Engineering and Characterisation of Bacterial Phosphopantetheinyl Transferases and their Peptide Substrates

Jack Alexander Sissons

2016-17

Submitted in fulfilment of the requirements for the degree of

Master of Science in Biotechnology

Victoria University of Wellington

Supervisor: Dr. David Ackerley

Abstract

Throughout all domains of life, phosphopantetheinyl transferase (PPTase) enzymes catalyse a post-translational modification that is important in both primary and secondary metabolism; the transfer of a phosphopantetheine (PPant) group derived from Coenzyme A to specific protein domains within large, multi-modular biosynthetic enzymes, thereby activating each module for biosynthesis. The short peptide motif of the protein to which this group is attached is known as a 'tag', and can be fused to other proteins, making them also substrates for post-translational modification by a PPTase. Additionally, it has been demonstrated that PPTases can utilise a diverse range of CoA analogues, such as biotin-linked or click-chemistry capable CoA derivatives, as substrates for tag attachment. Together, these characteristics make post-translational modification by PPTases an attractive system for many different biotechnological applications. Perhaps the most significant application is *in vivo* and *in vitro* site-specific labelling of proteins, for which current technologies are hindered by cumbersome fusion protein requirements, toxicity of the process, or limited reporter groups that can be attached. Confoundingly, most PPTases exhibit a high degree of substrate promiscuity which limits the number of PPTase-tag pairs that can be used simultaneously, and therefore the number of protein targets that can be simultaneously labelled. To address this, directed evolution at a single gene level was used in an attempt to generate multiple PPTase variants that have non-overlapping tag specificity which have applications in orthogonal labelling. Furthermore, assays for the rapid identification, characterisation and evolution of short, novel peptide motifs that are recognised by PPTases has further diversified the labelling toolkit. These developments have enhanced the utility of the PPTase system and potentially have a wide range of applications in a number of fields.

Acknowledgements

The first acknowledgement must go to my supervisor Dr. David Ackerley. Your superb lectures with their half-time funnies, and laboratories made up the most engaging and fascinating classes I've taken, and your genuine passion for the subjects came across and hooked many of us too. Thank you so much for letting totally green JMK run amok in the lab and then put our names on a published paper! I wonder how long the "when the undergrads get published before you" card hangs around in the cards against humanity set. And thank you for all your support and encouragement over the past two years. The lab environment you've created is second to none and it is an absolute pleasure to be part of it.

Thank you to Alistair and Katherine, the pyo pals, for all your help, particularly in the first six months when I couldn't do a miniprep right and was intimidated by the French Press. Alistair, its super cool how passionate you are about research, thanks for always spit balling ideas with (mostly at) me.

An enormous thank you to Kate. You've kept my chin up throughout 2016 even when nothing worked and things weren't so fun. It feels like we achieved this. Oh, and thanks for taking interest in all my pretty graphs that are(n't) so exciting.

Everyone who has been in the KK702 office, you're a great gang. Jasmine, Jacob, Euan, Izzie, Mitch, Varun (honorary), Vaughan, Melz, James, Sonja, Kyle, Diana, Sarah, Jen. I'll miss the provocative statements from Vaughan/Euan and the office-wide debates that ensue, the Crow Brothers, the star chart that Vaughan "doesn't care about", the Nut Council and its revoking of nut farmer Kyle's G-series aircrafts. And the chess. Big thanks to the rest of the Ackerlytes, good luck for the rest of your time in the lab and beyond.

Table of contents

List of Figures

List of Tables

Abbreviations used in this thesis

aa	Amino acid
ACP	Acyl carrier protein
ATP	Adenosine triphosphate
bp	Base pair
CoA	Coenzyme A
CP	Carrier protein
ddH_2O	Double distilled water
DMSO	Dimethyl sulfoxide
EtBr	Ethidium bromide
FMN	Flavin mononucleotide
Fw	Forward primer
GYT	Glycerol, yeast extract, tryptone
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LBA	Lysogeny broth & agar
MCS	Multiple cloning site
NRPS	Non-ribosomal peptide synthetase
OD_{600}	Optical density at 600 nm
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PPTase	Phosphopantetheinyl transferase
PTM	Post-translational modification
rev	Reverse primer
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TE	Tris-EDTA
WT	Wild-type

Chapter 1: Introduction

1.1 Phosphopantetheinyl transferases; function and classification

Since their first description as an enzyme superfamily in 1996 (Lambalot et al., 1996), study of phosphopantetheinyl transferases (PPTases) has grown into its own field. Analysis of their role in natural cell processes has not only yielded understanding of how these processes take place, but has also given rise to a number of potential biotechnological applications (Beld et al., 2014). These factors together make PPTases an exciting area of study.

PPTases are implicated in a large number of biological processes throughout all three domains of life (Beld et al., 2014). In general, they catalyse the transfer of a 4'-phosphopantetheine arm from coenzyme A (CoA) to a target carrier protein, via a phosphoester linkage to a conserved serine residue within the carrier protein (Mercer et al., 2007). The carrier proteins to which this PPant arm is added can reside within fatty acid synthases, non-ribosomal peptide synthetases (NRPSs) and polyketide synthases, or it can be a free standing enzyme (Paul et al., 2017). The post-translational modification undertaken by the PPTase upon these multi-modular biosynthetic enzymes converts them from their inactive *apo*- form to the active *holo*- form, permitting them to then carry out their function. Within each of these systems the PPant attachments act as "swinging arms" which pass the growing product from module to module, via thioester linkages throughout the elongation process (Beld et al., 2014). For example, within NRPS enzymes a PPant arm is attached at each peptidyl carrier protein domain, also known as the carrier protein (CP) or thiolation (T) domain (Figure 1.1). These arms receive an activated monomer from an associated adenylation domain and position it so that activated monomers from adjacent modules meet in the active site of a condensation domain. Covalent linkage of the two monomers (usually via

peptide bond formation) is catalysed by the condensation domain and the downstream PCP domain "swings" the newly formed intermediate onwards via its PPant arm to the following module of the NRPS. Once all constituents are joined, this final product is released from the terminal carrier protein domain by either hydrolysis or intramolecular cyclisation, catalysed by a thioesterase domain (Finking et al., 2002). An example of this system is the blue pigment synthetase encoding gene (*bpsA*) which encodes the single module NRPS, BpsA, of *Streptomyces lavendulae*. BpsA cyclises and joins two L-Glutamine monomers together to form indigoidine, a blue pigment (Takahashi et al., 2007). Many products synthesised by these three kinds of megasynthases are essential for life (e.g. fatty acids), thus some PPTases play an essential role in cell viability. In addition to the primary metabolites formed, numerous secondary metabolites also arise from these systems. These include pigments, antibiotics and antioxidants amongst others.

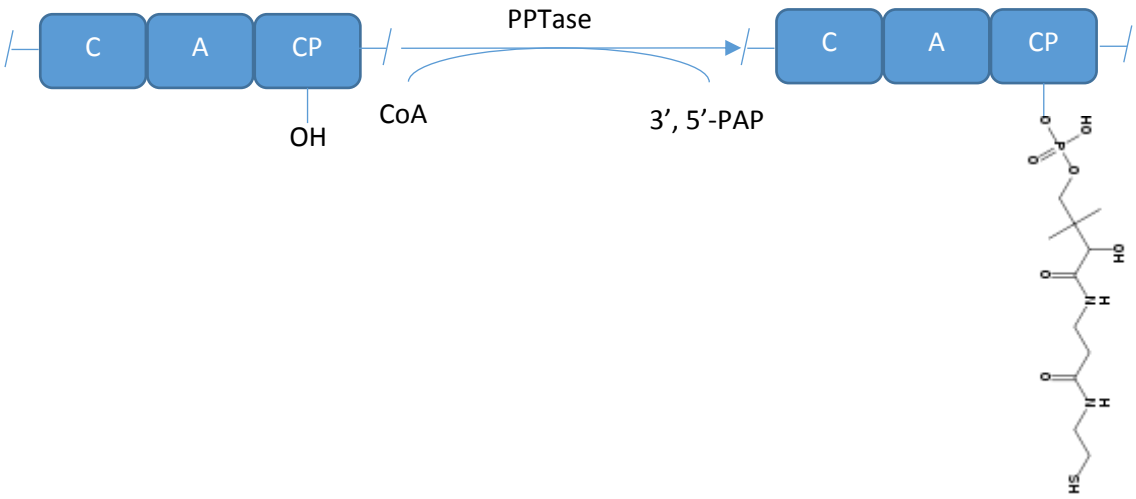

Figure 1.1: PPTase mediated addition of a 4'-phosphopantetheine arm from CoA to the carrier protein domain of a NRPS module.

3'-5'-phosphoadenosine phosphate (3'-5'-PAP) is produced as a by-product. C: Condensation domain, A: Adenylation domain, CP: carrier protein domain.

The diversity present in the CPs of NRPS, PKS and FAS enzymes is understandably mirrored by diversity in the amino acid sequences and tertiary structures of the PPTases able to catalyse their activation. This diversity manifests itself in three designated families.

1.1.1 Family 1: ACP synthase (AcpS)-type PPTases

The first PPTase to be described was the acyl carrier protein synthase (AcpS) of *Escherichia coli*, which catalyses the addition of a 4'phosphopantetheine arm to the carrier proteins of type II fatty acid synthase (Elovson et al., 1968), and typifies the first family of PPTases. These PPTases are typically involved in primary metabolism. The reaction catalysed by these enzymes requires Mg^{2+} or Mn^{2+} as a cofactor (Elovson et al., 1968). AcpS monomers are 120 aa in size and arrange into a functional homotrimer (Figure 1.2), at the interfaces of which lie the active sites (Beld et al., 2014). Initially, the enzyme binds Mg^{2+} and CoA, then the acyl carrier protein. Structurally, AcpS-type PPTases are the best understood of the families. The quaternary structure of *Bacillus subtilis* AcpS is largely stabilised by hydrophobic interactions, particularly from Ile5 on β-sheet 1 and residues on β-sheet 5. Hydrogen bonding networks provided by Gln113 residues at the centre of the trimer also play a significant role in stabilisation (Parris et al., 2000). Following discovery of *E. coli* AcpS, an AcpS-like enzyme was isolated from plants and shown to play a similar role in FAS (Elhussein et al., 1988). These findings indicated the importance of this family of PPTases in multiple domains of life. Following the ground-breaking Lambalot study describing PPTases as an enzyme superfamily (Lambalot et al., 1996), the same lab went on to show that *E. coli* AcpS had activity with several CPs from foreign fatty acid synthases (Gehring, et al., 1997). This promiscuity is of particular interest when considering potential applications of PPTases and is expanded upon in Section 1.4.

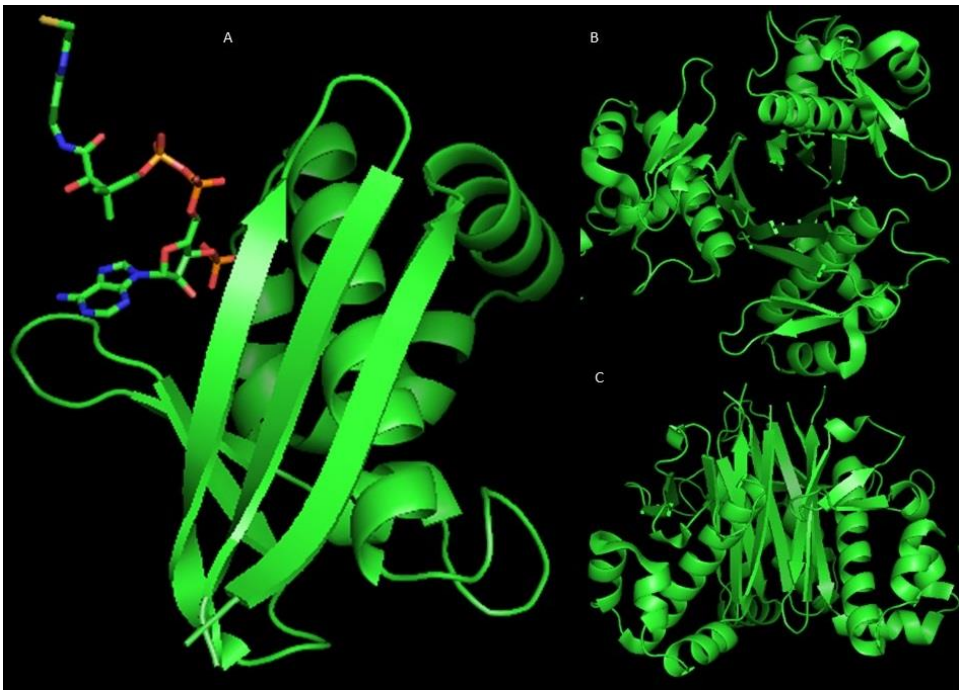

Figure 1.2: A: Monomeric AcpS in complex with CoA (PDB ID. 1F7L). B and C: Homotrimeric assembly of AcpS as viewed from two different angles (PDB ID. 1F7T).

1.1.2 Family 2: Surfactin-like phosphopantetheinyl transferase (Sfp)-type PPTases

Sfp is the PPTase responsible for the addition of the PPant arm to the PCPs involved in surfactin synthesis in *B. subtilis* (Grossman et al., 1993; Quadri et al., 1998) and is archetypal of the second family of PPTases (Lambalot et al., 1996). The enzymes belonging to this family function as pseudo-homodimers of around 240 aa (Figure 1.3). The genes responsible for these PPTases tend to reside within or nearby the biosynthetic operons that they activate, although this is not a strict rule. The Sfp-type PPTases are similar in appearance to an AcpS dimer, however they tend to exhibit much broader substrate specificity (Beld et al., 2014). EntD, the PPTase essential for synthesis of the siderophore enterobactin in *E. coli,* also belongs to this family. EntD modifies EntF, the NRPS involved in the biosynthetic pathway (Lambalot et al., 1996), as well as EntB, an isochromate lyase also belonging to this synthetic pathway (Beld et al., 2014). It has been demonstrated that EntD from *E. coli* and Sfp from *Bacillus subtilis* are interchangeable in the surfactin and enterobactin synthetic pathways (Grossman et al., 1993). Not only does Sfp accept many CP targets, it also

permits catalysis using a range of CoA analogues (Belshaw et al., 1999). This finding made the promise of putting these enzymes to biotechnological use all the more enticing.

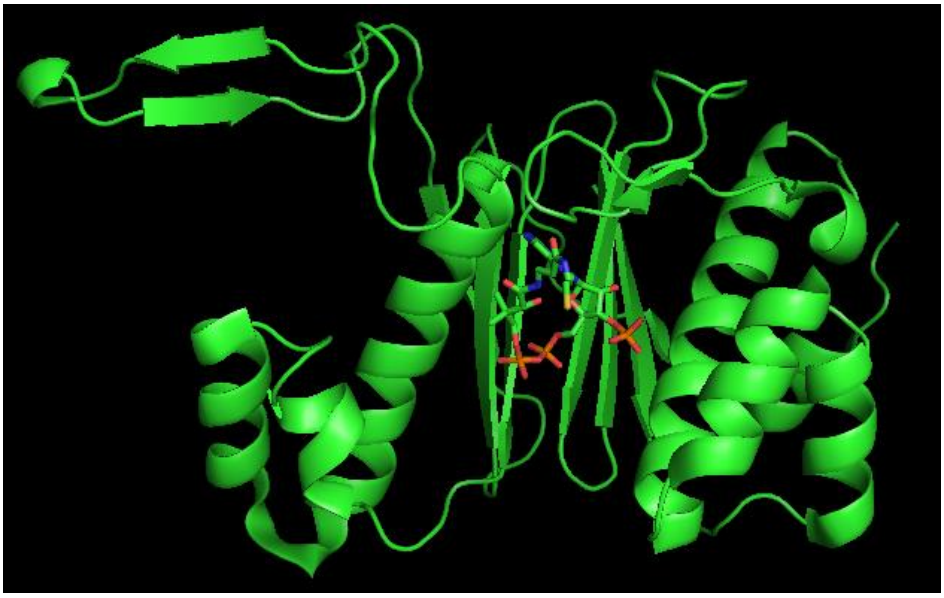

Figure 1.3: Pseudodimer of Sfp in complex with CoA (PDB ID. 1QR0).

Interestingly, the sole PPTase of *Pseudomonas aeruginosa*, PcpS, initially appears to belong to this family due to its monomeric nature and 242 aa size. Despite this, PcpS shares only 13.9% sequence identity with Sfp and is not encoded in an NRPS cluster. PcpS is the sole PPTase present in *P. aeruginosa* and has been found to catalyse modifications in both primary and secondary metabolism. It therefore belongs to its own subclass of PPTases (Finking et al., 2002).

1.1.3 Family 3: Type 1 integrated PPTases

In *Saccharomyces cerevisiae,* PPTases have been found translationally fused to cytosolic type I fatty acid synthase at the C-terminus. This arrangement is not conserved in other eukaryotes. These fatty acid synthase systems are large barrel shaped assemblies, at the centre of which synthesis occurs. Curiously, the PPTase domains are located on the outside of the barrel, separate from the CPs. It is proposed that during assembly, a transient dimer of PPTase domains forms and

performs the required modifications prior to assembly of the FAS complex (Johansson et al., 2009). When expressed and purified, these PPTase domains assemble into homotrimers similar to those seen with AcpS-type PPTases. Curiously, the PPTase domains of fully assembled fatty acid synthases retain activity with free CPs *in vitro*, and therefore may have other activities *in vivo*. This family of PPTases is the least well understood of the three identified to date.

1.1.4 PPTase classification

Although PPTases can be highly diverse in terms of their primary amino acid sequence, bioinformatic tools allow us to identify these enzymes and the family to which they belong with a high degree of confidence. It has been identified that each family of PPTases generally contains a signature sequence comprising two motifs: (V/I)G(V/I)D(x)$_{40–45}$(F/W)(S/C/T)xKE(A/S)hhK, where x is any amino acid and h is an amino acid with a hydrophobic sidechain (Lambalot et al., 1996) and, more specifically, bacterial PPTases contain the sequences (I/V/L)G(I/V/L/T)D(I/V/L/A)(x)n(F/W)(A/S/T/C)xKE(S/A)h(h/S)K(A/G) where n is 42–48 for the AcpS subfamily of PPTases and 38–41 for the Sfp-like PPTases (Asghar et al., 2011). Furthermore, Sfp-type PPTases can be divided into two subclasses by presence of either the WxxKEA or FxxKES motif within the larger signature sequence (Beld et al., 2014).

