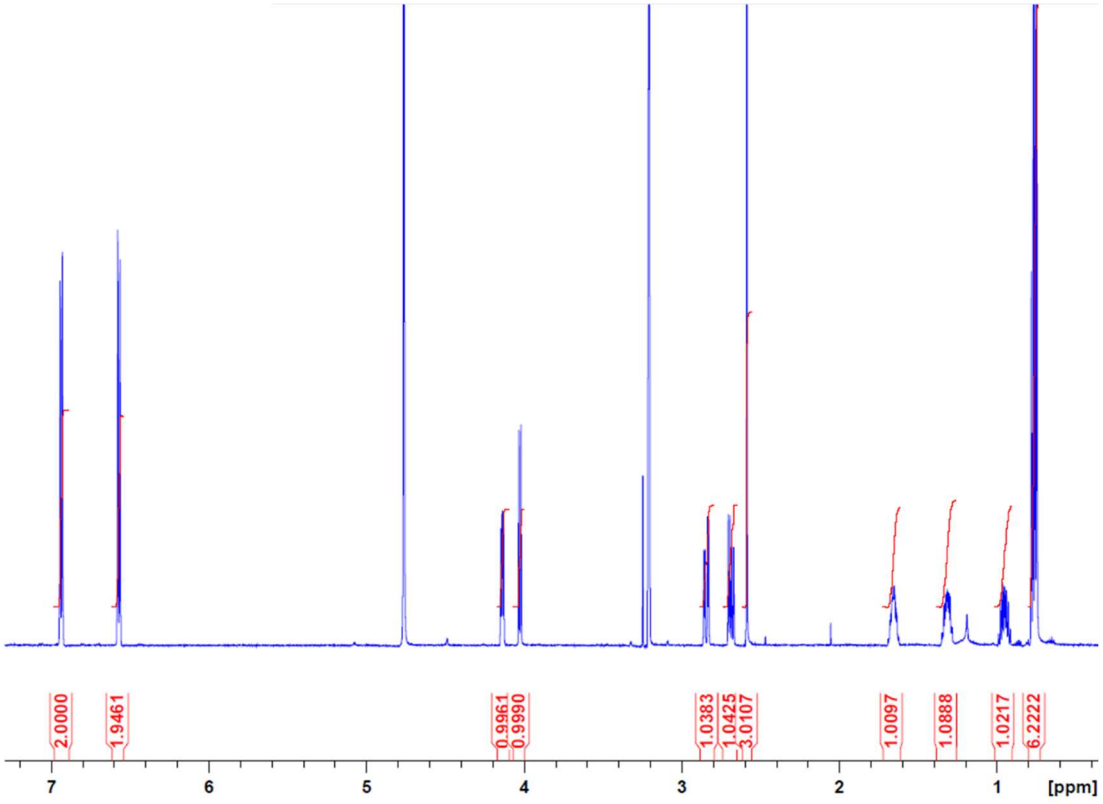
^{13}C



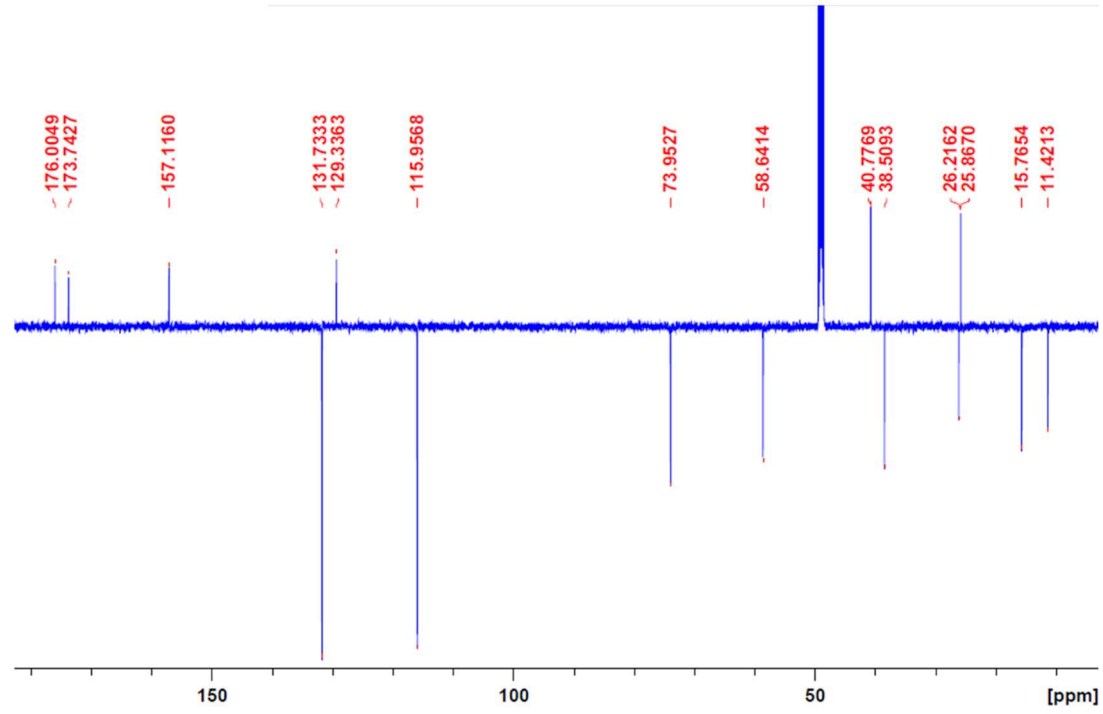
Appendix 9: ^{1}H and ^{13}C NMR spectra of 4-HPLA-L-Ile methyl amide (90) diastereomers

90, diastereomer 1:

^{1}H

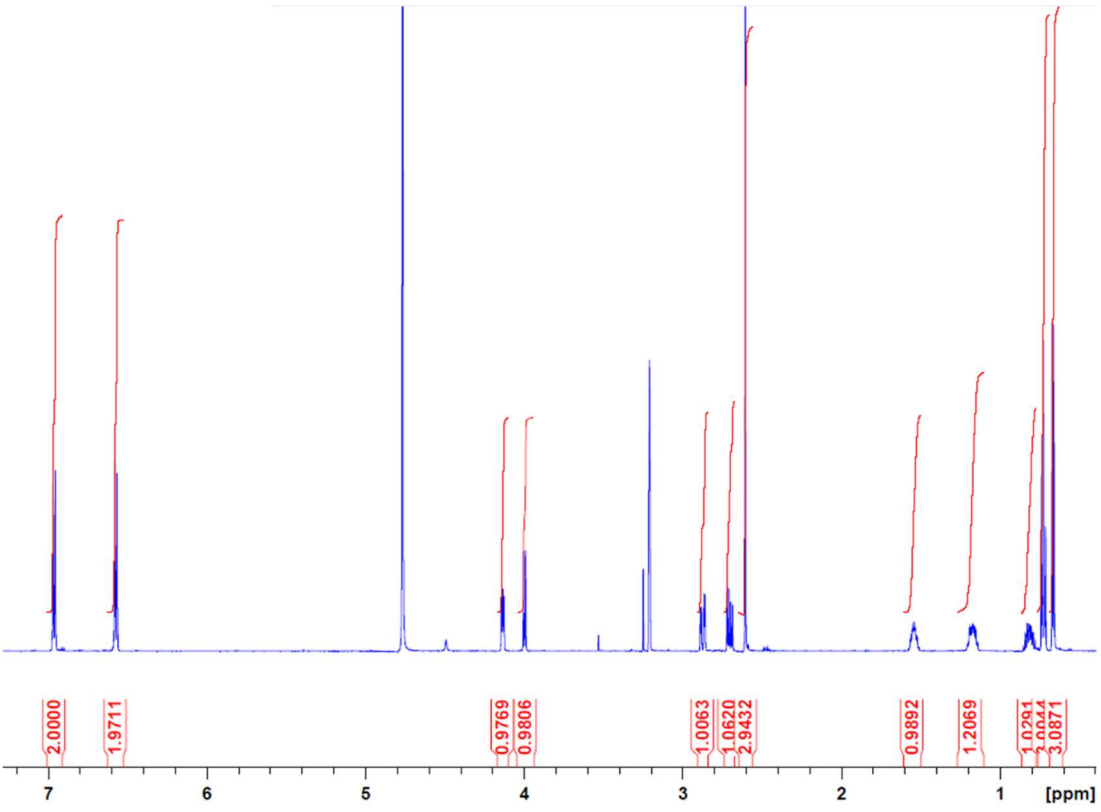


^{13}C

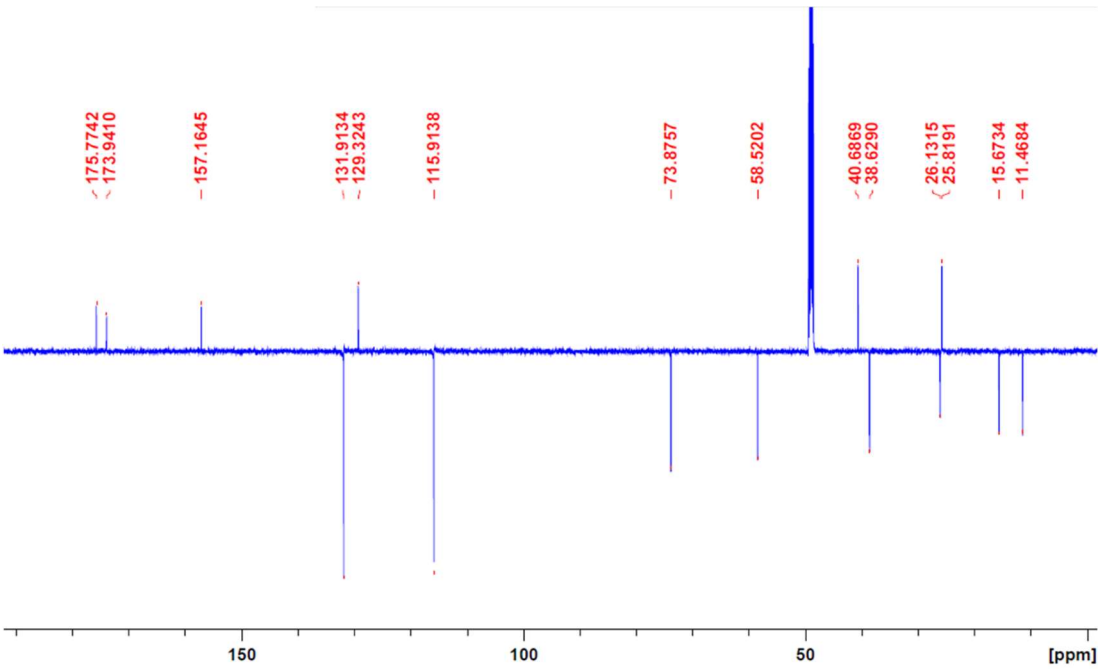


90, diastereomer 2:

^{1}H

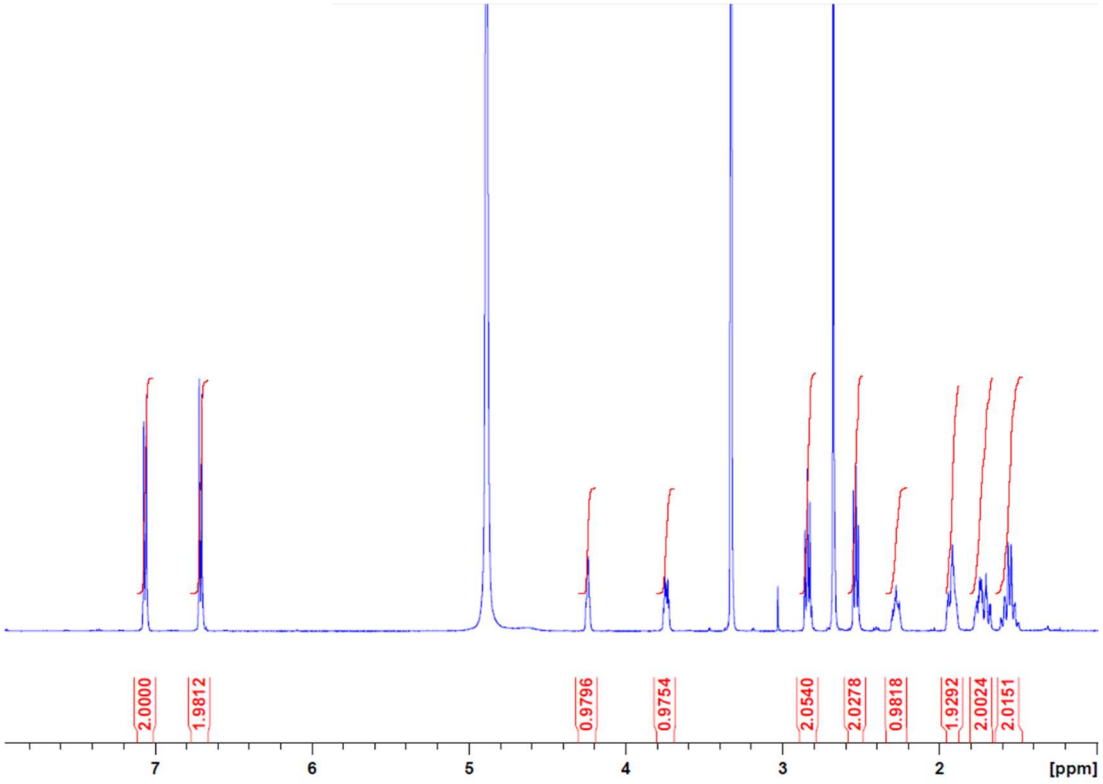


^{13}C

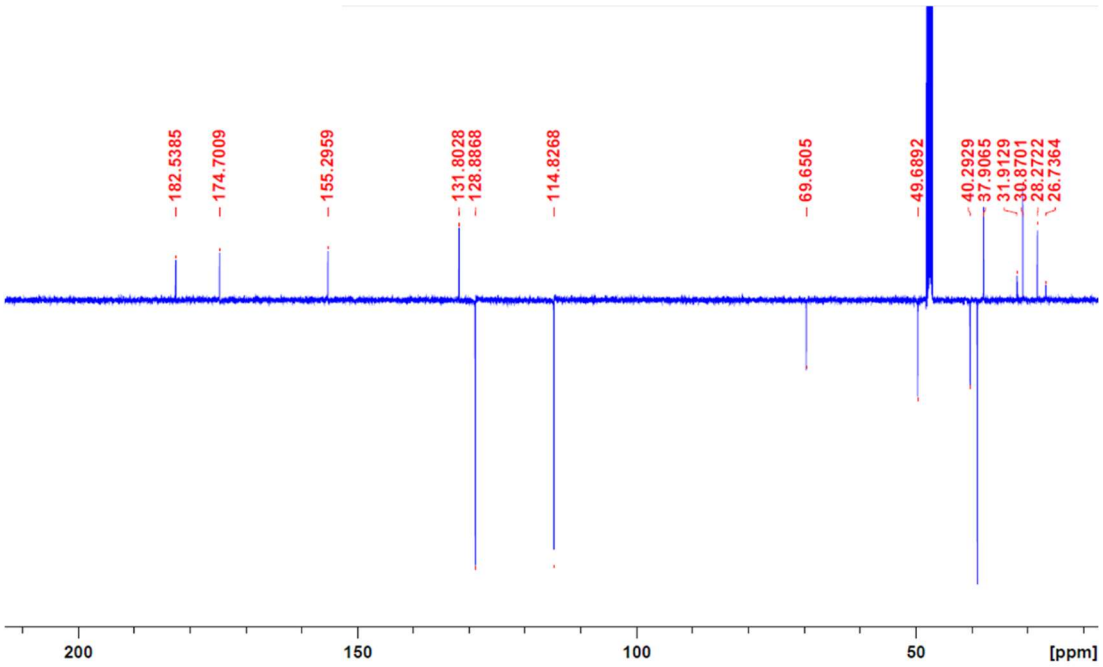


Appendix 10: ^{1}H and ^{13}C NMR spectra of 4-HPPA-AHCCA (97)