1.2 PPTase substrate specificity

1.2.1 Carrier protein specificity

Different PPTases exhibit varying degrees of specificity with regard to the carrier protein that they modify. For example, AcpS is relatively specific to CP domains of fatty acid synthase enzymes (Gehring, et al., 1997), whereas Sfp and EntD exhibit far greater CP promiscuity (Lambalot et al.,

1996). One study involving co-crystal structure and mutational analysis has elucidated numerous specific residues that play roles of varying importance in Sfp-CP interaction and recognition, particularly via hydrophobic interactions (Tufar et al., 2014). The core sequence to which the PPant arm is attached on the CP is a highly conserved DSL, the hydroxyl group of the serine sidechain bearing the arm via a phosphoester bond (Lambalot et al., 1996). The D of this triplet tolerates more variation than the S and L residues and is influenced by the location of the CP domain within the NRPS (Beld et al., 2014; Linne et al., 2001). The sequence surrounding the DSL contributes significantly to recognition, particularly a hydrophobic stretch of 4-6 amino acids following the DSL, as well as other tertiary features (Beld et al., 2014; Tufar et al., 2014). AcpS cannot act upon a PCP in nature, however, following the substitution of 14 aa from an acyl carrier protein domain into helix II of a CP domain, AcpS was found to be able to activate the hybrid (Finking et al., 2002). This indicates that there are residue specific interactions that allow interaction between the PPTase and the CP domain.

1.2.2 Carrier protein 'tag' regions

In an attempt to identify native substrates of Sfp, the Walsh laboratory created a phage display library from a fragmented *B. subtilis* genome (Yin et al., 2005). This library was enriched for PPTase substrates by treatment with Sfp and biotin-linked CoA, allowing streptavidin isolation of phage particles expressing a peptide that was post-translationally modified by Sfp. By this method, several different fragments of the 484 aa protein YbbR were identified as Sfp substrates. The phosphopantetheinylated serine in the shortest (49 residues) fragment was identified and various short peptides containing this serine were synthesised in an attempt to define the minimal sequence capable of recognition by a PPTase. One such peptide was the 11 aa sequence DSLEFIASKLA termed the "ybbR tag" (Yin et al., 2005). This tag forms a helical structure and functions as an efficient substrate for Sfp when fused to the C or N terminus of a protein. Addition

of residues to the N-terminus of the tag was found to have no effect on the tag's ability to function as a Sfp substrate, whereas truncation of just three residues from the C-terminus abolished activity with Sfp (Yin et al., 2005). The amino acid following the L of the DSL sequence was also found to play a determining role on which PPTase type can modify the CP domain. Peptidyl carrier proteins bear a positively charged residue in this position, whereas acyl carrier proteins predominantly have a negatively charged Asp. All these interactions and many more contribute to the interaction and specificity of PPTases for CP domains. Since the development of the ybbR tag, a few other short peptide sequences from CPs have been generated as substrates for PPTases via extensive library screening approaches (Zhou et al., 2007), for example, the A1 tag, a 12 residue substrate for AcpS. Development of multiple tags that act as substrates for different PPTases has enabled development of orthogonal labelling strategies which will be discussed later.

1.2.3 Coenzyme A specificity

PPTases exhibit promiscuity with the CoA substrate that they can bind (Belshaw et al., 1999). While the phosphoadenylate region of CoA is essential for binding the PPTase, the pantetheine arm has been shown to not interact with the enzyme. For example, the hydrophobic pocket of the PPTase which cradles CoA also allows binding of analogues such as malonyl-CoA and acetyl-CoA (La Clair et al., 2004). With this knowledge, it has been reasoned and demonstrated (Callegari et al., 2012; Marchetti et al., 2014; Sunbul et al., 2009; Vivero-Pol et al., 2005) that certain PPTases can accept many modified versions of CoA, opening the door for diverse PPTase modification of many substrates with additions such as click-chemistry functional groups and fluorophores, among others. Click chemistry-modified CoA derivatives (Figure 1.4) are of particular interest because click-chemistry permits the addition of virtually any functionalised molecule to the target (Kolb et al., 2001), thus greatly enhancing the utility of the PPTase labelling system.

Figure 1.4: Click chemistry capable CoA analogues. Image obtained from (Beld et al., 2014).

1.3 Measuring PPTase activity

Until recently, PPTase activity has been measured using cumbersome or time consuming techniques such as HPLC or radiolabelling (Lambalot et al., 1996; Mootz et al., 2001). However, using BpsA as a reporter, the Ackerley lab has developed a simple enzymatic assay to analyse the activity of a given PPTase (Owen et al., 2011). Conversion of BpsA from the *apo* to the *holo* form by a PPTase allows NRPS activity, converting L-Glutamine to detectable indigoidine. By measuring the rate of indigoidine synthesis, the kinetic parameters of the PPTase with BpsA can be determined. This rapid and flexible assay can also be adapted to assess activity of PPTase inhibitors or as a competition assay using alternative PPTase substrates (e.g. foreign carrier proteins) alongside BpsA to compete for a limited pool of CoA (Figure 1.5).

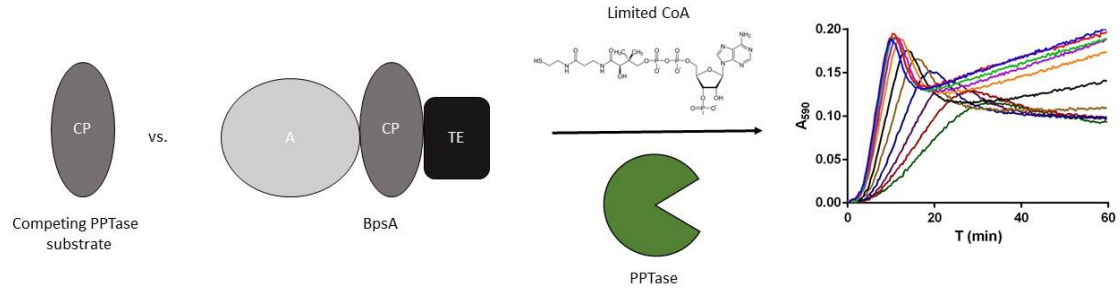

Figure 1.5: Schematic of a competition assay involving BpsA for the investigation of alternative PPTase peptide substrates.

An alternative PPTase substrate is titrated against BpsA so that the two proteins compete for PPTase activity in the presence of a limited pool of CoA. Indigoidine absorbance is measured at A_{590} and plotted against time. If the alternative substrate is able to compete, this is observed as a decrease in indigoidine production. IC_{50} values for inhibitors can be calculated from the measurements obtained at different inhibitor concentrations.

1.4 Biotechnological applications of PPTases

The ability of certain PPTases to add a large variety of small molecules to a range of carrier proteins has made them an attractive platform for a number of *in vitro* and *in vivo* applications, the majority involving site-specific labelling of target proteins (Beld et al., 2014). Other labelling techniques such as GFP fusion and biarsenical conjugated fluorophores tend to have the issue that the reporter group which must be fused into the target protein is too large for many applications, or the reactions involved are too slow, non-specific, toxic or labour-intensive (Muir et al., 1998; Sunbul et al., 2009; Yin et al., 2004). The PPTase system is attractive because, in theory, one can work down to a small (8-12 aa) tag region (such as the ybbR tag) that can then be integrated into a target protein. Subsequently, any small molecule conjugated to CoA can be added by a PPTase. Interestingly, quantum dots have also been conjugated (Sunbul et al., 2009), demonstrating the flexibility that is possible. Furthermore, the reaction is a single enzymatic step and gives a

quantitative result in the form of a 1:1 target protein to post-translational modification ratio, allowing straightforward optical analysis of the labelling reaction.

In 2004, Sfp was used with fluorescent and affinity-capable (biotin for capture by streptavidin, and maltose for capture by maltose-binding protein) CoA analogues to modify the CP domains of modular synthases and thereby visualise, identify by western blot, and affinity-purify those modular synthases *in vitro* (La Clair et al., 2004). This was a notable achievement because modular biosynthetic enzymes are difficult to purify and work with due to their large and dynamic nature. This study largely adhered to the system nature has evolved; the only artificial component being the CoA analogues. This, in part, played the role of a proof of concept experiment before modification of further variables in the system was explored. Yin et al delved deeper into the system by cloning the PCP domains from the NRPSs EntB and GrsA onto the N-termini of fluorescent target proteins (Yin et al., 2004). Biotin-linked CoA was then added to these proteins using Sfp, and addition of the biotin was analysed by enzyme linked immunosorbent assay (ELISA) using streptavidin. Fusion proteins were also expressed in liquid culture in 96-well plates to test whether the labelling method could be performed in a high throughput fashion. Fluorescent detection confirmed the incorporation had taken place and western blot confirmed that only the target proteins had been labelled. The labelling reaction was 95% complete after 30 minutes, highlighting the efficiency of the system. The tag was shown to not interfere with the activity of the enzyme, and furthermore, removal of a $(Ser\text{-}Gly\text{-}Gly\text{-}Gly\text{-}Gly)_3$ linker between the tag and the target protein that was used in initial tests did not alter activity, indicating a linker may not be necessary (Yin et al., 2004). Subsequently, the same group published results describing the “ybbR tag” mentioned previously in Section 1.2.2. Phage display of the *B. subtilis* genome using biotin conjugated CoA analogues to capture any phosphopantetheinylated peptides via streptavidin binding identified this 11 aa sequence DSLEFIASKLA originating from the putative YbbR protein

(Yin et al., 2005) which has an unknown function in nature (Barb et al., 2011). The identification of this short, specific sequence brought the PPTase system for site-specific labelling to a competitive standard relative to alternative labelling strategies because it showed the remainder of the relatively large carrier protein domain (80-100 aa) was not necessary for labelling. In 2007 the same lab expanded upon this work by identifying new tag sequences, again by phage display. This time, however, parallel phage display selections were used with each of Sfp and AcpS in order to find substrates that might be used orthogonally (i.e. could be expressed simultaneously to exclusively label different substrates). The search for orthogonality yielded a promising 12 aa peptide sequence for each enzyme, termed A1 and S6, which were relatively specific for AcpS and Sfp, respectively (Zhou et al., 2007). When NMR analysis was used to compare A1 alone vs. the A1:AcpS complex the specific interacting residues were identified, and from this an even smaller 8 aa tag was developed, termed A4 (Zhou et al., 2008). These peptides have been utilised more recently to track two membrane bound neutrophin receptors, TrkA and P75NTR, and nerve growth factor simultaneously, allowing study of their specific interactions (Marchetti et al., 2014). By use of phosphopantetheine analogues in culture media, intracellular labelling has also been achieved (Clarke et al., 2005). These analogues are designed to be taken up by the cell and incorporated into the CoA biosynthetic pathway, yielding a substrate for a PPTase to post-translationally modify proteins.

One of the reasons for interest in PPTases initially was their promiscuity with carrier protein substrates, however we now see potential advantages for multiple biotechnology applications in having a larger range of PPTase pairs with non-overlapping carrier protein tag specificity in order to add degrees of orthogonal labelling.

In addition to site-specific labelling, the PPTase-CoA-tag system has been adapted for the covalent immobilisation of recombinant proteins onto solid supports conjugated to CoA (Wong et al.,

2008). The rapid, simple, and quantitative nature of this system makes it very practical in the design and development of protein microarrays, biosensors and bioassays, amongst others. Another innovative application of PPTases was demonstrated by Mosiewicz et al. to create high performance hydrogels for potential use in 3D cell culture and tissue engineering (Mosiewicz et al., 2010).

A development pertinent to all of the above applications has been described in which the native *P. aeruginosa* phosphodiesterase, acyl carrier protein hydrolase (AcpH) is used to remove the PPant (or analogue) modification from an acyl carrier protein. This was shown to also be possible with acyl carrier protein-fusion proteins. Phosphopantetheine removal by AcpH was shown to leave behind a functional, fully folded acyl carrier protein which could be re-labelled (Kosa et al., 2014). Using this method, one can obtain uniformly labelled acyl carrier proteins by ensuring any undesired pantetheinylation is removed prior to addition of the label, a technology which has recently been patented (Burkart et al., 2015). This finding is valuable also because it allows iterative labelling of carrier proteins or carrier protein-tagged fusion proteins.

These adaptations of the natural PPTase mediated post-translational modification of carrier proteins have set a solid base for further development and innovation with the system. They highlight the strengths of the system as well as areas which require further investment of time and effort in order to maximise its utility.

1.5: Aims of this study

As mentioned earlier, the PPTase system for conjugating small molecules to recombinant proteins could greatly benefit from a larger range of tag sequences and PPTase variants that exhibit non-overlapping specificities for those tag sequences. For example, this would permit additional and/or more selective orthogonality for site-specific labelling. The major aim of this research was to engineer non-overlapping specificity by evolving two different PPTases to each be highly selective for two different carrier protein (CP) domain substrates. This was pursued via directed evolution, namely using error-prone PCR at a single gene level (Gillam et al., 2014), followed by screening and selection in an attempt to identify enhanced enzyme variants. The screening process utilised here was based on a pilot study performed by Dr. Katherine Robins of the Ackerley laboratory. This screening system was designed to identify EntD mutants which have lost the ability to recognise the BpsA CP domain as a substrate, while still recognising their native CP substrates, EntF and EntB (Robins, 2016). EntB and EntF are components of enterobactin synthase responsible for synthesis of the *E. coli* siderophore enterobactin (Lambalot et al., 1996). Therefore, Δ*entD E. coli* can be transformed with a gene encoding an EntD variant and plated on iron-restricted media to select for bacteria still capable of producing enterobactin. Our reasoning was that any transformants producing *holo*-BpsA were expected to appear blue as a result of indigoidine synthesis and were not selected. Katherine Robins demonstrated that this dual positive and negative screening strategy can be used to identify colonies that grow but do not produce indigoidine (i.e. appear white), indicating that they retained activity with EntF but no longer recognised the CP of BpsA.

Following Dr. Robins' preliminary work demonstrating proof-of-principle for this strategy, my goal was to expand the approach to evolve a non-overlapping (orthogonal) pair of PPTases originating from *E. coli* and *P. aeruginosa*, i.e. EntD and PcpS. More specifically, the intention was to generate

EntD and PcpS variants using directed evolution and then identify EntD mutants that had lost activity with the CP of PvdD, and PcpS mutants which lost activity with the CP of EntF, while both retaining activity with their native CP substrates. This was made possible via use of two BpsA variants that were previously generated by Dr. Jeremy Owen of the Ackerley laboratory, in which the native CP domain has been replaced with the CP domain of either EntF or PvdD, and termed BpsA(E) and BpsA(P), respectively (Owen, 2010). Like enterobactin, pyoverdine is a siderophore (Meyer et al., 1997), thus the above screening strategy accommodated this PPTase. This allowed screening in the same manner as described above; by selecting for growth without indigoidine production; in *P. aeruginosa* for PcpS evolution, and *E. coli* for EntD evolution.

Following this, the next goal of this study was to design and validate short peptide tag sequences derived from the native substrates of each PPTase and to assess how well these tags could perform as PPTase substrates *in vitro* by testing their ability to competitively inhibit BpsA by competing for PPTase activity. These experiments served to demonstrate a novel and simplistic method for designing such tags without a requirement for large randomised peptide libraries as have been used in previous studies (Yin et al., 2005; Zhou et al., 2007).

The aims of this research were:

1) *To use directed evolution to generate, identify and characterise an EntD variant that retains activity with the EntF CP domain but cannot activate the PvdD CP domain.*
2) *To generate by directed evolution, using an orthogonal screen in P. aeruginosa, a PcpS variant that retains activity with the PvdD CP domain but cannot activate EntF.*
3) *To design, validate and characterise short tag sequences within EntF and PvdD that can function as substrates for PcpS and EntD.*

Chapter 2: Methods

2.1 Media and reagents

2.1.1 Enzymes and chemicals

Restriction endonucleases were obtained from New England Biolabs or Thermo Fisher Scientific. Phusion™ high fidelity polymerase was obtained from Thermo Fisher Scientific. BioMix Red™ polymerase was obtained from Bioline. T4 DNA ligase was obtained from Invitrogen. Mutazyme epPCR kit was obtained from Stratagene. FastAP alkaline phosphatase was obtained from Thermo Fisher Scientific. Antarctic phosphatase was obtained from New England Biolabs. All chemicals were obtained from Sigma-Aldrich or Thermo Fisher Scientific, unless otherwise stated.

2.1.2 Media

All media was made to its final concentration in ddH_2O and sterilised by autoclaving, unless otherwise stated.

2.1.2.1 LB

Low salt LB was obtained from Duchefa Biochemie as a premixed powder. When dissolved, this formula contained 10 g/L tryptone, 5 g/L NaCl, 5 g/L yeast extract.

2.1.2.2 LBA

LB was prepared as stated above and microagar (obtained from Duchefa Biochemie) added to this mix to a final concentration of 1.5% w/v. This mix was poured into petri dishes to set for bacterial growth.

2.1.2.3 Indigoidine production media

LBA media containing appropriate antibiotics was supplemented with 100 mM L-glutamine to allow BpsA activity. Agar concentration in this media was 1.7% w/v.

2.1.2.4 Iron-restricted screening media for directed evolution

Screening media consisted of a mixture with final concentrations of 11 g/L M9 minimal salts (Sigma), 1.7% w/v agar, 100 mM L-Gln, 0.4% v/v glycerol, 2 mM $MgSO_4$, 100 µM 2,2'-dipyridyl, 666 µM L-arabinose, 100 µM $CaCl_2$ and appropriate antibiotics. For work in *P. aeruginosa,* the concentration of 2,2'-dipyridyl was increased to 800 µM**.** 25 mL of molten media was poured into each petri dish.

2.1.2.5 GYT media

GYT media for the preparation of electrocompetent cells was made to contain 10% v/v glycerol, 0.125% w/v yeast extract and 0.25% w/v tryptone in ddH_2O.

2.1.2.6 Media supplements

All antibiotics were prepared to 1000 x stocks of the working concentrations and added at a 1000 x dilution to the media volume. IPTG was obtained from Thermo Fisher. IPTG was added to plates via a "scoop method", in which the agar was lifted out of the petri dish with a sterilised spatula,

IPTG added underneath and the agar replaced. Antibiotics and their concentrations used in this study are presented in Table 2.1.

Antibiotic	Concentration (μg/mL)
Ampicillin	100
Kanamycin	50
Spectinomycin	100
Carbenicillin	100

Table 2.1: Final concentrations of antibiotics used in this study

2.1.2.7 Chrome azurol S media

Chrome azurol S (CAS) media was made for detection of siderophore production using published methods (Louden et al., 2011; Schwyn et al., 1987). 50 mL Chrome azurol S (CAS) was made by dissolving 60 mg CAS powder (Sigma) in 50 mL distilled water. To this, 5 mL of 1mM $FeCl_3$ was added. The entire solution was then poured slowly with stirring into 40 mL distilled water containing 73 mg dissolved HDTMA (Sigma) and autoclaved to sterilise. 100 mL 10x indigoidine production M9 solution (Section 2.1.2.4, minus 2,2'-dipyridyl, arabinose, agar) was added to 750 mL ddH_2O and then 32.24 g piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) was added. To dissolve PIPES, 12 M NaCl was slowly added to bring the pH up to 6.8. 1.5 g agar was then added to this solution and the mixture autoclaved to sterilise. When cooled to approximately 50 °C, 30 mL filter-sterilised casamino acid solution was added to a concentration of 0.3 % (w/v), and 100 mL of CAS solution added slowly with agitation. Plates were poured aseptically immediately after mixing. When cooled, CAS plates were a dark blue colour.

2.2 Bacterial strains, plasmids, oligonucleotide primers and gBlocks

2.2.1 Bacterial strains used in this study

The strains used in this study are presented in Table 2.2.

Strain	Characteristics	Source
DH5α	*supE*44 D*lacU*169 (Ø80 *lacZ* DM5) *hsdR*17	Invitrogen
BL21(DE3)	F^- *ompT* gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ	Novagen
BL21(DE3) Δ*entD*	Δ*entD*	Laboratory stock
EcoBlueE	BL21(DE3) Δ*entD* + pCDFDuet with BpsA containing the EntF T-domain	Laboratory stock
EcoBlueP	BL21(DE3) Δ*entD* + pCDFDuet with BpsA containing the PvdD T-domain	Laboratory stock

Table 2.2: E. coli strains used in this study

2.2.2 Plasmids used in this study

Plasmids used in this study were obtained from laboratory stock, a manufacturer or created through the course of this study.

Plasmid	Characteristics	Reference
pCDFDuet1	$LacI^q$,T7prom, $Spec^R$, CDFori.	Novagen
pTSWAP	pCDFDuet1 containing modified *bpsA* allowing replacement of native CP.	Laboratory stock
BpsA(E)	pTSWAP with *bpsA* containing the CP domain of *entF*	Laboratory stock
BpsA(P)	pTSWAP with *bpsA* containing the CP domain of *pvdD*	Laboratory stock
BpsA(E12)	pTSWAP with *bpsA* containing the tag region of the *entF* CP	This study
BpsA(P12)	pTSWAP with *bpsA* containing the tag region of the *pvdD* CP	This study
pET28a(+)	$LacI^q$,T7prom, Kan^R, ColE1ori	Novagen
pET::*entD*	pET28a(+) containing wild-type *entD*	Laboratory stock
pET::*pcpS*	pET28a(+) containing wild-type *pcpS*	Laboratory stock
pET::*malE*	pET28a(+) containing truncated *malE* lacking membrane localisation signal sequence	Laboratory stock
pBAD/His A	araBADprom, Amp^R, His tag	Life Technologies

Table 2.3: Plasmids used in this study

2.2.3 Primers used in this study:

Primers were designed using Geneious software and ordered from IDT or obtained from laboratory stocks. Primers were resuspended in 1 x TE buffer (10 mM Tris-HCl, 1 mM EDTA) to a final concentration of 100 µM. Working stocks of 10 µM were prepared from this in ddH_2O. All primers were stored at -20 °C.

Primer	Sequence (5' – 3'). Restriction sites in bold.
entD_pBAD_Fwd_XhoI	AGCT**CTCGAG**ATGGTCGATATG
entD_pBAD_Rev_HindIII	AGCT**AAGCTT**TTAATCGTGTTGG
bpsA_t_mutant_screen	ACGAGCAGATCGGCCACGA
slBPSAT_Fwd	AGCT**CTGCAG**AGCTCGTCGAGCGCCCCTTCGTCGCCCCGCGCACG
slBPSAT_Rev	AGCT**TCTAGA**CTCCTGGGCGACCTCGCGCTCCAGGCGGCGGGCCAG
pBAD_F	ATGCCATAGCATTTTTATCCA
pBAD_R	GATTTAATCTGTATCAGG
MBP_Fwd_NdeI	ATCG**CATATG**AAAATCGAAGAAGGTAAACTGGTAATCTGGATTAA
P12_MBP_Rev_HindIII	ATCG**AAGCTT**ACCGCTCCTTGAGCATCAGCAGCAACAATGAGTGACCGCCGCCCTTGGTGATACGAGT
P11-MBP_Rev_HindIII	ATCG**AAGCTT**ACTCCTTGAGCATCAGCAGCAACAATGAGTGACCGCCGCCCTTGGTGATACGAGT
P10-MBP_Rev_HindIII	ATCG**AAGCTT**ACTTGAGCATCAGCAGCAACAATGAGTGACCGCCGCCCTTGGTGATACGA
P9-MBP_Rev_HindIII	ATCG**AAGCTT**AGAGCATCAGCAGCAACAATGAGTGACCGCCGCCCTTGGTGATACGAGTC
P8A-MBP_Rev_HindIII	ATCG**AAGCTT**ACATCAGCAGCAACAATGAGTGACCGCCGCCCTTGGTGATACGAGTCTGA
P8B-MBP_Rev_HindIII	ATCG**AAGCTT**AGAGCATCAGCAGCAACAATGAGTGGCCGCCCTTGGTGATACGAGTCTGA
E12_MBP_Rev_HindIII	ATCG**AAGCTT**ACTGCGCTGCTAGTTTCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGAGT
E11_MBP_Rev_HindIII	ATCG**AAGCTT**ACGCTGCTAGTTTCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGAGT
E10_MBP_Rev_HindIII	ATCG**AAGCTT**ATGCTAGTTTCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGA
E9_MBP_Rev_HindIII	ATCG**AAGCTT**ATAGTTTCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGAGTC
E8A_MBP_Rev_HindIII	ATCG**AAGCTT**ATTTCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGAGTCTGA
E8B_MBP_Rev_HindIII	ATCG**AAGCTT**ATAGTTTCATTGCCAGTAGCGAATGGCCGCCCTTGGTGATACGAGTCTGA
E12mut1_MBP_Rev_HindIII	ATCG**AAGCTT**ACTGCGCTGCTAGTATCATTGCCAGTAGCGAATGACCGCCGCCCTTGGTGATACGAGT

Table 2.4: Primers used in this study

2.2.4 gBlocks

gBlocks were ordered from IDT and reconstituted in 20 µL 1 x TE buffer to a stock concentration of 10 ng/µL. When required, entFbpsA and S6bpsA were amplified with Phusion™ polymerase as

per the manufacturer's instructions. Primers for these amplification reactions were slBPSAT_Fwd and slBPSA_Rev.

gBlock	Sequence (5' – 3'). Restriction sites in bold, tag sequences underlined.
E12BpsAT	TTTT**CTGCAG**AGCTCGTCGAGCGCCCCTTCGTCGCCCCGCGCACGGAGACGGAGAAGGA GATCGCGGCGGTCTGGGAGAAGGCCCTGCGCCGCGAGAACGCCTCCGTCCAGGACGAC TTCTTCGAGTCGGGCGGTCATTCGCTACTGGCAATGAAACTAGCAGCGCAGCTCAACGC GCGCCTGGGCGTCTCCCTGCCGCTGCAATCCGTCCTGGAGTCCCCGACCATCGAGAAGCT GGCCCGCCGCCTGGAGCGCGAGGTCGCCCAGGAG**TCTAGA**TTT
P12BpsAT	TTTTTTTTCTGCAGAGCTCGTCGAGCGCCCCTTCGTCGCCCCGCGCACGGAGACGGAGAA GGAGATCGCGGCGGTCTGGGAGAAGGCCCTGCGCCGCGAGAACGCCTCCGTCCAGGAC GACTTCTTCGAGTCGGGCGGTCACTCATTGTTGCTGCTGATGCTCAAGGAGCGGCTCAAC GCGCGCCTGGGCGTCTCCCTGCCGCTGCAATCCGTCCTGGAGTCCCCGACCATCGAGAA GCTGGCCCGCCGCCTGGAGCGCGAGGTCGCCCAGGAGTCTAGATTTAGCTCGTCGAGCG CCCCTTCGTCGCCCCGCGCACGGAGACGGAGAAGGAGATCGCGGCGGTCTGGGAGAAG GCCCTGCGCCGCGAGAACGCCTCCGTCCAGGACGACTTCTTCGAGTCGGGCGGTCACTC ATTGTTGCTGCTGATGCTCAAGGAGCGGCTCAACGCGCGCCTGGGCGTCTCCCTGCCGCT GCAATCCGTCCTGGAGTCCCCGACCATCGAGAAGCTGGCCCGCCGCCTGGAGCGCGAGG TCGCCCAGGAG**TCTAGA**TTT
E941KBpsAT	TTTT**CTGCAG**AGCTCGTCGAGCGCCCCTTCGTCGCCCCGCGCACGAAGACGGAGAAGGA GATCGCGGCGGTCTGGGAGAAGGCCCTGCGCCGCGAGAACGCCTCCGTCCAGGACGAC TTCTTCGAGTCGGGCGGCAACTCGCTGATCGCCGTCGGCCTCGTCCGCGAGCTCAACGC GCGCCTGGGCGTCTCCCTGCCGCTGCAATCCGTCCTGGAGTCCCCGACCATCGAGAAGCT GGCCCGCCGCCTGGAGCGCGAGGTCGCCCAGGAG**TCTAGA**TTT
Q962HBpsAT	TTTT**CTGCAGAG**CTCGTCGAGCGCCCCTTCGTCGCCCCGCGCACGGAGACGGAGAAGGA GATCGCGGCGGTCTGGGAGAAGGCCCTGCGCCGCGAGAACGCCTCCGTCCATGACGAC TTCTTCGAGTCGGGCGGCAACTCGCTGATCGCCGTCGGCCTCGTCCGCGAGCTCAACGC GCGCCTGGGCGTCTCCCTGCCGCTGCAATCCGTCCTGGAGTCCCCGACCATCGAGAAGCT GGCCCGCCGCCTGGAGCGCGAGGTCGCCCAGGAG**TCTAGA**TTT

Table 2.5: gBlocks used in this study

2.3 Culturing and general molecular biology

2.3.1 Growth and maintenance of bacteria

When required, samples from -80 °C freezer stocks were picked and streaked onto LB agar media and grown at 37 °C, then stored at 4 °C. These colonies were used to inoculate further cultures. Liquid cultures for downstream applications were made in sealed 50 mL falcon tubes (up to 10 mL cultures), sealed 250 mL conical flasks (10 – 50 mL cultures) or sealed 2500 mL conical flasks (up

to 500 mL cultures). These cultures contained LB and appropriate antibiotics when required. Prior to inoculation, 1 mL of media containing all components was removed and placed in a microcentrifuge tube to serve as a negative control to ensure no contamination was present. Incubation conditions for all liquid cultures were 37 °C, 200 rpm shaking for 12-16 h or until a desired OD_{600} was reached, unless otherwise stated. When a desired strain harbouring one or more plasmids was obtained, a 0.5 mL aliquot of an overnight culture was frozen in 0.5 mL 80% v/v glycerol (40 % v/v glycerol final) at -80 °C for long-term storage.

2.3.2 PCR protocols

Phusion™ high-fidelity polymerase was used according to the manufacturer's directions for accurate amplification of DNA. In some cases, adjustment of the amount of primers, template, and/or the addition of DMSO was required.

Step	Temp	Duration	Cycles
Initial denaturation	98 °C	60 s	
Denaturation	98 °C	10 s	30 cycles
Annealing	As per primer T_m	20 s	
Extension	72 °C	30 s/kb	
Final extension	72 °C	10 m	
Incubation	16 °C	Hold	

Table 2.6: General Phusion™ PCR protocol

In cases where two primers with significant difference in T_m were used and a one-step protocol did not yield a satisfactory result, a two-step protocol was used.

Step	Temp	Duration	Cycles
Initial denaturation	98 °C	60 s	
Denaturation	98 °C	10 s	10 cycles
Annealing	As per higher primer T_m	20 s	
Extension	72 °C	30 s/kb	
Denaturation	98 °C	10 s	20 cycles
Annealing	As per lower primer T_m	20 s	
Extension	72 °C	30 s	
Final extension	72 °C	10 m	
Incubation	16 °C	Hold	

Table 2.7: General two-cycle Phusion™ protocol

BioMix Red™ PCR kit was used for colony PCR.

Step	Temp	Duration	Cycles
Initial denaturation	98 °C	3 m	
Denaturation	98 °C	10 s	30 cycles
Annealing	As per primer T_m	30 s	
Extension	72 °C	30 s/kb	
Final extension	72 °C	10 m	
Incubation	16 °C	Hold	

Table 2.8: BioMix Red™ protocol

PCR protocols were adjusted subjectively when required.

2.3.3 Plasmid preparation for cloning

Overnight cultures of the vector bearing strain were grown and harvested by centrifugation. Plasmid DNA was isolated from the harvested cells using a GeneJet™ plasmid mini-prep kit obtained from Thermo Fisher, following the manufacturer's instructions. The only amendment to these directions was that elution was achieved using ddH_2O heated to 70 °C rather than elution buffer. Plasmid DNA was stored at -20 °C prior to use.

Plasmid DNA was digested using restriction enzymes from either New England Biolabs or Thermo Fisher as per the respective manufacturer's instructions. Prior to digestion, plasmid DNA was heated to 70 °C for 20 min to relieve supercoiling. Following incubation at appropriate conditions,

enzymes were heat inactivated at 80 °C for 20 min. Digested DNA was purified using a Clean and Concentrator kit (Zymo Research) and stored at -20 °C if not used immediately.

2.3.4 Insert preparation for cloning

Purified PCR product was digested with the necessary enzymes in buffer composition and conditions as recommended by the manufacturer. After the required incubation time (2-14 h, dependant on enzyme and quantity of insert) enzymes were heat inactivated at 80 °C for 20 min. DNA from reactions was purified using a Clean and Concentrator kit (Zymo Research). Digested DNA was stored at -20 °C if not used immediately.

2.3.5 Ligation

Ligations were performed using T4 DNA ligase obtained from Invitrogen. Typically it was found that the best results were obtained by using a 1:6 molar ratio of vector:insert. Buffer composition was set as per the manufacturer's directions. The concentration of DNA in the reaction was kept below 10 ng/µL. A vector-only control (i.e. containing no insert DNA) was always set up and ligated in parallel to control for uncut/partially uncut vector able to ligate to itself. Ligations were incubated for 12-16 h at 16 °C.

2.3.6 DNA quantification

DNA concentrations of 1 µL samples were quantified using a Nanodrop ND-1000 spectrophotometer. Samples assessed by this method were also assessed by agarose gel electrophoresis, upon which the brightness of the corresponding band could be used as a rough confirmation of the Nanodrop reading.

2.3.6 Agarose gel electrophoresis

For rapid assessment of the quality of DNA products of the reactions described in this section, 2 µL samples were diluted 1:1 with ddH_2O and mixed with 1 µL 5x DNA loading buffer obtained from Bioline. These samples were loaded into wells of a 1 % w/v agarose gel supplemented with 1 µg/mL ethidium bromide. The DNA ladder used as a size standard was HyperLadder™ I obtained from Bioline. Gels were run at 125 V, 400 mA in 1 x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) for 30 – 60 min depending on the size of the DNA fragment being assessed. UviTec software connected to a transilluminator was used for gel visualisation.

2.3.7 Gene sequence data

The NCBI Gene ID numbers for sequences used in alignments are presented in Table 2.9.

Name	**ID #**
entD	945194
pcpS	878433
entF	945184
pvdD	879750
bpsA	20473004

Table 2.9: Genbank sequences used in this study

2.3.8 DNA sequencing

Sequencing of nucleic acids was performed by Macrogen Inc. Samples were transported in conditions according to the company's directions. Sequence data was analysed using Geneious software and aligned to Genbank sequence data.

2.4. Competent cells and transformation

2.4.1 Preparation of electrocompetent *E. coli*

The strain to be made competent was grown overnight on LBA at 37 °C. From this a single colony was picked to inoculate 50 mL LB broth which was grown overnight at 37 °C, 200 rpm shaking. 10 mL of this culture was subsequently used to inoculate a 400 mL LB culture in a sealed 2 L conical flask which was grown at 37 °C, 200 rpm to an OD_{600} of 0.4 – 0.6 and subsequently placed on ice for 20 – 30 min. Cells were transferred to chilled 50 mL Falcon tubes and harvested by centrifugation at 4000 rpm, 4 °C for 20 minutes. Cells were washed by sequential resuspension and centrifugation using 100 % culture volume of chilled ddH_2O, 50 % culture volume of 10 % v/v glycerol, 5 % culture volume of 10 % v/v glycerol and finally suspended in a volume of ice-cold GYT media to give a final concentration of 2 – 3 x 10^{10} cells/mL, as calculated from OD_{600} values. 60 µL aliquots were snap frozen and stored at -80 °C prior to use.

2.4.2 Transformation of electrocompetent *E. coli*

60 µL aliquots of competent *E. coli* cells were defrosted on ice and mixed with 20 µL ddH_2O containing 0.1 – 50 ng DNA. The mixture was then transferred to a pre-chilled electroporation cuvette which was electroporated at 2.5 kV, 200 Ω for 4.7 ms. 1 mL of LB was then immediately added to the cuvette and it was incubated at 37 °C, 200 rpm for 1 h. Recovered cells were then plated on appropriate solid media, or mixed 1:1 with sterile 80 % v/v glycerol and stored at -80 °C if not used immediately.

2.4.3 Preparation of chemically competent *E. coli*

For transformations not requiring a high efficiency, chemically competent cells were used. These were prepared by inoculating a 3 mL overnight culture of LB supplemented with 10 mM $MgCl_2$ and appropriate antibiotics with the desired strain from either glycerol stock or agar plate. Following approximately 16 h incubation at 37 °C, 200 rpm shaking this culture was used to inoculate a 50 mL LB culture also supplemented with 10 mM $MgCl_2$ and appropriate antibiotics to a starting OD_{600} of 0.05. This culture was then returned to the same incubation conditions and grown to an OD_{600} of 0.4-0.6 when it was harvested by centrifugation at 2700 x g, 4 °C for 15 min. The resulting cell pellet was resuspended in 50 mL ice-cold TFBI buffer (Table 2.10) and incubated on ice for 2 h, following which the cells were pelleted again by centrifugation at 2700 x g for 15 min at 4 °C. The cell pellet was then resuspended in 5 mL ice-cold TFBII buffer (Table 2.10). Aliquots of 50 – 200 µL were snap frozen in pre-chilled 1.5 mL microcentrifuge tubes and stored at -80 °C until use.

Component	Final concentration
TFBI	
Potassium acetate	30 mM
$MnCl_2$	50 mM
$CaCl_2$	10 mM
Glycerol	15% (v/v)
TFBII	
Na-MOPS, pH 7.0	10 mM
$CaCl_2$	75 mM
KCl	10 mM
Glycerol	15% (v/v)

Table 2.10: Buffers used in preparation of chemically competent E. coli

2.4.4 Transformation of chemically competent *E. coli*

For transformation of chemically competent *E. coli* cells, aliquoted cells were thawed on ice and mixed with a volume of DNA no greater than 10% of the cell volume (i.e. ≤10 µL DNA in 100 µL

cells) in 1.5 mL microcentrifuge tubes and incubated on ice for 20 min. Following this incubation, cells were heat shocked at 42 °C for 90 s using a heat block, then returned to ice for 2 min. Next, LB was added to cells to make the volume up to 1 mL and the cells recovered at 37 °C, 200 rpm shaking for 1 h prior to plating on appropriate media.

2.4.5 Preparation of Mix & Go™ competent cells

For large scale transformation in which many different plasmids needed to be transformed, Mix & Go™ (Zymo Research) competent cells were employed due to their simplicity of preparation. Preparation was performed according to the manufacturer's specifications.