^{1}H

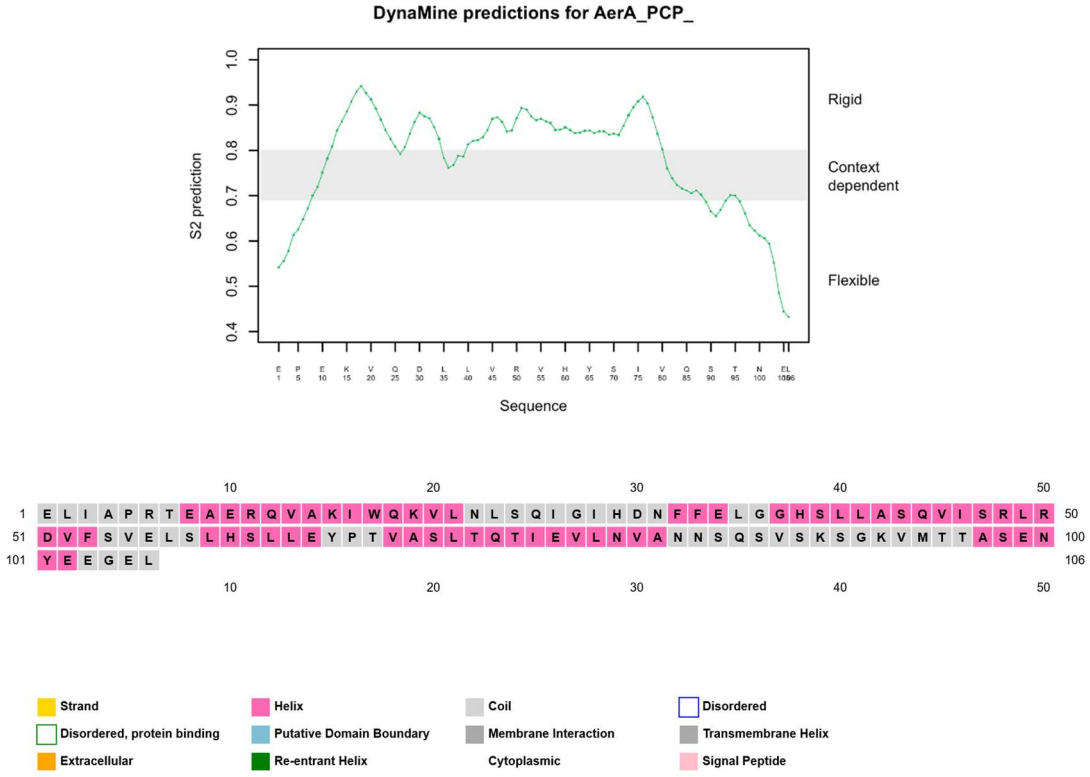


^{13}C

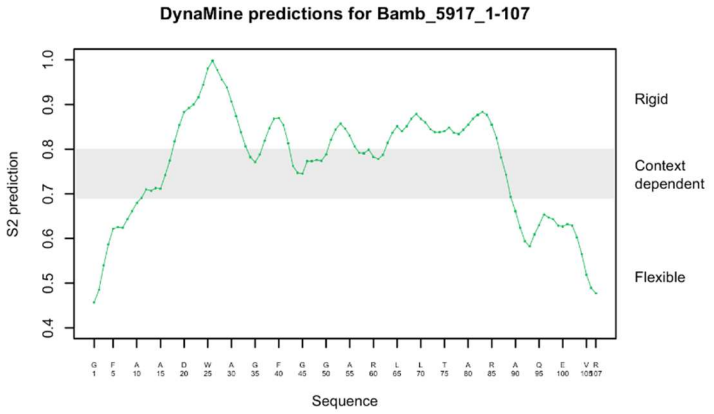


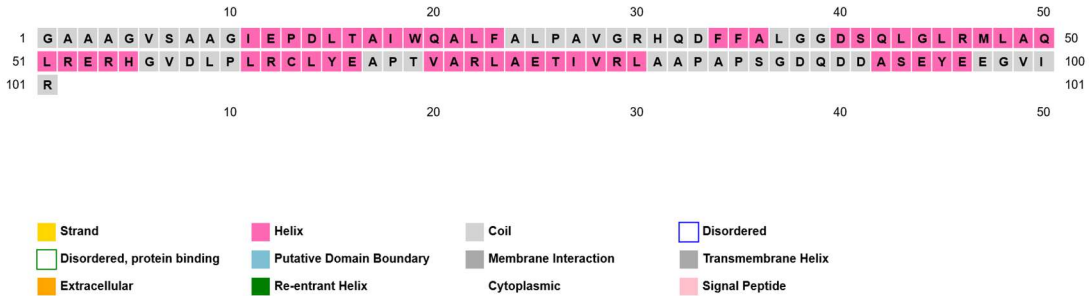
Appendix 11: Dynamine and PsiPred 4.0 predictions of carrier protein domains

AerA(PCP)

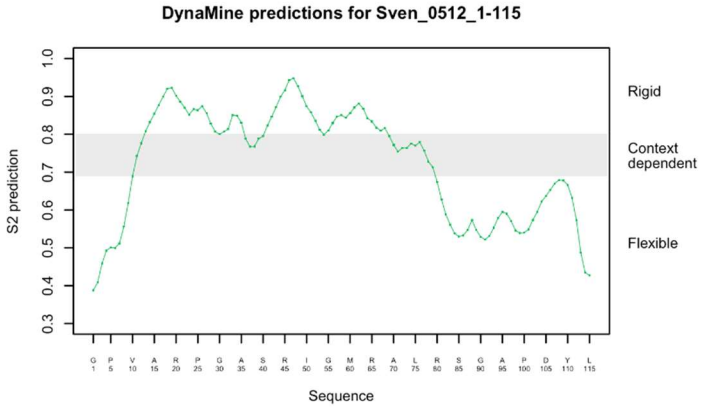


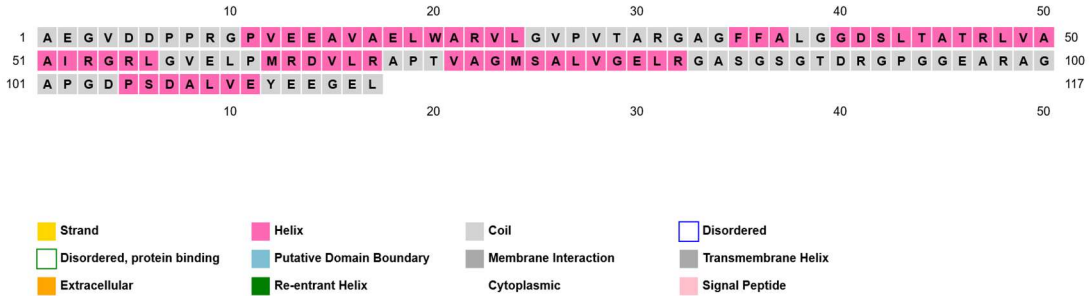
Bamb_5917(PCP)





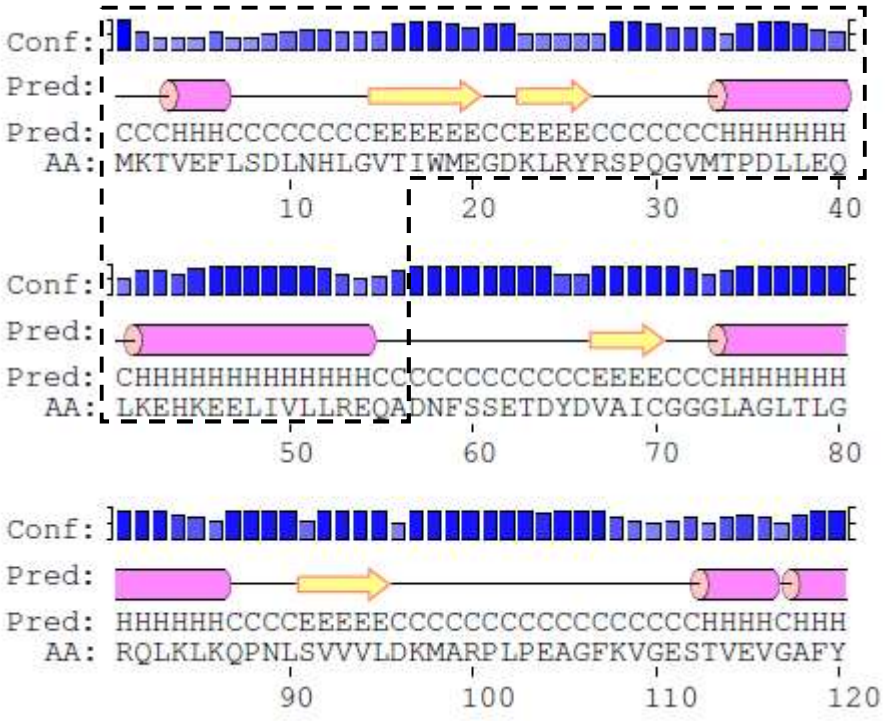
Sven_0512(PCP)





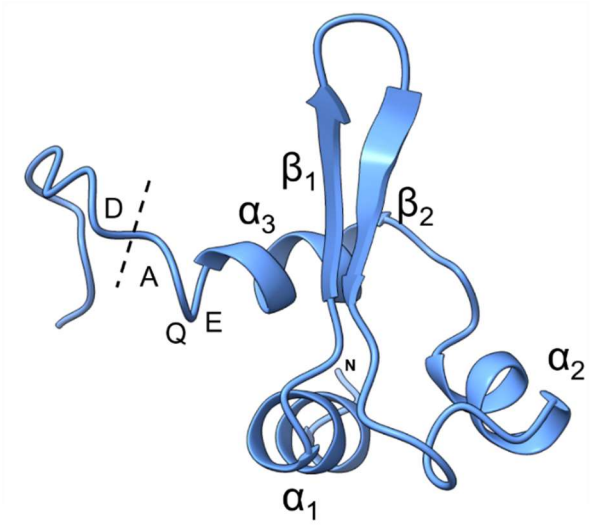
Appendix 12: Psipred 4.0 secondary structure prediction and Phyre2 N-terminal homology model of halogenase AerJ for determining the site for βHDD excision