2.4.6 Transformation of Mix & Go™ competent cells

Twenty-five microliter aliquots of competent cells were transformed in a 96-well plate via mixing with ≤ 5% (v/v) plasmid DNA. The mixtures were then diluted in LB to a final volume of 100 µL and spread onto solid media containing appropriate antibiotics and incubated at 37 °C for 12-16 h.

2.5 Protein expression and purification

2.5.1 Protein expression

A culture of the desired strain was grown for 12 – 14 h in LB supplemented with appropriate antibiotics and used to seed a larger culture, the size of which was dependant on the downstream application. This culture was grown to OD_{600} 0.6 - 0.8 at which point the culture was chilled on ice and expression was induced with IPTG to a final concentration of 0.5 mM. The culture was then

incubated for a further 24 h at 18 °C, 200 rpm. Cells were then harvested by centrifugation and cells pellets frozen at -80 °C until use.

2.5.2 Protein purification

2.5.2.1 Cell lysis and solubility based separation

Cell pellets prepared as described in Section 2.5.1 were resuspended in 5 % culture volume of chilled 1 x protein binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) and mechanically lysed using a French press or sonication. Lysis by French press was achieved using two passages at 1000 psi. The French Press cell was cooled on ice for 10 minutes prior to use. Sonication was used when access to the French press was unavailable. Lysis by sonication was achieved at 70 % power with a 50 % duty cycle. 10 s pulses were used to a point of sufficient lysis as estimated by consistency of the bacterial suspension (usually 2-3 minutes). Fractions resulting from lysis were separated by centrifugation (20 min, 2 °C, 4000 x g). The soluble fraction was removed and kept on ice prior to purification.

2.5.2.2 Ni-NTA purification of His6-tagged proteins

His_6-tagged protein purification was performed using a Novagen Hisbind™ Ni-NTA chromatography kit as per the manufacturer's directions. Purification was typically achieved using 1.5 mL total settled resin. 1.5 mL eluent fractions were collected off the column for SDS-PAGE analysis.

2.5.2.2.1 BpsA purification

This method was based on the BpsA purification method described by Owen et al (Owen et al., 2011). To achieve good BpsA purification it was necessary to use a total of 8 mL settled resin and to add an excess of soluble cell lysate (the entire soluble fraction obtained from a 400 mL expression culture) to saturate the column's binding capacity. Flow-through was added to the column a second time. The initial wash step was performed three times and the second wash step recommended by the manufacturer was not used. Following elution, the buffer was exchanged for 50 mM sodium phosphate buffer and 12.5 % v/v glycerol using a 100 kDa cut-off Millipore Amicon® ultra-15 centrifugal filter tube.

2.5.2.2.2 PPTase purification

This method was based on the PPTase purification method described (Owen et al., 2011). To prevent precipitation of PPTases directly following elution, all Hisbind™ buffers were supplemented with 25 % v/v glycerol and kept ice-cold. Once eluted, fractions were desalted and stored in storage buffer (50 mM Tris-HCl, pH 7.5) supplemented with 50 % v/v glycerol at -80 °C.

2.5.3 SDS-PAGE analysis of protein expression

SDS-PAGE gels were set up and run as described in (Laemmli, 1970). The stacking gel was made to 5 % acrylamide and resolving gels made to 12.5 %, 15 % or 17.5 % depending on the size of the target protein. Gels were set and run with a Bio-Rad Protean II dock as per the manufacturer's instructions. Staining of gels was performed as described by Green et al. (Green et al., 2012). Pierce™ Unstained Protein Molecular Weight Marker (ThermoScientific) was used for size comparison (Figure 2.1).

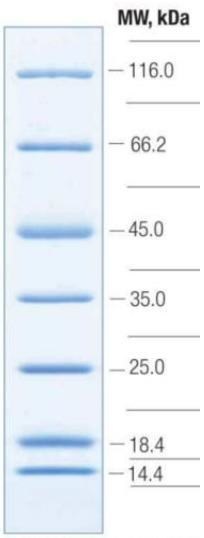

Figure 2.1: Unstained Protein Molecular Weight Marker for SDS-PAGE.

2.5.3 Peptide synthesis

Peptide tags were artificially synthesised rather than purified. Peptides were either synthesised at the Ferrier Research Institute or purchased from GenScript. Upon arrival, peptides were solubilised in 100 % DMSO, then diluted and aliquoted at 100 µM and stored at -20 °C.

2.6 Library generation and screening for directed evolution

2.6.1 Library generation

2.6.1.1 Vector digestion

16 µg of plasmid DNA was digested using 50 U of enzyme in a 400 µL reaction. For work in *E. coli,* the vector used was pBAD. This was digested using HindIII High-Fidelity and XhoI (New England Biolabs. After 5 h, an additional 20 U of enzyme was added and the reaction left overnight. Following this incubation period, Antarctic phosphatase (Thermo) was added to the reaction as recommended by the manufacturer. This reaction was left to proceed for 4 h, following which the

enzymes were heat inactivated at 80 °C for 20 min and the DNA cleaned using a Clean and Concentrator 25 kit (Zymo Research). Digested vector was eluted from the column in 30 µL ddH_2O and stored at -20 °C. A 200 ng sample of the vector was analysed by agarose gel electrophoresis to check for degradation.

2.6.1.2 Determining digested vector quality

To assess the quality of the vector following digestion, the wild-type *entD* gene was amplified with Phusion™ high-fidelity polymerase and digested using the same enzymes as the vector. This was cloned into the digested vector and transformed (as described in Section 2.4.2) into electrocompetent Δ*entD* cells harbouring a plasmid containing functional *bpsA*. These cells were plated onto indigoidine production media (LB, 1.7% w/v agar, 1 M L-glutamine, appropriate antibiotics). Once colonies had formed, *bpsA* expression was induced with 100 µL of 100 mM IPTG added beneath the agar within the petri dish and plates were incubated at room temperature. Expression of both BpsA and wild-type EntD resulted in indigoidine production; if this occurred in over 70% of colonies the vector was deemed to be of sufficient quality for library generation.

2.6.1.3 Error prone PCR

epPCR reactions were performed using a Stratagene Mutazyme II kit as per the manufacturer's instructions. An optimal error rate was found by empirically assessing the rate of improved clone generation. The template DNA for mutation was generated using Phusion™ high-fidelity polymerase. 100 ng template DNA was used for library generation. Prior to running epPCR, reactions were split into separate reactions to lower the rate of clonal mutants generated. Quality of amplicons generated was assessed by analysis of a 2 µL sample by agarose gel electrophoresis.

2.6.1.4 Preparation of insert

epPCR reactions were cleaned using a Clean and Concentrator kit (Zymo Research) and eluted in 20 µL ddH_2O. This was digested using 30 U of each enzyme for 5 h in a 50 µL reaction. The components of these reactions were as per the instructions of the manufacturer. Following digestion, reactions were heat inactivated at 80 °C for 20 min and cleaned using a Clean and Concentrator kit (Zymo Research).

2.6.1.5 Ligation

Ligation of the digested insert into the digested vector was performed as described in Section 2.3.5, however the total amount of DNA in each ligation was 1.0 µg and following overnight incubation, an additional 2 U ligase was added and the reaction incubated for a further 8 h.

2.6.1.6 Transformation of library into screening strain

Transformations were carried out as described in Section 2.4.2 with strain EcoBlueP. Recovered transformants were then mixed 1:1 with sterile 80 % glycerol and stored in 0.5 mL aliquots at -80 °C. It was found that plating 200 µL of transformants resulted in manageable numbers of colonies upon a single screening plate.

2.6.1.7 Control transformations

Cells electroporated with ddH_2O only served as the first negative control. The second negative control involved transformation with uncut vector. This confirmed that cells without a PPTase could not grow in the Fe-restricted conditions. Cells transformed with the target WT PPTase served as the positive control.

2.6.2 Library screening

2.6.2.1 First-tier screening

The aim of this directed evolution study was to generate PPTase variants unable to activate the foreign CP target, whilst retaining activity with their native CP substrates. The enzymatic effect the PPTases have on the native CPs allows the cell to produce a siderophore, granting ability to grow in Fe limited conditions. To this end, an orthogonal screen was used. This consisted of minimal media containing the iron chelator 2,2'-dipyridyl to select for clones containing a PPTase able to activate their WT substrate (and thus able to produce a siderophore, allowing growth), and the use of a screening strain containing the reporter gene *bpsA* with its CP substituted by the foreign CP target. It was reasoned that a PPTase able to activate the foreign CP would result in indigoidine production and therefore blue colonies, whereas PPTases that had lost this ability as a result of mutation would not activate BpsA and would therefore generate white colonies. 200 µL of transformants were plated onto each screening plate (as described in section 1.2.3) and incubated at 37 °C and colonies allowed to develop. This typically took 48-72 h due to the restrictive nature of the medium. Here it was found that use of ampicillin to select for plasmid pBAD resulted in a number of satellite colonies which interfered with screening, therefore carbenicillin was adopted in place of ampicillin for future screening. Upon presence of medium sized colonies (~1 mm diameter) *bpsA* gene expression was induced by adding 100 µL 100 mM IPTG via the scoop method (describe in section 1.2.6), after which the plates were incubated at room temperature for a further 2 – 3 days. After this period, colonies which remained white were streaked onto both a new screening plate and indigoidine production media (Section 2.1.2.3), alongside a positive control of the screening strain expressing the wild-type PPTase. Re-plating on screening media is important because siderophore compounds secreted from active PPTase containing cells can be taken up by cells with a non-functional PPTase, thus resulting in satellite colonies that may appear as hits. Re-plating on indigoidine production media helps to distinguish

the colour of colonies because variation in inducer (L-arabinose and IPTG) concentrations in the screening media can result in the inaccurate appearance of colonies. Colonies which still grew well and remained white were then selected for the second-tier screening. Overnight cultures were inoculated from these colonies and duplicate glycerol stocks frozen in a 96-well plate. 10 mL overnight cultures (LB, appropriate antibiotics) were also created from these glycerol stocks for plasmid isolation to enable DNA sequencing.

2.6.2.2 Second-tier screening of PPTase variants

To augment screening on solid media, second-tier screening was performed in liquid media which allowed more accurate quantification of PPTase activity. The improved clones from first-tier screening were used to inoculate 200 µL overnight cultures (LB, appropriate antibiotics) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate wells in a new 96-well plate. These plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader at 24 and 48 h time points. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. This was chosen so that indigoidine production would not interfere with the OD_{800} reading. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values.

2.6.2.3 Kinetic determination of PPTases

To determine the kinetic parameters of PPTases, top performing mutants were expressed and purified (as described in section 2.5.2). Activity assays with these were conducted in 96-well

plates. Activity assays consisted of triplicate reactions spanning serial dilutions of CoA from 100 – 0.09765 µM. In addition to CoA, the 200 µL reactions contained 4 mM L-Gln, 1 mM ATP, 1 µM *apo*-BpsA (containing the appropriate CP), 1 µM variant PPTase. To begin the reaction, PPTase was added last and the plate shaken to mix at 900 rpm for 10s. A_{590} readings were taken at 20 s intervals for 1-2h depending on the PPTase-BpsA combination. The slope function of Microsoft Excel was used to calculate to the velocity of the reaction, and then again applied to these velocity values and the maximum values taken as maximum PPTase velocity. Prism GraphPad was then used to derive kinetic parameters as published (Owen et al., 2011)

2.7 BpsA competition assays

Competition assays were performed to assess how well isolated tag peptides or tagged proteins could act as substrates for PPTases. Assays were performed in triplicate in 96-well plates. First, 50 µL volumes of peptide tag/tagged protein were serially diluted 2x with ddH_2O across the plate, after which 50 µL of reaction mix containing 2 µM BpsA, 0.75 µM CoA, 40 mM $MgCl_2$, 400 mM TrisCl pH 8.0 was added to each well. To begin BpsA activation, 50 µL PPTase was then added to each well at a final concentration of 0.25 µM (PcpS) or 0.5 µM (EntD). This strategy ensured that the tag construct and BpsA would compete for the PPTase activity, given a limited amount of CoA. Reactions were then incubated at 30 °C, 200 rpm for 40 min (PcpS reactions) or 75 min (EntD reactions), after which 50 µL of an ATP (4 mM) and L-Gln (4 mM) mix was added and the plate mixed at 1000 rpm for 10 s to initiate indigoidine production. Two control reactions were used, one containing no PPTase (-ve) and one containing no competing peptides (+ve).

Indigoidine production was quantified by measuring absorbance at 590 nm at 10s intervals for 30 – 60 mins with a microplate reader. The slope function of Microsoft Excel was used to calculate the velocity of the reactions and Prism GraphPad used to derive kinetic data. In order to calculate

IC_{50} values, an additional data point was added at 0.1 M tag to represent complete inhibition to complete the dose-response curve.

Chapter 3: Directed evolution of *entD* towards substrate-selectivity

3.1 Introduction

3.1.1 *entD*

As described in Section 1.1.2, *entD* is an *E. coli* gene that encodes the 209 amino acid PPTase EntD required for production of the siderophore enterobactin. The *entD* gene is located closely upstream of the enterobactin gene cluster, which encodes enzymes involved in the synthesis (*ent* genes), transport of ferric enterobactin (*fep* genes) and release of iron from the complex (*fes* gene) (Walsh et al., 1990) (Figure 3.1). Enterobactin is synthesised in a stepwise manner involving enzymes EntB, EntC, EntE and EntF (Raymond et al., 2003). The role of EntD in this pathway is to pantetheinylate (i.e. activate) the CP domains of EntB and the NRPS EntF, allowing them to bind and modify the growing enterobactin molecule (Gehring, et al., 1997). Without this post-translational modification EntF remains in its inactive state and enterobactin is not synthesised. Although *E. coli* expresses two other PPTases, AcpS and AcpT, neither of these is able to activate the CP domains of the enterobactin biosynthetic machinery. As enterobactin is the sole siderophore of *E. coli,* loss of EntD function prevents growth in an iron-restricted environment.

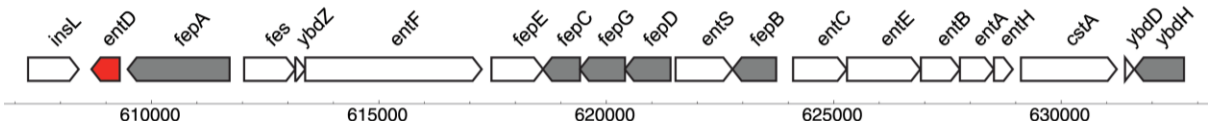

Figure 3.1: Genetic structure of enterobactin gene cluster. (Beld et al., 2014)

EntD belongs to the Sfp-type (type II) family of PPTases and although it has not been extensively studied, is thought to exhibit relatively broad substrate specificity as is characteristic of this family. For example, EntD is interchangeable with Sfp in the surfactin synthesis pathway of *B. subtilis*. Furthermore, there is evidence that EntD is capable of modifying at least some PKS carrier proteins

(Chalut et al., 2006). Of particular relevance to the work performed in this thesis, EntD is capable of activating the foreign CP domain of PvdD from *P. aeruginosa* (Figure 3.2), part of the reason for its selection as a partner-PPTase for PcpS. Compared to other type II PPTase family members, however, EntD is somewhat more specific. For example, heterologous expression of active NRPS/PKS enzymes in *E. coli* usually requires co-expression of a PPTase with broader activity, often Sfp (Kealey et al., 1998; Ku et al., 1997). Moreover, EntD is inefficient at activating the native BpsA CP domain (Owen et al., 2011). As with other members of the type II PPTase family, EntD is capable of using CoA analogues as substrate for post-translational modification of its target CP, and this is thought to happen in nature when CoA supply is limited (Chen et al., 2009).

Figure 3.2: Indigoidine production resulting from EntD activity with the PvdD CP within BpsA.

E. coli ΔentD expressing *entD* from pET28a(+) and *bpsA(P)* from pCDFDuet. Both plasmids induced via addition of 150 µL 10 % (v/v) IPTG underneath agar and incubated at room temperature for 24 h.

3.1.2 The promise of PPTases as site-specific labellers.

As described in Section 1.4, the substrate promiscuity of Sfp-type PPTases for carrier protein domains of different biosynthetic enzymes limits their practicality for multiplex labelling uses in

biotechnology. For example, PPTases have been used to identify the cellular locations of proteins on the surfaces of living cells, and to study the interactions of those proteins with one other protein of interest in any given experiment (Section 1.4). Previous efforts to achieve multiplex labelling have typically employed a relatively specific PPTase to label one substrate and subsequently a non-selective PPTase to label the unlabelled substrates that remain. There is a demand for a greater range of substrate-selective PPTases to broaden the scope of such labelling studies and thereby allow study of multiple cellular components in a multiplexed approach.

To this end, we predicted that the PPTases EntD and PcpS would be a good pair of enzymes with which to begin the evolution strategy because they are both activators of NRPS CP domains, have relatively high levels of activity, and PcpS naturally shows a preference for its native substrate, PvdD, over the substrate of EntD, EntF (Figure 4.3). Critically, EntD and PcpS are both required for the biosynthesis of siderophores in their native strains, which means that Dr. Robins' screen could be applied to directed evolution of both enzymes. PcpS is a particularly interesting candidate PPTase for evolution because it is so distinct from other PPTases in sequence and no structure is available for it. Comparison of *pcpS* evolution to *entD* evolution, therefore, may shed light on the mechanisms by which this enzyme acts in relation to better characterised PPTases.

3.1.3 Previously developed screen for substrate-selective PPTases.

During her PhD studies, Katherine Robins of the Ackerley lab developed a screen to evolve the *entD* gene to lose activity with wild-type BpsA while retaining its native activity with EntB and EntF of the enterobactin synthesis pathway (Robins, 2016). The *E. coli* strain used for this screen was named EcoBlue, which contains BpsA encoded on the IPTG-inducible plasmid pCDFDuet and has had the chromosomal *entD* gene knocked out. The screen relies on an iron-restricted, nutritionally minimal solid media that permits only the growth of cells expressing a PPTase capable of activating EntF and EntB, i.e cells that can synthesise enterobactin, scavenge iron and survive. To then assess

the BpsA reactivity of the *entD* variants, BpsA expression was induced via IPTG addition and subsequent incubation at room temperature. Following this process, it was anticipated that white colonies, or colonies with a reduced blueness (compared to a WT *entD* control) would have been evolved towards the desired phenotype and could be picked for further characterisation and evolution.

In a small proof-of-principle study, Dr. Robins found that using the Fe^{2+} chelator, 2'2-dipyridyl, at a concentration of 100 µM allowed positive selection of EntF-active *entD* variants after 72 hours of incubation at 37 ˚C.

The induction method for *entD* expression was by arabinose inclusion in the solid media, dictated by the plasmid pBAD into which the *entD* library was cloned. The arabinose concentration used was 0.02% (w/v) which allowed rapid BpsA activation (if possible) as reported by indigoidine production. To ensure sufficient raw material to support high-level indigoidine synthesis, 100 mM L-Glutamine, the typical concentration for indigoidine production, was used in this media.

Using her screen, Dr Robins achieved loss of EntD – BpsA(WT) activity with retention of EntD – BpsA(E) activity, although the catalytic efficiency of the latter reaction was approximately 200-fold lower than when performed by WT EntD. When speaking to this result, Katherine stated how an overall reduction in PPTase function could be interpreted as substrate-selectivity since there is already a large difference in EntD activity with BpsA(WT) versus with BpsA(E) ((Robins, 2016), Figure 4.3), and therefore this result may not be have been as successful as it first appeared. The work of this project will hopefully add to these findings and clarify the true activity profiles of any desirable variants that are generated and thereby also clarify the findings of Dr. Robins' work.

3.1.4 Aims

- To employ the screening strategy of Dr. Robins, to
- Generate an *entD* variant which cannot modify the foreign CP of PvdD whilst retaining its ability to modify the CP of EntF.
- To characterise the lead variant(s) by kinetic determination.

3.2 Results

3.2.1 Generation of *entD* epPCR libraries

Using a GeneMorph Mutazyme II kit, an epPCR library of *entD* variants was created from a WT *entD* template. The epPCR reaction was designed (as per the manufacturer's instructions) to introduce an average of 3 – 4 point mutations in the 621 bp gene. This library was cloned into the arabinose inducible plasmid pBAD using restriction sites XhoI and HindIII. The vector used for library generation was tested for quality prior to ligation as described in Section 2.6.1.2. The library was subsequently transformed into the screening strain EcoBlueP. The mutation rate of the library was checked by sequencing ten randomly chosen clones and was found to be 1.8 mutations/gene. This mutation rate is lower than the rate predicted by the Mutazyme II protocol, an observation commonly made in our laboratory. One clone sequenced contained zero mutations.

3.2.2 Library screening

3.2.2.1 First-tier screening

Between 1.5 x 10^4 – 1.6 x 10^4 clones were initially subjected to first-tier screening (Section 2.6.2.1). Following 72h incubation at 37 ˚C the screening plates had grown colonies varying in diameter

from 0.1 – 0.5 mm. In addition to these colonies, thousands of satellite colonies were present encircling the larger colonies. This observation is typical of siderophore-dependent growth and was not confounding as it was not difficult to distinguish between the two types of colony. At this point the screening plates were induced with IPTG and incubated at room temperature for a further 48 – 72 h. Approximately 30% of non-satellite colonies had turned blue following this induction period, indicating that the remaining 70% were preliminary hits – an unexpectedly high hit rate. Due to the abundance of white colonies, only the largest, healthiest white colonies were selected for further investigation. These clones were grown overnight and their plasmid DNA purified for preparation for second-tier screening. Separation of the PPTase encoding plasmid and the first-tier BpsA(P) screening plasmid had to be performed to ensure that upon transformation into the second-tier screening strains indigoidine production (or lack of) would report on EntD activity with only the BpsA variant inherent to those strains. This isolation was performed by transformation of BL21 Δ*entD* using ampicillin selection, followed by single colony restreaking on ampicillin and spectinomycin; clones unable to grow on the dual antibiotic media had not taken up a BpsA encoding plasmid. These clones were day-cultured and their plasmid DNA harvested for transformation into the second-tier screening strains EcoBlueE and EcoBlueP.

3.2.2.2 Second-tier screening

To confirm the characteristics of the promising clones from first-tier screening, 200 µL cultures of each second-tier screening strain expressing an *entD* variant were incubated overnight and subsequently used to inoculate second-tier screening reactions (Section 2.6.2.2).

Following incubation of these *in vivo* screening reactions, it was revealed that all clones investigated were false positives and fell into three categories:

a) Variants exhibiting drastically reduced activity with both EntF and PvdD, i.e. their overall activity had been reduced to either a trace or undetectable level by BpsA analysis (see

variants 1A, 1F, 1G, IH, 1I, 1J, 2J, 3B, 3E, 3F, 3G, 3J, 4G, KL, 2F, 7A, 4.1, 11.11, 15.3; Figure 3.3), but was nevertheless sufficient for enterobactin synthesis, providing them an alternate path through the screen to selection. Variants of this kind are a major pitfall of this screening strategy; it is clear that only a trace level of EntD activity is required for the synthesis of sufficient enterobactin to permit growth on the selection media and thus much screening effort and characterisation goes into identifying and removing these false positives. The variants shown in Figure 3.3 belonged to a single screening attempt and are presented as an example group that represent the phenotypes shown by initially selected *entD* variants from all screening attempts that were made. Continued screening of the library consistently revealed this activity profile, with 40 – 60% of selected variants exhibiting the loss of activity described here.

b) Variants having a similar level of activity to WT *entD*, i.e. the second-tier screen of these variants yielded a high level of indigoidine production in both EcoBlueE and EcoBlueP. Examples of such *entD* variants are 2B, 3D, 4B, 4D, 4F, 8.3, 12.7, 12.9 (Figure 3.3). The first-tier screen was designed to prevent selection of variants that had retained PvdD reactivity by labelling these colonies blue. The failure of this step is likely to be due to absence of BpsA(E) induction or expression of non-functional BpsA(E) caused by chance mutation (this is not unprecedented; it has been noted that the mild toxicity of indigoidine to the producing cell can establish a selective pressure for mutations in the *bpsA* gene (Owen et al., 2011)). This activity profile consistently accounted for the remainder of first-tier false positives that did not exhibit the phenotype described in (a).

c) Variants falsely reported by the second-tier screen. This situation occurred infrequently but was nevertheless confounding; these variants appeared to have evolved a certain degree of selectivity, however further investigation revealed that the activity profile was

as described in point b), i.e. comparable to WT EntD. An example of such an *entD* variant is ID which appeared to have evolved specificity for BpsA(P) - curiously, as this is the inverse of the activity the screen sought. This variant was retransformed to the second-tier screening strains and subjected to repetition of the second-tier screen which revealed its true activity – the same as WT.

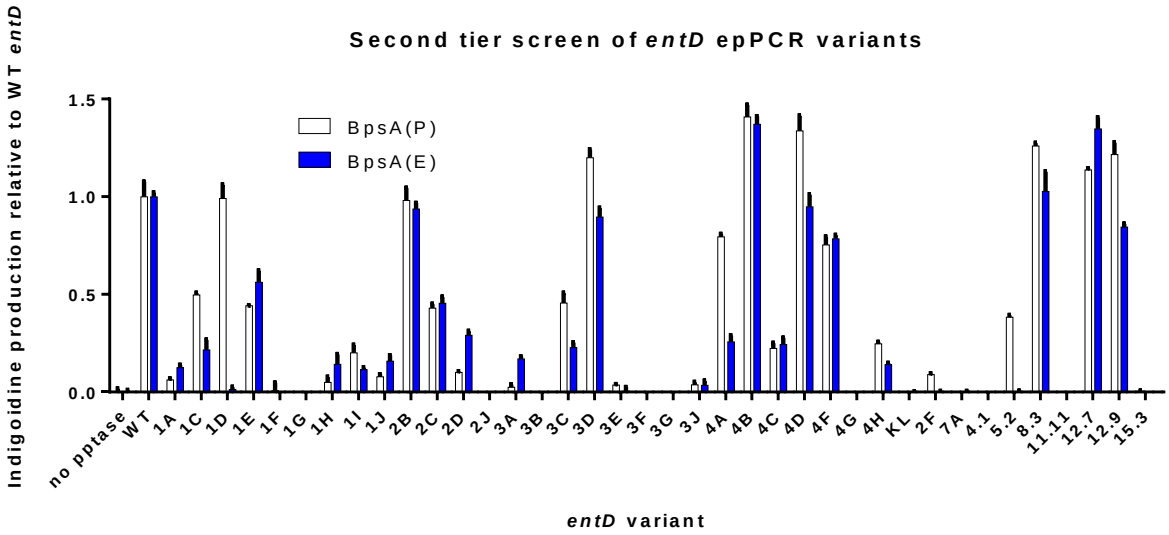

Figure 3.3: Second-tier screen results of a subset of promising entD variants from first-tier screening.

Colonies selected from first-tier screening were used to inoculate 200 µL overnight cultures (LB, appropriate antibiotics) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate wells in a new 96-well plate. Plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader at 24 and 48 h time points. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values.

Another drawback of this screening strategy was that the period of time between the first plating of the library and generating results from subsequent second-tier screening could last up to two weeks due to numerous long incubation periods. Following these initial screening efforts, it was concluded that the screening strategy required modification, not only to avoid selection of false positive variants, but also to streamline the workflow and to reduce the time required to identify and confirm hits from the library.

3.2.3 Screen developments

A number of different approaches were tested in order to prevent selection of *entD* variants possessing non-specific, trace-level CP recognition, as well as to increase the rate of screening.

3.2.3.1 Solid media optimisation for increased selectivity and growth rate

In the initial library screening attempts, L-arabinose had been included to induce the expression of the PPTase library from its host plasmid, pBAD. It has been consistently observed in our laboratory that pBAD exhibits a basal level of 'leaky' expression and thus it was hypothesised that relying solely on this low-level expression would enrich for PPTases with high levels of activity that could compensate for their low molecular concentration. To test this hypothesis, PPTase variants with known activity were grown on first-tier screening media with or without L-arabinose. PPTases chosen for this experiment were *entD* 8.3 (having activity equivalent to WT) and *entD* 4.1 (having trace activity undetectable by BpsA), as well as an empty plasmid (no PPTase) control. Overnight cultures of *E. coli* BL21 Δ*entD* cells transformed with the different plasmids were diluted to yield about 100 colonies/plate and were plated on the test media.

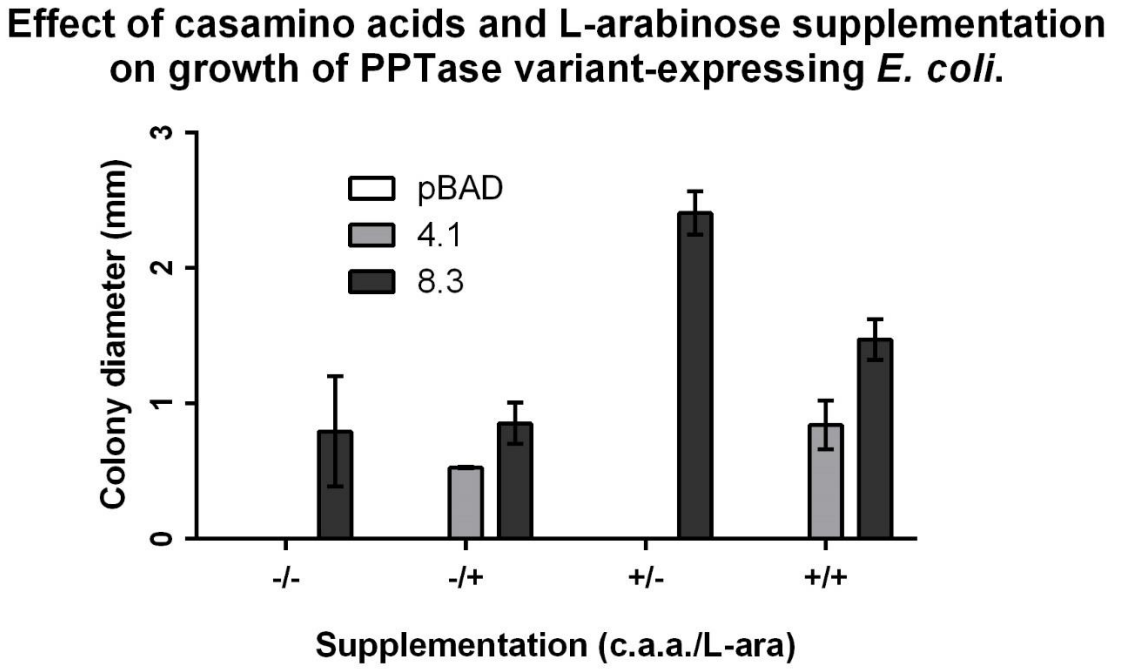

Figure 3.4: Testing the effect of L-arabinose (L-ara) and casamino acid (c.a.a.) supplemented media on growth of E. coli expressing PPTases entD 4.1 and 8.3 or an empty plasmid (pBAD) control.

Base media was iron-restricted screening media (Section 2.1.2.4). + indicates inclusion, - indicates omission. Plates were incubated for 45 hours at 37 °C, after which measurements were made.

Measurements are reported (Figure 3.4) following 45 hours of incubation at 37 ˚C because maximal colony numbers had been achieved by this time. Plates were incubated for 72 hours in total; plates with no growth at 45 hours still had no growth at 72 h. Each condition was tested in triplicate. In no condition were cells lacking a PPTase (i.e. the pBAD control) able to grow. Screening media without arabinose (-/- in Figure 3.4) permitted exclusively the growth of *E. coli* expressing a highly active PPTase; *entD* 4.1 was unable to grow in this condition. A slow growth rate was observed on this screening media containing arabinose but no other supplements (-/+ in Figure 3.4); this was attributed to a lack of general cellular nutrients in addition to the limited iron. Therefore, it was theorised that supplementation with casamino acids might aid cell growth by providing an amino acid source, while not interfering with the screening method. Inclusion of 1 gL^{-1} casamino acids (+/- and +/+ in Figure 3.4) supported this theory; colonies were visible upon all +/- plates after 24 hours, a threefold improvement on the original incubation time allowed for

this step. Presence of both casamino acids and L-arabinose (+/+ in Figure 3.4) permitted growth of *entD* 4.1 expressing cells as well as *entD* 8.3 expressing cells. These colonies required 30 hours of incubation, indicating that the casamino acids again permitted faster growth, but that the L-arabinose had a negative effect and slowed growth somewhat.

Following these results, it was concluded that L-arabinose was detrimental to the screening media in two ways: it promoted the growth of cells expressing a PPTase that had very low activity, and it slowed the rate of bacterial growth. For these reasons, arabinose was omitted from all future solid screening media. Conversely, 1 gL^{-1} casamino acids were included in all subsequent solid screening media due to its ability to improve growth rate without impairing the selectivity of the media.

Media made with these two modifications was then used to screen approximately 1.5 x 10^4 library variants (from Section 3.2.1). Of the 93 variants taken to second-tier screening, roughly 40% had activity undetectable by BpsA (not shown) which demonstrated that the omission of L-arabinose had not solved the issue of these false-positives navigating the screen towards selection. The failure of this modified screen despite its promising control results appears to be a function of plating a mixed population of cells, wherein the low activity of false-positive clones may be augmented by uptake of extracellular enterobactin secreted by neighbouring cells. The definitive reason for these observations remains unclear. The remaining 60% of variants screened had activity equivalent to WT *entD,* indicating that the BpsA(P) counter-screening step was still imperfect.

3.2.3.2 CAS based screening

Screening experience thus far led to the conclusion that relying on growth as a measure of PPTase activity with the EntF CP was an inadequate method. Instead it was theorised that detecting the siderophore molecules directly might give a more accurate representation of the abundance of siderophore, and thereby levels of PPTase-mediated CP activation also. The effect of incorporating

a chrome azurol S (CAS) siderophore assay into the library's first-tier screening protocol was tested. CAS media basically consists of iron bound to the dye, CAS. In this bound state the dye is a blue colour. If a colony on this media produces siderophore, it will absorb the iron from the dye, causing the colour to change to orange. CAS media was made as described in Section 2.1.2.7. The protocol was followed as described by Louden et al, with the addition of screening components $MgSO_4$ and $CaCl_2$ (Section 2.1.2.4). Following the findings described in Section 3.2.3.1, L-arabinose was omitted from this media and casaminoacids were included. The strategy for this screen was to plate the library as usual, induce BpsA(P) expression and to select colonies that possess an orange halo but do not produce indigoidine. A concern with using media was that the trace PPTase activity that could not be detected by BpsA would continue to confound screening by providing sufficient enterobactin production to create a halo. To test the viability of this media, three control *entD* variants obtained from initial library screening were tested for their ability to form colony halos. These variants were; *entD* 4.1 (representing PPTase variants with trace level activity), 1I (representing a variant with reduced activity but still detectable by BpsA screening), and 8.3 (representing a variant with equivalent activity to WT). A no-PPTase control (empty pBAD vector) was also used as a negative control.

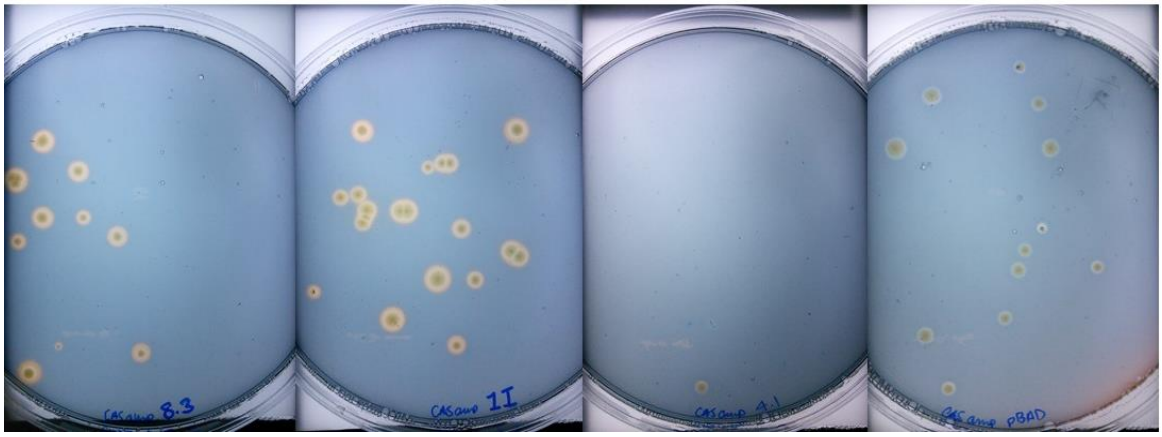

Figure 3.5: Observed phenotypes of entD variant controls on CAS media.

Left to right: 8.3, 1I, 4.1, empty pBAD. CAS media made as described by Louden et al with the addition of 2 mM $MgSO_4$ and 100 µM $CaCl_2$. Plates were incubated at 37 °C for 72 h and then BpsA(P) expression induced via addition of 150 µL of 10 % IPTG (v/v) beneath the agar and incubation at room temperature for 24 h.

All controls exhibited the desired interaction with the CAS media; variants 8.3 and 1I had large, bright halos that were easily distinguished from the smaller, paler halos of the variants with trace/no activity, 4.1 and empty pBAD, respectively (Figure 3.5). Following this apparent validation of the CAS screen, approximately 3 x 10^4 library clones were plated onto the media and incubated. When growing on this media it was observed that there was a large discrepancy between the number of cells plated and the resulting colony count. Using a volume of transformants that yielded approximately 1000 colonies on the control media, only between 10 – 50 colonies were observed on CAS plates, even after prolonged 37 ˚C incubation beyond 72 h (at which point colonies were already overgrown). As a result of this, only about 700 variants were screened during this attempt. This was very detrimental to the screen as it meant that the majority of the diversity present in the library was not being accessed and thus the chance of hit discovery was substantially decreased. Reflecting this, only 10 colonies (three examples are shown in Figure 3.6) had the phenotype required for selection and were taken to second-tier screening.