PsiPred 4.0



βHDD boundary chosen by presence of αββαα topology

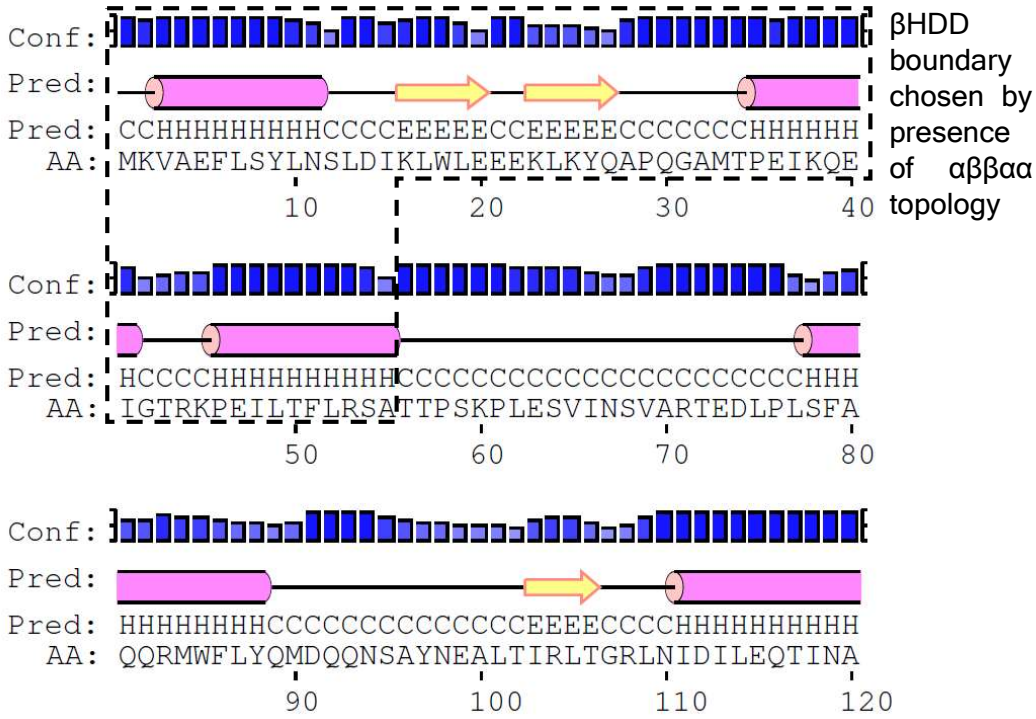
Phyre2 homology model



Note: Phyre2 homology model was generated using the TubC βHDD (PDB accession = 2JUG) which was given the best confidence of homology ranking by Phyre2.

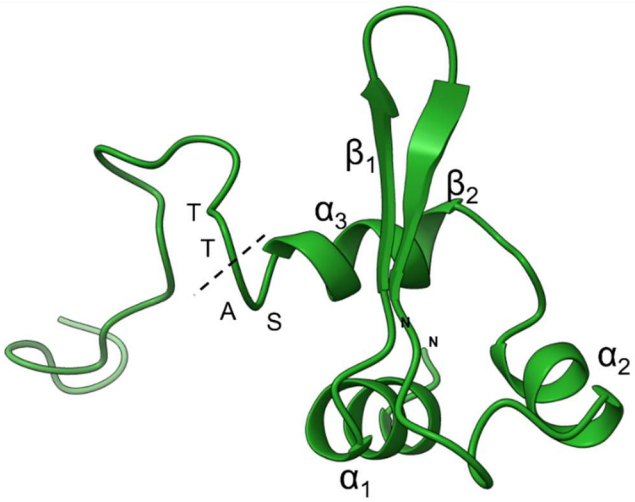
Appendix 13: Psipred 4.0 predictions and Phyre2 N-terminal homology model of AerB(C) for determining the site for βHDD excision

PsiPred 4.0



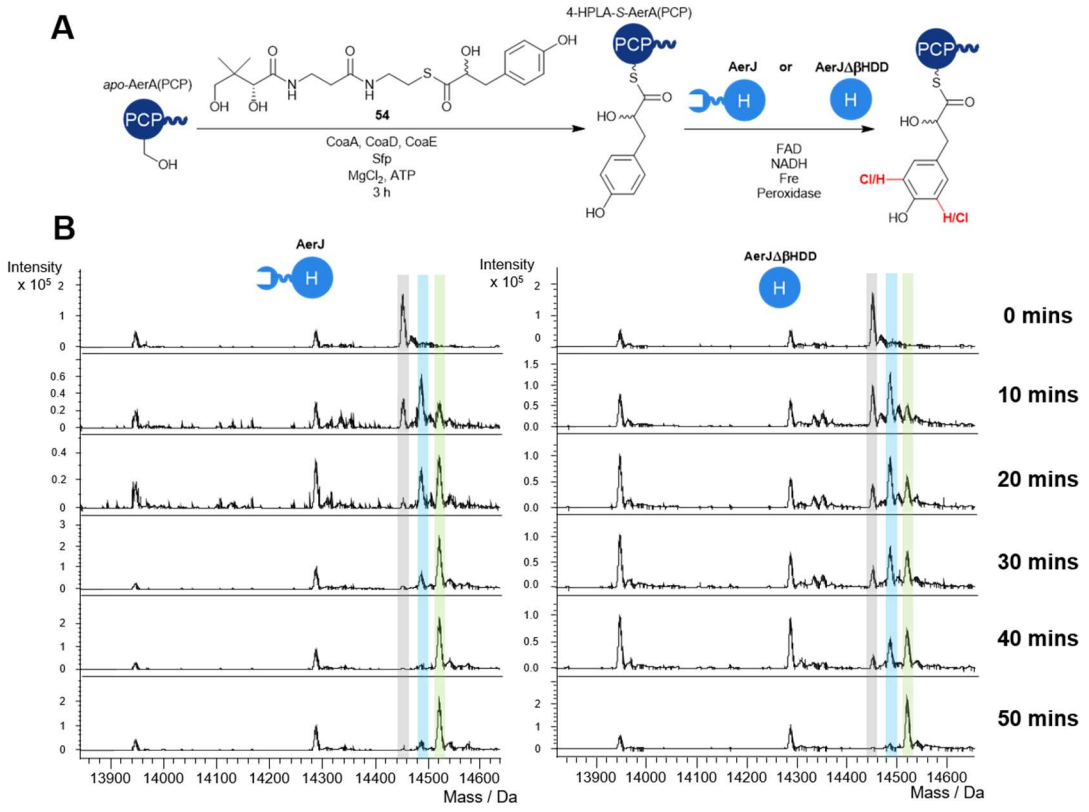
βHDD boundary chosen by presence of αββαα topology