Figure 3.6: From left to right: entD variants C1.1,1.2 and 1.3 selected using CAS media for second-tier screening.

CAS media made as described by Louden et al with the addition of 2 mM $MgSO_4$ and 100 µM $CaCl_2$. Plates were incubated at 37 °C for 72 h and then BpsA(P) expression induced via addition of 150 µL of 10 % IPTG (v/v) beneath the agar and incubation at room temperature for 24 h.

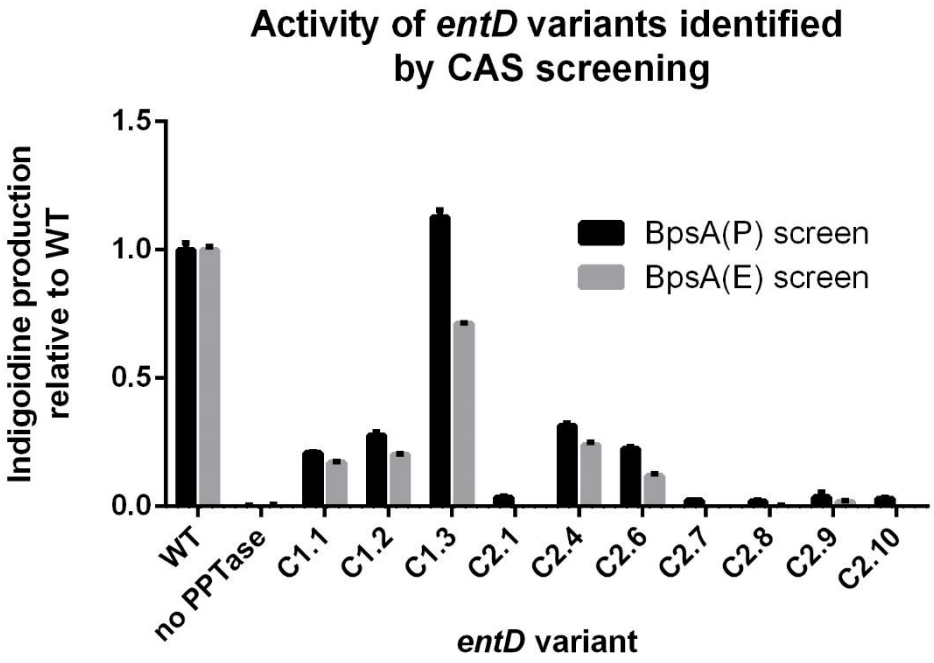

Figure 3.7: Second-tier screening entD variants selected from CAS screening.

Colonies selected from first-tier screening were used to inoculate 200 µL overnight cultures (LB, appropriate antibiotics) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate wells in a new 96-well plate. Plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader at 24 and 48 h time points. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values.

Of these 10 variants, 5 had again escaped the selection process and been selected because they possessed a trace level of activity undetectable by BpsA (Figure 3.7). This was surprising following the pilot trials of this screening method in which *entD* 4.1 (which showed trace activity undetectable by BpsA) lacked a siderophore-indicating halo. It appears that within this subset of PPTase variants there remains a range of activity levels and that CAS media screening has a resolution capable of removing the more active of these variants, but not sufficient to remove all false-positives.

3.2.3.3 Pre-screening for BpsA(E) active *entD* variants

In an attempt to prevent the selection of *entD* variants lacking BpsA(E) activity, a more stringent two-step approach was taken. This involved the plating of 3 x 10^4 clones, followed by the picking into a single 10 mL LB culture of every colony that remained white following induction of BpsA expression. This was roughly equal to 1.4 x 10^4 clones, or ~47% of the total number of colonies. This mixed culture was incubated for just a short period of 2 h (37 ˚C, 200 rpm shaking) in order to retain the high level of diversity. The mixed culture was then harvested and its plasmid DNA isolated, following which it was collectively used to transform EcoBlueE. The resulting transformants were plated to indigoidine production media and BpsA expression was induced following the appearance of colonies. Upon completion of this step, it was anticipated that colonies that turned blue would have no/reduced activity with BpsA(P) but have retained BpsA(E) activity, and thus the pool of variants would be pre-enriched by using indigoidine production as a reporter for *entD* – BpsA(E) activity rather than growth on iron-restricted media. About 300 colonies appeared blue in this screen and 94 of the darkest blue colonies were selected for second-tier screening (a number that could be easily miniprepped and transformed using 96-well plates, with room for positive and negative controls).

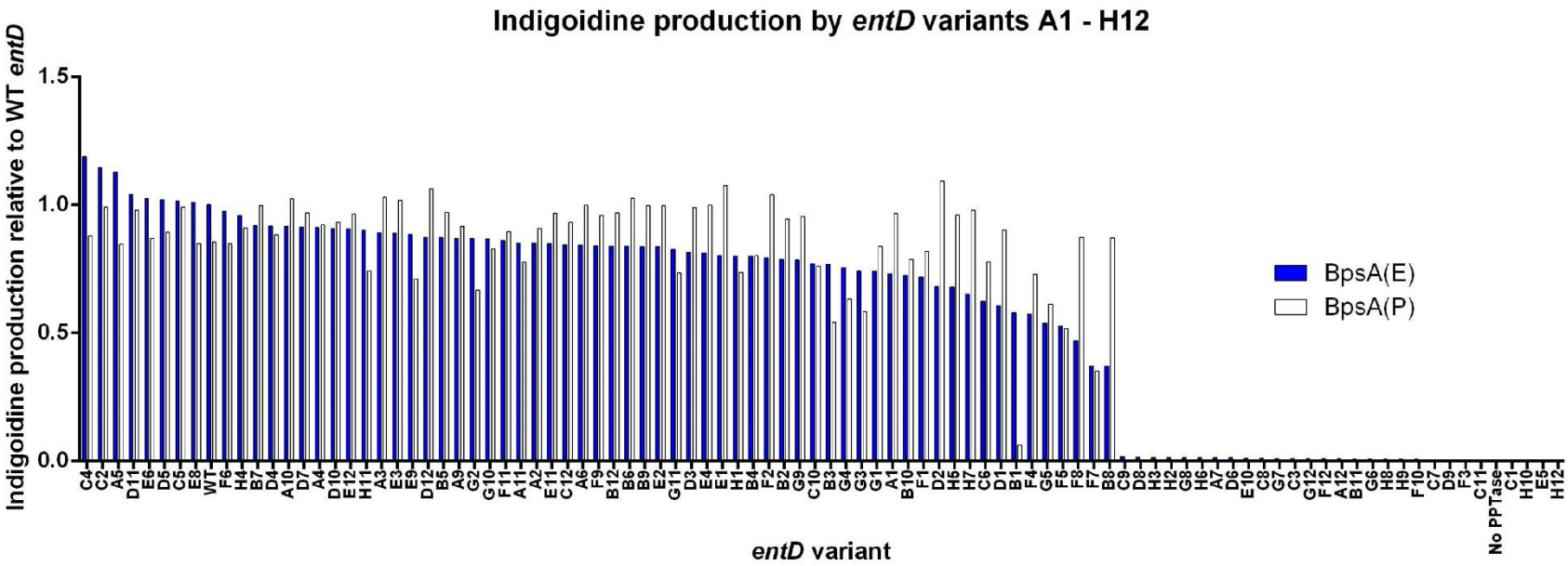

Figure 3.8: Second tier screening (Section 2.6.2.2) of *entD* variants identified following selection by BpsA(E) activity. Clones ordered by BpsA(E) activity.

Blue colonies selected from BpsA(E) screening were used to inoculate 200 µL overnight cultures (LB, Amp, Spec) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate 96-well plates. Plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader at 24 and 48 h time points. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values.

Despite this theoretically stringent approach, the profiles of the clones investigated (Figure 3.8) resembled those of the previous library screening attempts. Sixty-six clones reported high levels of activity with both BpsA substrates and 28 clones were indistinguishable from the no-PPTase negative control. One clone, B1, appeared to have the desired activity but was revealed to be a false-positive upon retransformation and re-screening (Figure 3.9). Clone 1D from an earlier screening round was also investigated here as it appeared to possess substrate specificity, although in reverse, i.e. activity with BpsA(P) but not BpsA(E).

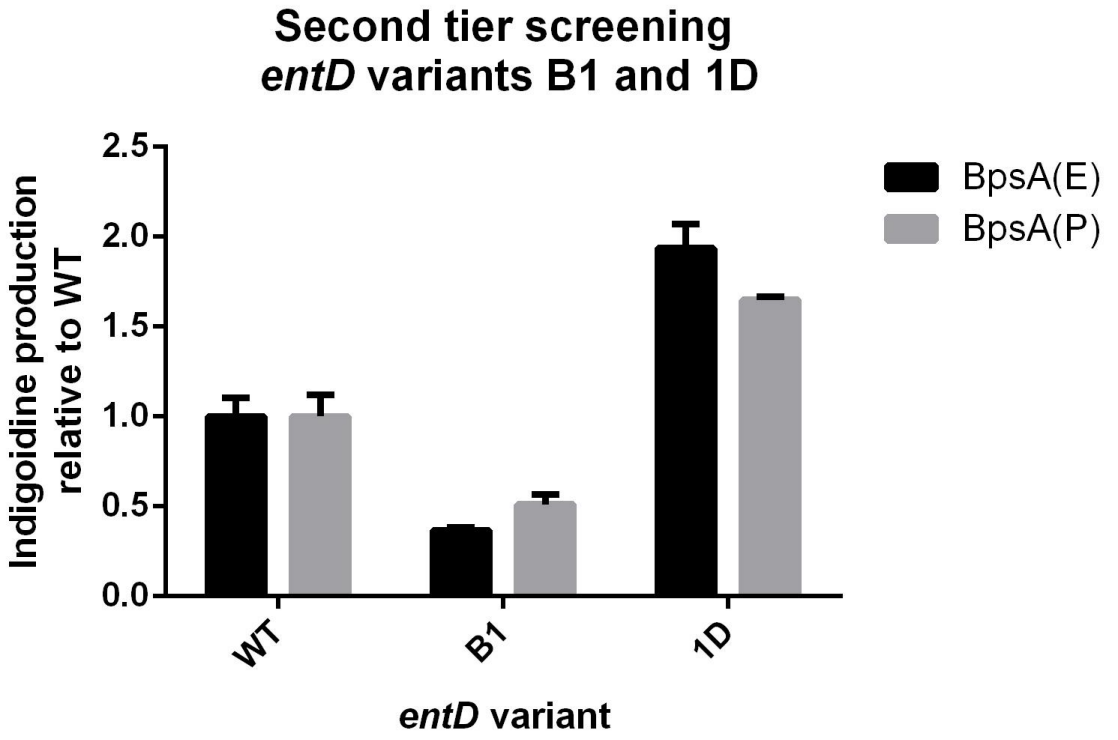

Figure 3.9: Second-tier screening of entD variants B1 and 1D following retransformation to screening strains. Neither variant showed the activity it appeared to have in the first round of screening.

Glycerol stocks of EcoBlue(E) and EcoBlue(P) each containing *entD* variant B1 or 1D were streaked to LBA plates (Amp + Spec) and incubated at 37 °C overnight. Single colonies from these plates were then used to inoculate 200 µL overnight cultures (LB, appropriate antibiotics) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate wells in a new 96-well plate. Plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader at 24 and 48 h time points. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values.

The rescreening of clones B1 and 1D to prove their false-selection highlights the inadequacies of the screening method applied to these libraries. At this point it was decided that the strategy for PPTase evolution needed considerable re-evaluation and therefore no more libraries were constructed.

3.3 Conclusions and Discussion

A directed evolution library can only be as good as the screen used to identify hits therein. The work in this chapter highlighted this statement, as it was consistently observed that the implemented screen had two major drawbacks that allowed non-desirable variants to be selected for further rounds of screening. The first drawback is a result of the nature of the interaction between PPTase and NRPS enzymes; only a very low level of PPTase activity with an NRPS is necessary to achieve a biologically relevant level of NRPS product formation *in vivo*. Here it was observed that a level of PPTase activity that was undetectable by BpsA activation provided sufficient EntF activation to enable a level of enterobactin production that could support growth under iron-restricted conditions. Colonies were selected on this basis and the second-tier screen revealed that many of these "hits" were, in fact, almost non-functional due to introduction of deleterious mutations. This "leak" in the selection strategy led to considerable wasted effort and must be prevented if the efficient selection of variants from a PPTase mutant library is to prove successful. In an odyssey of troubleshooting, several screen modifications were implemented in attempts to mitigate this problem, yet none reduced the proportion of activity-reduced variants being identified in second-tier screening. It became clear that the scale of the window of required activity is much smaller the scale of change that can be made by varying the parameters of the screen used in this project. We predict that the reason such a low level of PPTase activity is required is because a single activated NRPS can continually produce its product until it is inhibited by either lack of substrates, loss of the PPant arm, or inhibition by the product itself (Finking et al., 2004). The lesson learned here regarding persistent NRPS activation is perhaps one that should be considered whenever performing experiments that utilise an NRPS product as a reporter for a certain activity. Possible future strategies to overcome this problem are considered in Chapter 5.

The second flaw which was of detriment to the screening of *entD* libraries occurred in both first- and second-tier screens and was the variance in indigoidine production that was observed variants possessing comparable levels of activity. In the first-tier screen, a high proportion of colonies that appeared white on plates (suggesting loss of BpsA(P) reactivity), and were carried through to second-tier screening in EcoBlueP, produced substantial levels of indigoidine via BpsA(P) activation, a phenotype conflicting with the results of the first-tier screen. Variants of this type comprised 40 – 60% of all variants that were screened in liquid media. The cause of the colonies that remained white on the first-tier screening plates is unclear. It is unlikely that BpsA expression was not being induced because the method of induction that was used is routinely used in our lab as a reliable method for inducing all colonies on a plate. A possible explanation is that deleterious mutations could be occurring in the BpsA gene which prevent indigoidine production. The rationale for this proposal lies in the fact that indigoidine is mildly toxic to *E. coli* which provides a selection pressure for the cells to lose the ability to produce the blue compound (Owen et al., 2011). The rate at which these falsely-white colonies occurred, however, was far greater than the typical rate of BpsA loss-of-function that is observed in our laboratory (Unpublished Observations). Without external mutagenesis, it seems unlikely that all these colonies could have mutated to provide themselves a slight growth advantage. To analytically rule out this possibility, however, sequencing of the BpsA gene and its expressional control regions from a large number of these variants would be required. Issues with apparent variance in indigoidine production between equivalent strains also arose in the second-tier screen, albeit to a lesser extent (e.g. *entD* variants 1D and B1 (Figure 3.9), which appeared to have substrate-selective activity in second-tier screening but did not retain their respective profiles upon re-screening).

In its current state, the screening strategy used to investigate PPTase libraries in this project is yet to be shown to work. Both flaws described so far in this section must be addressed to functionalise the screen, as both cause substantial waste of screening effort via generation of false-positive hits.

As described in Section 3.1.3, using the screen that she had developed, Dr. Robins generated what appeared to be substrate-selectivity of *entD*. Due to the 200-fold loss of activity associated with this selectivity, however, she theorised that the level of PPTase activity might have been reduced non-specifically to the edge of the BpsA detection range and in combination with the natural lack of EntD affinity for BpsA, this loss of activity would appear as substrate-selectivity (i.e. EntD – BpsA activity below detectable range; EntD-BpsA(E) activity marginally above detectable range). The work performed in this chapter lends evidence to this theory as the only time any (supposed) substrate selectivity was observed, a drastic loss of activity accompanied it (see *entD* variants 2F, 3A Figure 3.3). The combination of these two separate projects demonstrate the need for a novel screen that does not rely on a comparison of enzyme activity, but rather one that can quantitatively measure the activity of individual variants at a greater resolution than is achievable using BpsA. Development of a novel screen addressing the issues associated with the screen used in this project could realise the value of some of the variants generated in this project. Those that seemed to exhibit a slight shift towards substrate-selectivity could be the starting point for secondary libraries to be fed through the new protocol.

We foresee that an *in vitro* screening system may be the best way forward for developing a novel screen for substrate-selective PPTases. *In vitro* screens tend to benefit from more precise detection ranges and quantitative reporting – two factors lacking in the BpsA-based experiments. An outline for one possible screen is described in Chapter 5.

The carrier proteins of EntF and PvdD were chosen for use in this study as they are both activated by type II PPTases and were both available already in our laboratory. However, their selection could have been a limiting factor to the endeavours of this project. The tag regions of the EntF and PvdD are naturally very similar (see tags E12 and P12, Table 4.1) and it is interaction of hydrophobic PPTase residues with hydrophobic residues inside this tag region that are largely responsible for PPTase-CP-domain interaction (Tufar et al., 2014). Due to the lack of exploitable

sequence divergence between these two substrates, the low numbers of point mutations introduced by the epPCR protocol used in the current project may be insufficient to generate the change required to develop specific interactions outside this region. Furthermore, a mutational strategy targeting PPTase tertiary structure may be a better approach, to exploit differences that are not at the protein sequence level. For future evolution attempts aiming to generate substrate-selectivity, starting with PPTases that naturally activate more sequence-distinct CP domain substrates may provide a better template for evolution.

At this point, it was decided not to begin work on the second aim of this project (i.e. directed evolution of PcpS towards substrate-selectivity) because the screen was clearly not capable of the task required of it. If the proposed methods of Section 5.1 prove effective, this aim will be investigated then.

Chapter 4: Design, analysis, and evolution of small peptide tags for post-translational modification by PPTases.

4.1 Introduction

Perhaps the most beneficial discovery for the future prospects of PPTase labelling applications was the identification of the YbbR tag which, despite its small size (11 aa), is able to act as an efficient substrate for the PPTase Sfp (Yin et al., 2005). This discovery subsequently inspired a bottom-up synthetic library based method for generation of novel tags functional as substrates for either Sfp or the type 1 *E. coli* PPTase, AcpS (Zhou et al., 2007). This library contained 1 x 10^9 variants and ultimately yielded two tags, A1 and S6 for use with AcpS and Sfp, respectively. This method of tag discovery required considerable screening effort using phage display techniques. Importantly, because the two PPTases used in screening this library belong to different families, a level of orthogonality was achieved, allowing the iterative labelling of two distinct protein targets. The work in this chapter sought to build on these discoveries by taking note of the structure and sequence motifs of the current tags and to utilise this information to design a novel tag discovery method without the need for extensive peptide libraries.

4.1.1 Designing tags from carrier protein domains

The YbbR tag that was discovered via *B. subtilis* genome fragmentation and phage display demonstrates that it is possible for native components of proteins to act as substrates for PPTases. Subsequently, the library from which the A1 and S6 tags were identified was constructed from the base sequence GDS(L/I)XXXXXXXX (X = any of the 20 protein residues). The motif at the N terminus of the tag library is largely conserved in natural carrier proteins of NRPS, PKS, and FAS enzymes

(Beld et al., 2014). The aspartic acid residue at position two is, however, often exchanged for a histidine residue, or less commonly, an asparagine residue.