Phyre2 homology model

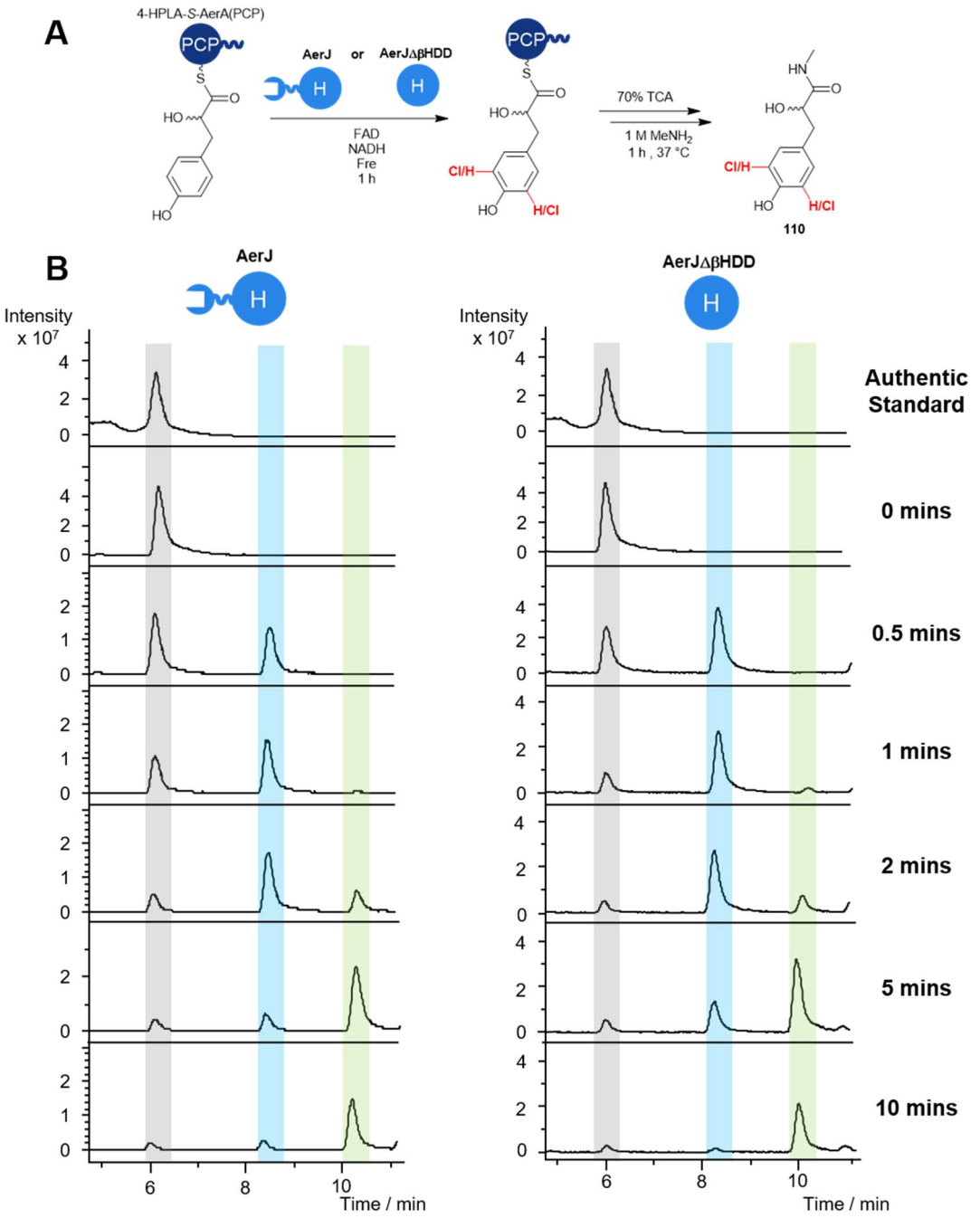


Note: Phyre2 homology model was generated using the TubC βHDD (PDB accession = 2JUG) which was given the best confidence of homology ranking by Phyre2.

Appendix 14: Attempted comparison of the rate of chlorination of 4-HPLA-*S*-AerA(PCP) by AerJ and AerJΔβHDD.



Note: This experiment was performed and analysed as described in section 6.1.6.9.



Grey = non-chlorinated **110**, blue = mono-chlorinated **110**, green = di-chlorinated **110**.

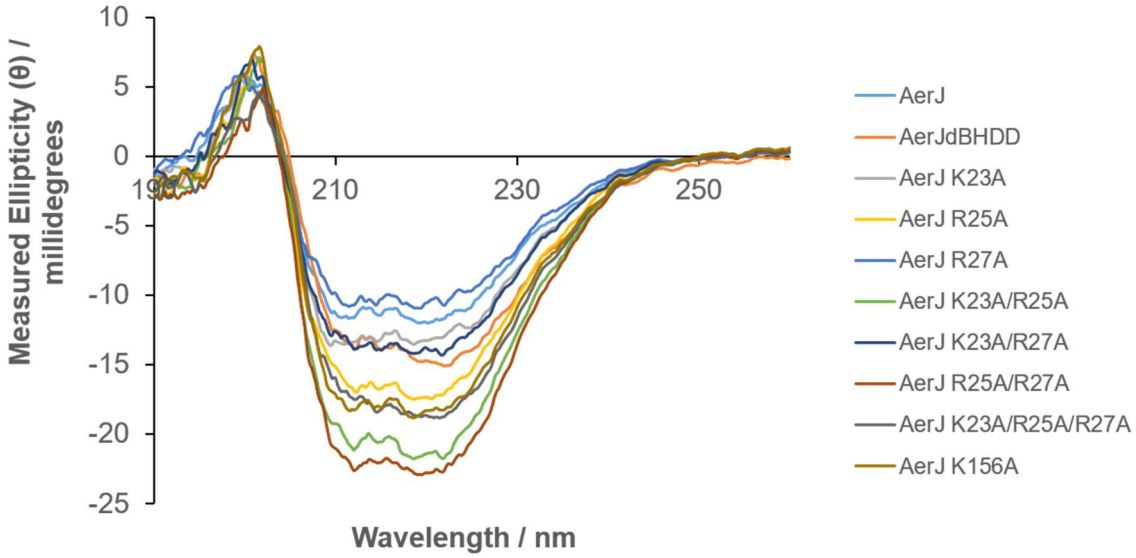
Note: The non-chlorinated authentic standard was synthesised as described in section 6.2.2.15. The biochemical reaction was performed and analysed as described in section 6.1.6.10.