Interestingly, functional tags identified within this library significantly differed in sequence from any sequences found within the native carrier proteins with which Sfp or AcpS interact. Surrounding the invariant serine that receives the PPT arm, however, there is a strong conservation of polarity at the majority of positions within these tags (Figure 4.1).

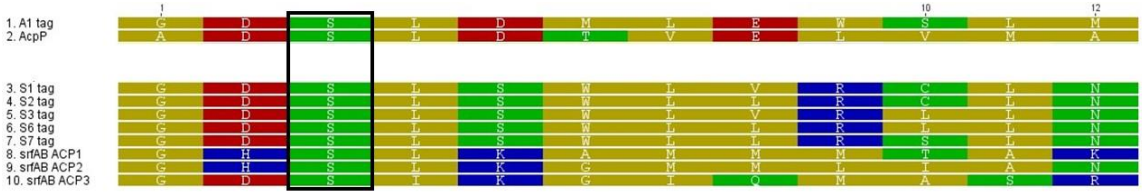

Figure 4.1: Geneious alignments of functional tags to the native substrate of their respective PPTase showing the conservation of polarity in these tag sequences.

Seq. 2: AcpS native substrate, seq. 9 & 10: native substrates of Sfp. Gold: Non-polar; green: polar, uncharged; red: negatively charged; blue: positively charged. The boxed residue is the pantetheinylated serine.

As an alternative to this bottom-up library approach we theorised that it may be possible to excise candidate tag sequences directly from the native carrier protein substrate of any given PPTase.

4.1.2 BpsA as a model NRPS for testing of candidate tags

BpsA has been used extensively in our lab for understanding the function and interaction of NRPS domains (Owen, 2010). The characteristics of BpsA as a relatively small, simple, and easily purifiable NRPS, coupled with its ability to produce a coloured product, indigoidine, make it an excellent platform for manipulation and functional investigation of NRPS components. Here we envisaged that it could be used as a proof-of-concept platform to show that a putative tag could be recognised and modified by a PPTase; an event visually reported by indigoidine production. The experiments described in this chapter were the first time that the tag region has in isolation been modified in BpsA, therefore these experiments also served to give a better understanding of

the flexibility of BpsA to perform its function with this relatively small, yet significant alteration. Of particular interest was the importance of the CP domain surrounding the tag region for PPTase recognition and interaction. Because PPTases and CP domains have evolved cooperatively, we predicted that removal of the tag sequence from this CP domain scaffolding would reduce the affinity of the tags for a PPTase partner.

4.2 Aims

1. To design a range of candidate tag sequences derived from the CP domains of the NRPS enzymes PvdD and EntF.
2. To assess the practicality of, and then develop, BpsA as a tag-validation system.
3. To assess the ability of the candidate tags within BpsA to act as substrates for their partner PPTases.
4. To evaluate the importance of rest of the CP domain outside the tag region for PPTase binding.

4.3 Results

4.3.1 Tag design

From each enzyme, PvdD and EntF, six peptide tags of varying length were designed (Figure 4.1). These tags comprise parts of the second alpha helix of their respective carrier protein origin.

Origin	Tag	Sequence
PvdD	P12	GH**S**LLLLMLKER
	P11	GH**S**LLLLMLKE
	P10	GH**S**LLLLMLK
	P9	GH**S**LLLLML
	P8A	GH**S**LLLLM
	P8B	H**S**LLLLML
EntF	E12	GH**S**LLAMKLAAQ
	E11	GH**S**LLAMKLAA
	E10	GH**S**LLAMKLA
	E9	GH**S**LLAMKL
	E8A	GH**S**LLAMK
	E8B	H**S**LLAMKL

Table 4.1: Putative tag sequences designed for this project.

Shown in bold is the serine residue which receives the PPant arm.

PvdD and EntF were selected for designing tag sequences from because they are both activated by type II PPTases that can be purified and tested *in vitro.* Type II PPTases are known to be able to activate BpsA, whereas AcpS cannot, a characteristic we predicted to extend to other type I PPTases. Furthermore, the best characterised tag is A1; NMR studies have revealed the specific residues within the tag that directly interact with the associated type I PPTase, AcpS. Using PvdD and EntF templates contrasts with this, as they are both associated with type II PPTases and so it will be interesting to see how the lessons learned with A1 translate to the tags generated in this chapter.

The longest tag length was chosen to be 12 amino acids because all current functional tags are ≤12 amino acids. Also contributing to this decision was the fact that one of the major advantages of the tag labelling system is that short tags can provide a non-disruptive addition to the protein of interest, thus this upper limit adhered to this principle. The shorter limit was set at 8 amino acids because previous tag testing revealed that tags shorter than this had undetectable or severely impaired reactivity with PPTases (Yin et al., 2005; Zhou et al., 2008).

4.3.2 Using BpsA to validate tag function

pTSWAP was used to generate BpsA variants in which the native tag region was replaced with a putative tag. This construct lacks a functional T-domain and therefore cannot produce indigoidine. Two gBlocks™ encoding the BpsA T-domain with the P12 or E12 tag were cloned into this construct at NsiI and SpeI restriction sites. Following sequence verification, the BpsA(P12) and BpsA(E12) constructs were co-expressed in BL21 *ΔentD* cells with a partner PPTase (either PcpS or EntD, respectively, from pET28a(+)). Following induction of BpsA and PPTase expression via addition of IPTG, colonies from both BpsA variant plates turned blue, indicating successful recognition of the tags by their respective PPTase and functional activity within the recombinant BpsA enzymes (Figure 4.2). This result was highly encouraging for the intended use of BpsA as a tag validation platform.

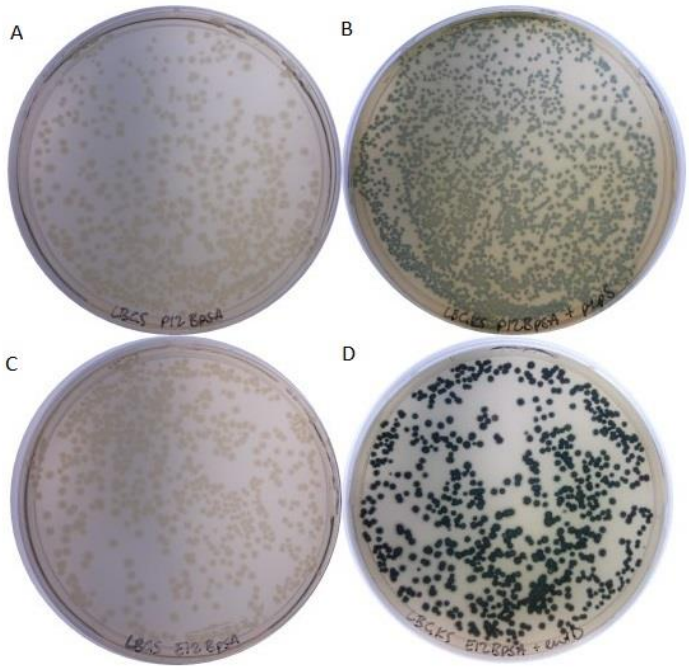

Figure 4.2: Testing E12 and P12 tag function using BpsA and PcpS or EntD.

Function reported by indigoidine production following IPTG induction of BpsA and PPTase expression. A: BpsA(P12), no PPTase, B: BpsA(P12) with PcpS, C: BpsA(E12), no PPTase, D: BpsA(E12) with EntD. Plates were incubated at 37 °C overnight and then at room temperature for 32 h following induction.

The interaction of BpsA(P12) with PcpS yielded less indigoidine production than BpsA(E12) activated by EntD. This was interesting because PcpS typically activates BpsA(WT) efficiently and a high level of indigoidine is rapidly produced.

To probe the function of these tag-swapped BpsA variants more comprehensively, the BpsA(E12) and (P12) constructs as well as EntD and PcpS were purified and their kinetic parameters determined *in vitro*. Positive control reactions for comparison were with BpsA(P) and BpsA(E), in which the entire CP domain has been substituted for the CP domain of PvdD or EntF, respectively (followed by directed evolution to optimise CP domain function (Owen et al., 2016; Owen et al., 2012)), and with WT BpsA. All reactions contained 1 µM BpsA variant and 1 µM PPTase except PcpS + BpsA(P), where 1 µM PcpS caused too rapid a reaction to measure accurately and therefore the PPTase concentration was dropped to 0.25 µM.

The results of the *in vitro* assays (Figure 4.3, Table 4.2, Table 4.3) were consistent with the previous i*n vivo* plate-based test (Figure 4.2); both (E12) and (P12) BpsA variants were activated by both EntD and PcpS and subsequently produced indigoidine. The level of indigoidine produced was also consistent with the previous plate-based observations, in that BpsA(P12) performed worse than BpsA(E12) when activated by either PPTase. In each instance, the BpsA-tag constructs were out-performed by their corresponding positive control CP domain-swapped BpsA constructs when activated by their native PPTase, i.e. BpsA(E12) < BpsA(E) when activated by EntD, and BpsA(P12) < BpsA(P) when activated by PcpS. A surprising result was that the combination of PcpS and BpsA(E12) exhibited greater indigoidine production than PcpS with BpsA(E).

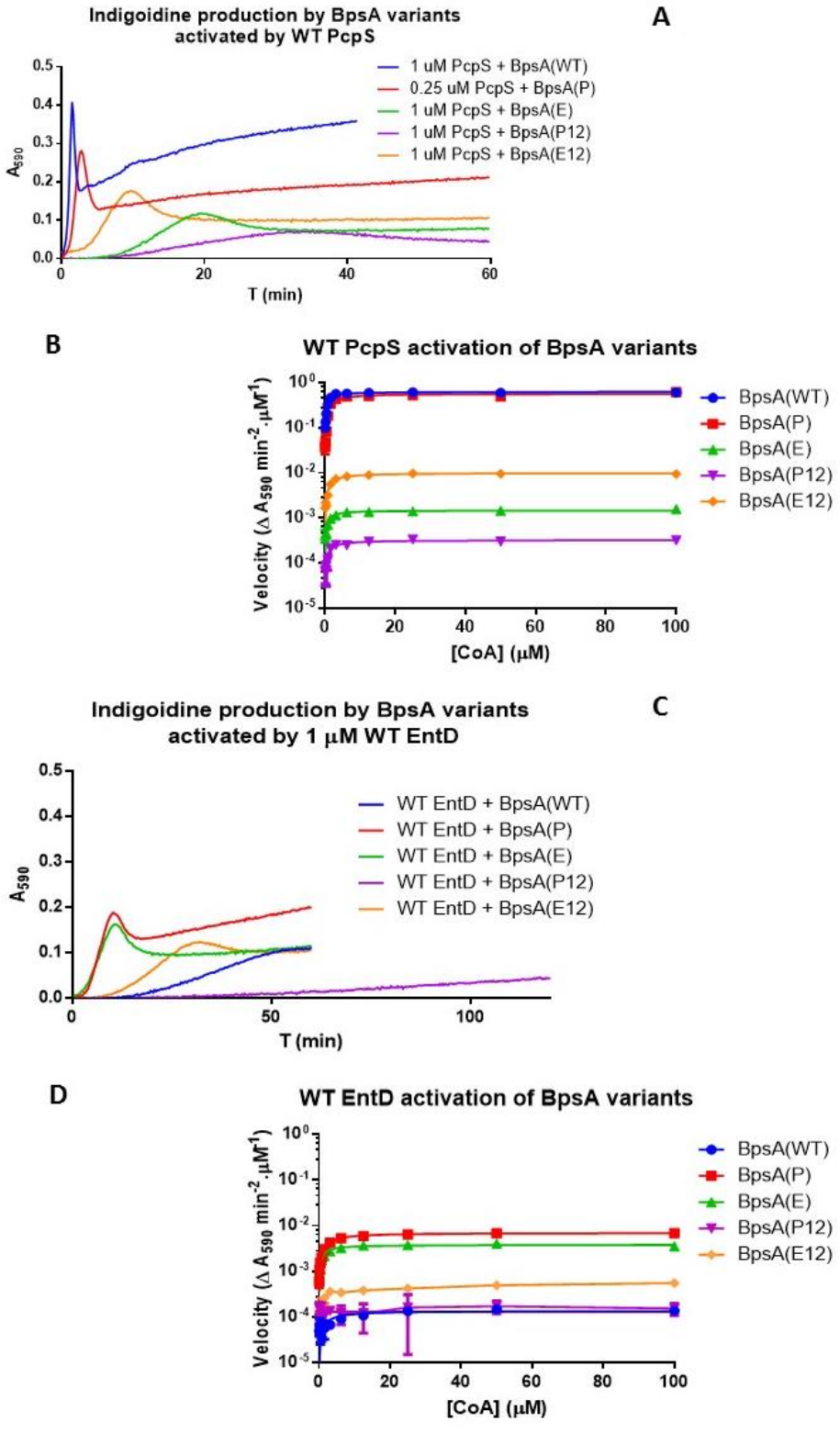

Figure 4.3: In vitro indigoidine production and kinetic profiles of BpsA variants.

Assays were conducted and data analysed as described in Section 2.6.2.3 and in triplicate. A + C: Indigoidine production by BpsA variants with 100 μM CoA activated by PcpS or EntD. B + D: Michaelis-Menten kinetic profiles of BpsA variants activated by either PcpS or EntD. Error bars show std. error.

PcpS	K_m (nM)	V_{max} (ΔA_{590} $min^{-2}.\mu M^{-1}$)
BpsA(WT)	650 ± 42	0.63 ± 0.007
BpsA(P)	1300 ± 180 *	0.57 ± 0.016 *
BpsA(E)	720 ± 110	0.0014 ± 0.00004
BpsA(P12)	1060 ± 170	0.00032 ± 0.00001
BpsA(E12)	1200 ± 130	0.0099 ± 0.0002

Table 4.2: Kinetic values of PcpS activation of BpsA variants titrated against CoA concentration.

Values were calculated from A_{590} values using GraphPad Prism. Standard error shown. *BpsA(P) + PcpS reaction was performed using 0.25 uM PcpS. All values rounded to 2 s.f.

EntD	K_m (nM)	V_{max} (ΔA_{590} $min^{-2}.\mu M^{-1}$)
BpsA(WT)	1700 ± 560	$0.00013 \pm 9.5e^{-6}$
BpsA(P)	1850 ± 109	$0.0069 \pm 9.1e^{-6}$
BpsA(E)	880 ± 96	$0.0037 \pm 8.2e^{-5}$
BpsA(P12)	**	**
BpsA(E12)	950 ± 190	$0.00046 \pm 1.9e^{-5}$

Table 4.3: Kinetic values of EntD activation of BpsA variants titrated against CoA concentration.

Values were calculated from A_{590} values using GraphPad Prism. Standard error shown. ** Reaction was too slow to calculate these values accurately. All values rounded to 2 s.f.

These assays were based upon PPTase kinetic determination assays (Section 2.6.2.3) and therefore represent a biphasic reaction consisting of PPTase activation of BpsA, followed by indigoidine synthesis. This means that there are two major factors determining the outcome of each assay: the ability of the PPTase to recognise and modify the substrate, and the ability of BpsA to produce indigoidine. The tag region is involved in both reactions and therefore the exact reason for the level of indigoidine produced is likely to be a combination of the two factors and therefore

is unclear. Despite this ambiguity, it can still be concluded that the E12 and P12 tags can be substituted into the BpsA tag region and function is retained, albeit impaired by varying amounts relative to the positive controls.

4.3.3 BpsA competition assays to assess PPTase recognition of tags

In order to exclusively focus on PPTase – tag reactivity, a competition assay was implemented (Section 2.7). Isolated E12 and P12 peptide tags were generously synthesised by Prof. Gavin Painter and provided to this study. These tags were co-incubated with BpsA variants, a PPTase and a limited pool of CoA and after a period of time indigoidine synthesis was initiated via the addition of L-Gln and ATP.

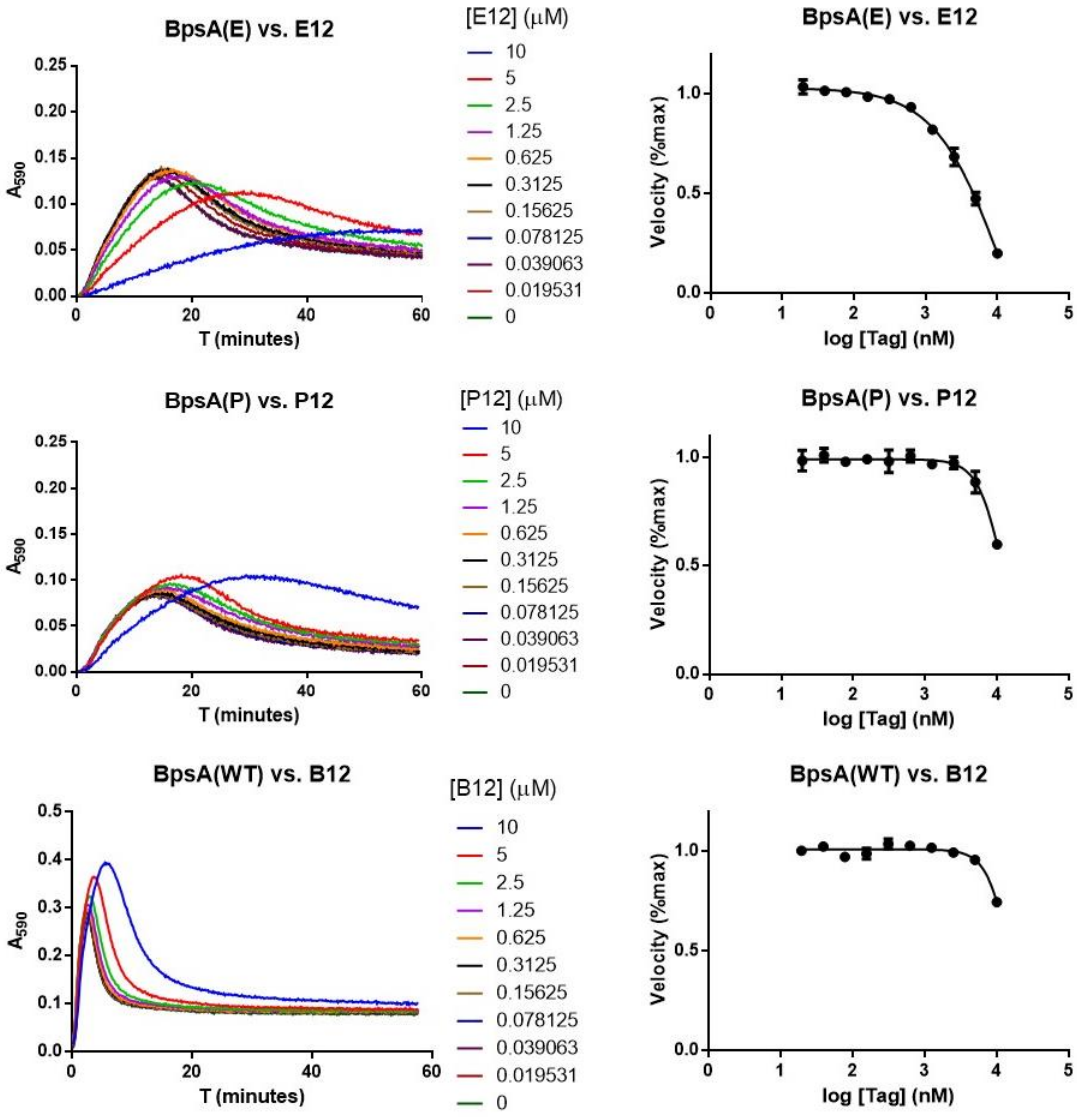

Figure 4.4: Competition assays between isolated peptide tags and BpsA variants.