Appendix 15: Comparison of the concentrations of AerJ mutants measured by Nanodrop spectrophotometry and Bradford assay

Halogenase Point Mutant	Concentration of stock determined by Nanodrop measurements / mg mL^{-1}	Ratio of concentrations determined by Nanodrop measurements	Concentration of stock determined by Bradford assay / mg mL^{-1}	Ratio of concentrations determined by Bradford assay
WT AerJ	30.5 ± 3.0	1.00	23.0 ± 2.4	1.00
AerJΔβHDD	21.3 ± 0.4	0.70	15.4 ± 0.6	0.67
AerJ K23A	38.2 ± 0.3	1.25	37.3 ± 2.0	1.63
AerJ R25A	47.2 ± 1.0	1.55	47.2 ± 4.6	2.06
AerJ R27A	96.9 ± 7.9	3.18	82.1 ± 2.9	3.57
AerJ K23A/R25A	21.2 ± 0.1	0.70	35.4 ± 6.3	1.54
AerJ K23A/R27A	24.5 ± 5.6	0.80	13.6 ± 3.9	0.59
AerJ R25A/R27A	19.8 ± 2.7	0.65	21.2 ± 2.3	0.93
AerJ K23A/R25A/R27A	20.7 ± 0.4	0.68	21.1 ± 0.9	0.92

Note: Nanodrop readings were measured from 60x dilutions of the stock solutions. Bradford assays were measured from 30x dilutions of the stock solutions. In both cases errors were calculated from the standard deviations of triplicate measurements.

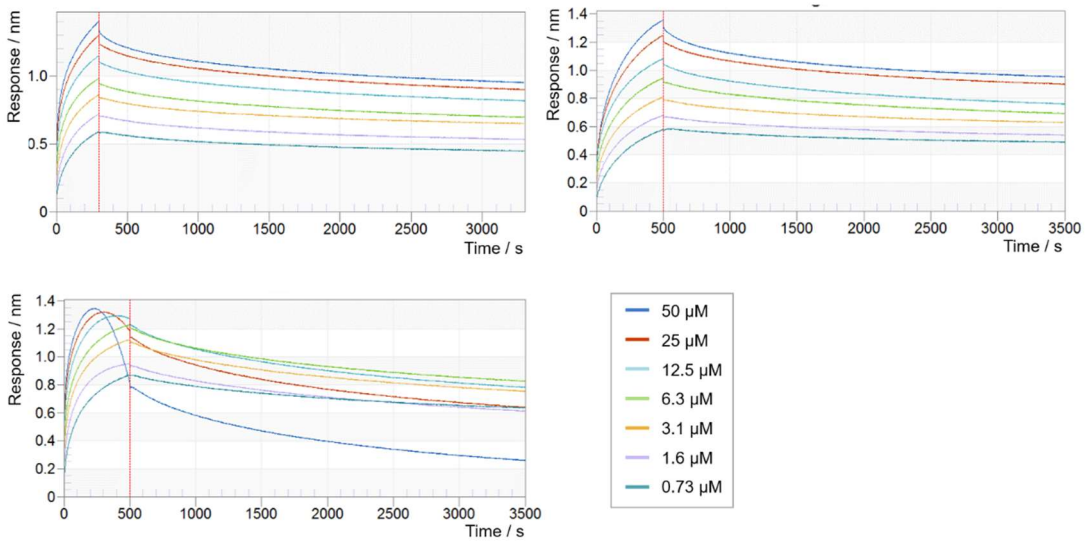
Appendix 16: Circular dichromism spectra of halogenase AerJ and mutants used in this work. Spectra were normalised with respect to a blank run.



Note: Spectra were normalised with respect to a blank run.

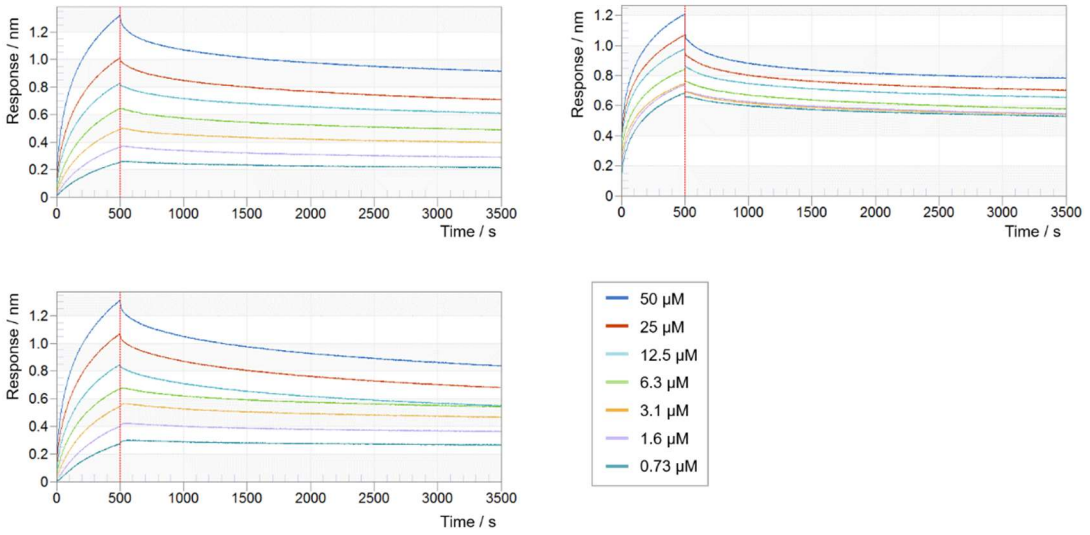
Appendix 17: Association and dissociation curves from Bio-Layer Interferometry titrations of AerA(PCP) with AerJ, AerJΔβHDD, AerB(C), AerB(C)ΔβHDD

AerA(PCP) + AerJ

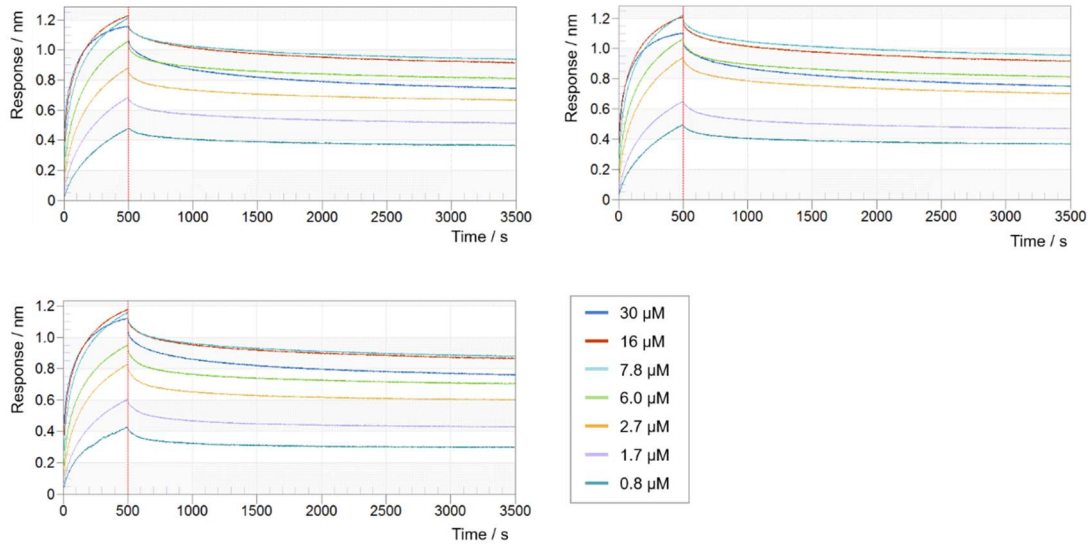


Note: *The 50 and 25 µM results for the 3rd replicate were not used for steady state fitting.*

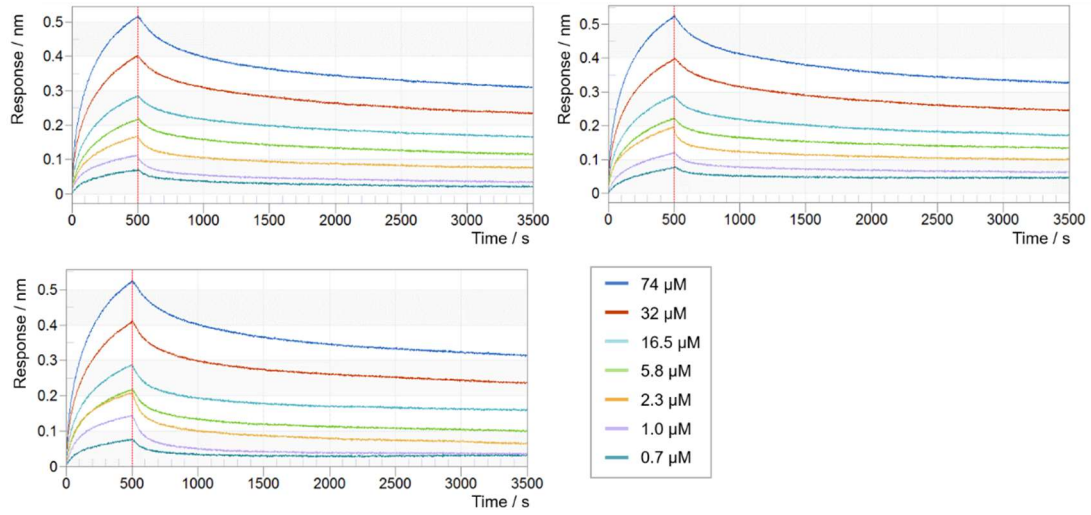
AerA(PCP) + AerJΔβHDD



AerA(PCP) + AerB(C)



AerA(PCP) + AerB(C)ΔβHDD



Appendix 18: Sequences of AerJ and AerA(PCP) displaying carbene footprinting sequence coverage

Detected, masked

Detected, no effect

Not detected

AerA(PCP)

```
         10         20         30         40         50         60
MHHHHHHHHG ENLYFQGSEL IAPRTEAERQ VAKIWQKVLN LSQIGIHDNF FELGGHSLLA

         70         80         90        100        110        120
SQVISRLRDV FSVELSLHSL LEYPTVASLT QTIEVLNVAN NSQSVSKSGK VMTTASENYE

EGEL
```

AerJ

```
         10         20         30         40         50         60
MHHHHHHHHG ENLYFQGSMK TVEFLSDLNH LGVTIWMEGD KLRYRSPQGV MTPDLLEQLK

         70         80         90        100        110        120
EHKEELIVLL REQADNFSSE TDYDVAICGG GLAGLTLGRQ LKLKQPNLSV VVLDKMARPL

        130        140        150        160        170        180
PEAGFKVGES TVEVGAFYLA NTLQLTDYFD HQHLPKLGLR YFFKPQETEF HKRPELGLSE

        190        200        210        220        230        240
FHAPKSYQID RGKLENDLRQ FNIEAGIDLK ENCSVKDIEF AEGLQQQHKI IYTQGSGANQ

        250        260        270        280        290        300
KTHCIKSRWV VDAMGRRRFI QKKLGLAKPN HNNYSAVWFR VEGRFDVSNF VPASEEKWHR

        310        320        330        340        350        360
RVPNNNRYYS TNHLCGEGYW VWLIPLSTGY TSIGIVARQD IHPLKNYHNY ELAFQWLREN

        370        380        390        400        410        420
EPVLAAHLEG KSPEDFRKMP KYSYSSKQVF SFNRWACVGE AGLFPDPFYS PGSDSIGFGN

        430        440        450        460        470        480
SLTTQMIELD LKGQLTPERV DDANHFYLTY HDGLTFNIQN SYNCMGNGIV MATKMIWDTL

        490        500        510        520        530        540
AGWTFGCLMM FNSIFLDPEL KMKVQQINAE FFPLSYRIQQ LFRDWANQSL GRVSFEFIDY

        550        560        570        580        590        600
LAIPFVNELR TRNLQSNKTE AEIIEAYRAS LKLLEEFAQV IFQIALEDTS PELLTKINEH

        610        620        630        640
PWLNAWAISL DASKWETDGL FSPKNEPRNL HSIKKQYLEA INR
```