Assays were performed in triplicate in 96-well plates. 50 µL volumes of peptide tag/tagged protein were serially diluted 2x with ddH_2O across the plate, after which 50 µL of reaction mix containing 2 µM BpsA, 0.75 µM CoA, 40 mM $MgCl_2$, 400 mM TrisCl pH 8.0 was added to each well. To begin BpsA activation, 50 µL PPTase was then added to each well at a final concentration of 0.25 µM (PcpS) or 0.5 µM (EntD). Reactions were incubated at 30 °C, 200 rpm for 40 min (PcpS reactions) or 75 min (EntD reactions), after which 50 µL of an ATP (4 mM) and L-Gln (4 mM) mix was added and the plate mixed at 1000 rpm for 10 s to initiate indigoidine production. Two control reactions were used, one containing no PPTase (-ve) and one containing no competing peptides (+ve). Indigoidine production was quantified by measuring absorbance at 590 nm at 10s intervals for 30 – 60 mins with a microplate reader. The slope function of Microsoft Excel was used to calculate the velocity of the reactions and Prism GraphPad used to derive kinetic data. Error bars show the standard error.

Tag	IC_{50} (nM) vs. BpsA with entire CP swap	Ratio of tag:BpsA required for 50% inhibition (IC_{50}/[BpsA])
E12	4340 (3940 – 4770)	8.7:1
P12	11900 (10900 – 13000)	23.8:1
B12	14700 (12700 – 17100)	29.5:1

Table 4.4: IC_{50} values for peptide tags (95% confidence intervals)

Values (rounded to 3 s.f.) calculated from the rate of A_{590} change using GraphPad Prism and Microsoft Excel.

Both the E12 and P12 tags were able to inhibit BpsA activation, although it was observed that a relatively high proportion of tag:BpsA was required for this inhibition to become apparent (Table 4.4).

In accordance with the previous section's results, the E12 tag performed better as a PPTase substrate than the P12 tag, which required an approximately 3 x higher ratio of tag:BpsA for a similar level of inhibition.

An additional tag, B12 (GNSLIAVGLVRE), comprising the WT BpsA tag region was also designed and ordered from GenScript. A competition assay between this tag and BpsA(WT) using PcpS as the activating PPTase was employed as a control to assess the involvement of the CP domain outside the tag region in PPTase binding. As with the other competition assays, this assay showed that a high proportion of tag is required to compete with a native CP domain.

Together these experiments suggest that the CP domain surrounding the tag region plays an important role in PPTase binding. Were this not the case, one would expect to observe 50% inhibition at a tag:BpsA ratio near 1:1. The variance in ratio between the tags also suggests that there is variation in the level of contribution that different CP domains provide in different NRPS systems.

From these results, it was concluded that the E12 and P12 CP domain-derived tags can indeed function as PPTase substrates, however their isolation from the CP domain results in a loss of affinity for PPTase binding.

In a real-world context, these tags would be used fused to other proteins, therefore, to more realistically assess tag function the tags were fused to the C-terminus of *E. coli* maltose-binding protein (MalE) by a PCR method using primers that fully encoded each candidate tag sequence. All 12 tags were fused this way and the resulting constructs purified using French Press and Ni affinity column methods. These fusion proteins were then assessed in BpsA competition assays. At the highest concentration tested (20 times the molar concentration of BpsA) no inhibition was observed in any assay. The reason for this negative result remains unclear and future optimisation work is planned to address it.

4.3.4 Directed evolution of BpsA containing tag-swapped CP domains to improve indigoidine production

The tag-swapped BpsA kinetic results reported above served the purpose of validating tag function, however reaction rates for the two tags investigated were very slow and not directly quantitative because the rate is potentially influenced by two distinct factors as described in Section 4.3.2. It was therefore of interest to investigate whether the observed impairment of BpsA function could be recovered by introduction of additional mutations outside the tag sequence. To address this, epPCR libraries were constructed from the BpsA CP domain into which the E12 and P12 tags had been substituted. It was hoped that any mutations associated with improved function might give insight into the mechanism(s) resulting in impaired improvement, and possible methods for improving the tag validation system. These libraries were then screened in *E. coli* Δ*entD* over-expressing either *entD* or *pcpS* (encoded on pET28a(+)) on solid indigoidine production media. Library plates were compared to plates upon which BpsA(E12) or (P12) “template”

expressing cells grew. Colonies that turned visibly blue before the control plates were picked for second-tier screening in liquid media. Top performing variants were then sequenced and aligned.

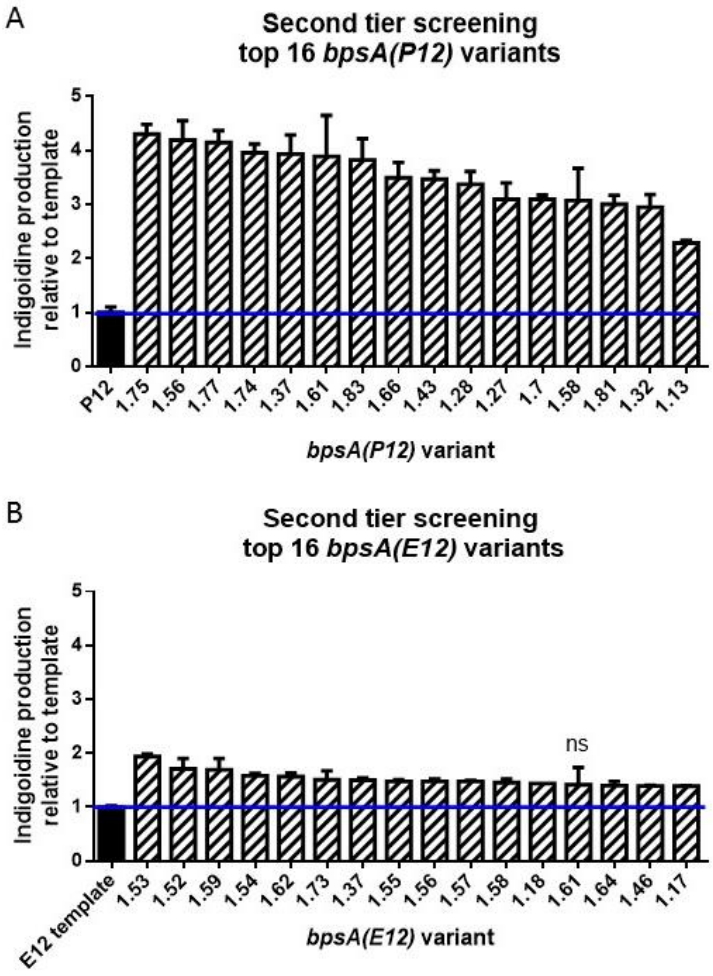

Figure 4.5: Indigoidine production by top evolved (A) bpsA(P12) and (B) bpsA(E12) variants shown as fold increase over template level.

Colonies selected from first-tier screening were used to inoculate 200 µL overnight cultures (LB, appropriate antibiotics) in a 96-well plate. Following incubation at 37 °C, 200 rpm shaking for 12-16 h, 20 µL of each overnight culture was transferred to 130 µL fresh indigoidine production media (LB, 115 mM L-Gln, 0.02% (w/v) L-arabinose, 600 µM IPTG) in triplicate wells in a new 96-well plate. Plates were incubated at 18 °C, 200 rpm. OD_{590} and OD_{800} readings were taken using a microplate reader after 24 h. OD_{800} readings allowed correction for cell density and meant that the same plates could be used for both readings. The WT PPTase being evolved served as the positive control and point of comparison in these reactions. A no-PPTase control containing an empty pBAD plasmid provided the OD_{800} values for subtraction. Triplicate sterile wells allowed subtraction of media OD_{590} and OD_{800} values. ($P<0.05$ for all variant data compared to template level unless indicated on graph, student's t test)

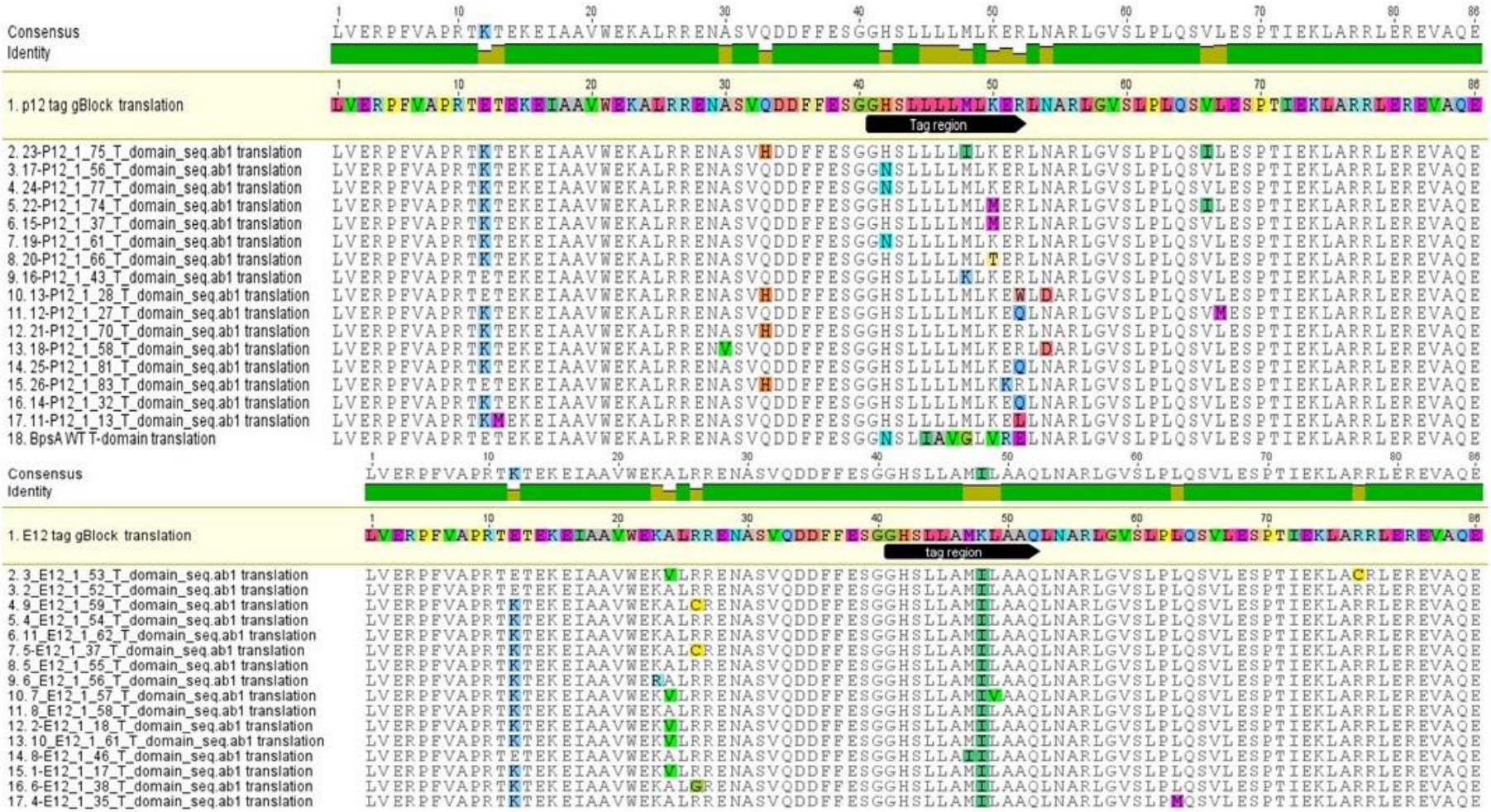

Figure 4.6: Geneious alignments of top variant sequences to template sequences.

WT BpsA CP domain sequence is shown for additional reference below P12 variant sequences. Differences from the template sequence resulting from the epPCR mutagenesis are highlighted (colours are arbitrary and simply distinguish different amino acids from one another).

After a single round of mutagenesis, substantial improvement in both libraries was observed (Figure 4.5). The top *bpsA(P12)* variants all exhibited a 3 to 4-fold increase in indigoidine production whereas the top performing *bpsA(E12)* variant had an almost 2-fold increase from the template level. The discrepancy in the scale of improvement between the two libraries is likely to be due to two major factors; first, PcpS activity is considerably higher than that of EntD, and the BpsA(P12) template activity was already more drastically impaired than the BpsA(E12) template. The combination of these characteristics meant that there was more room for improvement above the template level of indigoidine production for BpsA(P12) than for BpsA(E12). Second, PcpS naturally has greater activity with BpsA(WT) than EntD (Figure 4.3, graphs A and C). Earlier results and previous studies (Owen et al., 2011; Owen et al., 2012; Robins, 2016)(Section 4.3.3) show that interaction with the CP domain outside the tag region is important for recognition by a PPTase, an interaction that naturally occurs efficiently between PcpS and BpsA, but not as efficiently between EntD and BpsA. The results here demonstrated that BpsA function following tag swapping could be somewhat restored by random mutagenesis to a level dependent on the PPTase being used. An additional round of evolution was employed for each enzyme, using pooled top variants as the template sequences, however no significant improvement above round 1 selection was observed. Furthermore, the level of activity that top round 1 variants possessed was near the maximum activity that the *in vivo* screen could identify, therefore a new screening strategy would be required to take this work further.

The sequences of the top variants in both libraries (Figure 4.6) contained several highly conserved mutations, suggesting that they were important to the observed increases in activity. The mutation E941K occurred almost in saturation in both libraries, A954V occurred 5 times among the top E12 variants and Q962H occurred 4 times among the top P12 variants. All of these mutations fall outside the tag region. Only a single mutation, K978I, was conserved within the tag region, occurring in all top *bpsA(E12)* variants. Currently validated tags typically contain a non-polar residue in this position which is likely to be involved in PPTase recognition and binding

through hydrophobic interactions, therefore this shift is not particularly surprising. OF note, the tag region of the EntF CP domain naturally bears a positive group in this position. Due to the saturation of this tag-region mutation in the *bpsA(E12)* library hits it was theorised that this mutation may improve E12 tag function when present in isolation. The E12 tag region containing the K978I mutation was termed E12mut1 and ordered from GenScript as an isolated peptide tag for use in a BpsA competition assay. Disappointingly, the E12mut1 tag did not show a decrease in IC_{50} compared to the native E12 tag (Figure 4.7).

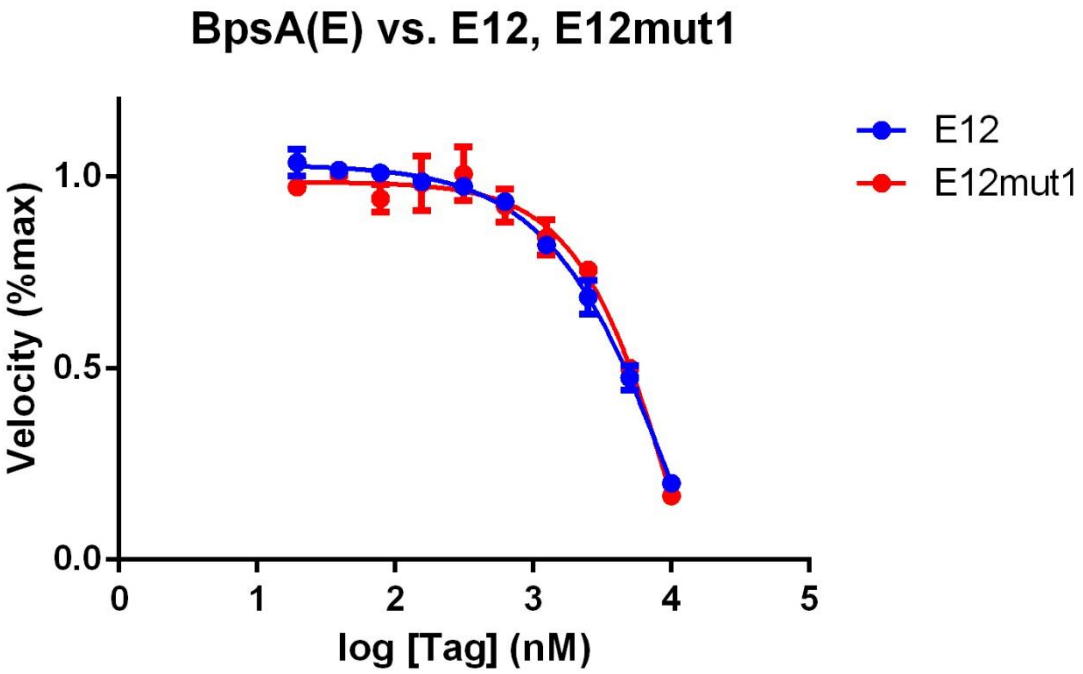

Figure 4.7: Competitive inhibition of BpsA(E) by E12 and E12mut1 peptide tags.

Assays were performed in triplicate in 96-well plates. 50 µL volumes of peptide tag/tagged protein were serially diluted 2x with ddH_2O across the plate, after which 50 µL of reaction mix containing 2 µM BpsA, 0.75 µM CoA, 40 mM $MgCl_2$, 400 mM TrisCl pH 8.0 was added to each well. 50 µL PPTase was then added to each well at a final concentration of 0.25 µM (PcpS) or 0.5 µM (EntD). Reactions were incubated at 30 °C, 200 rpm for 40 min (PcpS reactions) or 75 min (EntD reactions), after which 50 µL of an ATP (4 mM) and L-Gln (4 mM) mix was added and the plate mixed at 1000 rpm for 10 s to initiate indigoidine production. Two control reactions were used, one containing no PPTase (-ve) and one containing no competing peptides (+ve). Indigoidine production was quantified by measuring absorbance at 590 nm at 10s intervals for 30 – 60 mins with a microplate reader. The slope function of Microsoft Excel was used to calculate the velocity of the reactions and Prism GraphPad used to derive kinetic data. Error bars show the standard error. No significant difference was observed between the two reactions ($p<0.05$, multiple t tests).

This finding suggested that mutation K978I alone was not sufficient to increase the E12 tag's ability to act as a substrate for EntD and its saturation in the library is likely due to an alternative mechanism.

To investigate two of the other common mutations, E941K and Q962H, two additional BpsA constructs were created. Each of these constructs contained the WT BpsA CP domain scaffold bearing one of these two mutations. EntD activity with these constructs was measured to determine whether either mutation alone enhanced function. Interestingly, both constructs performed worse than WT BpsA (Figure 4.8), suggesting that these mutations are most likely helping BpsA to accommodate the E12 tag into its reaction mechanism, rather than promoting EntD binding, or else they have a synergistic effect with other mutations occurring in the library and improved activity is not observed when the mutations are present in isolation.

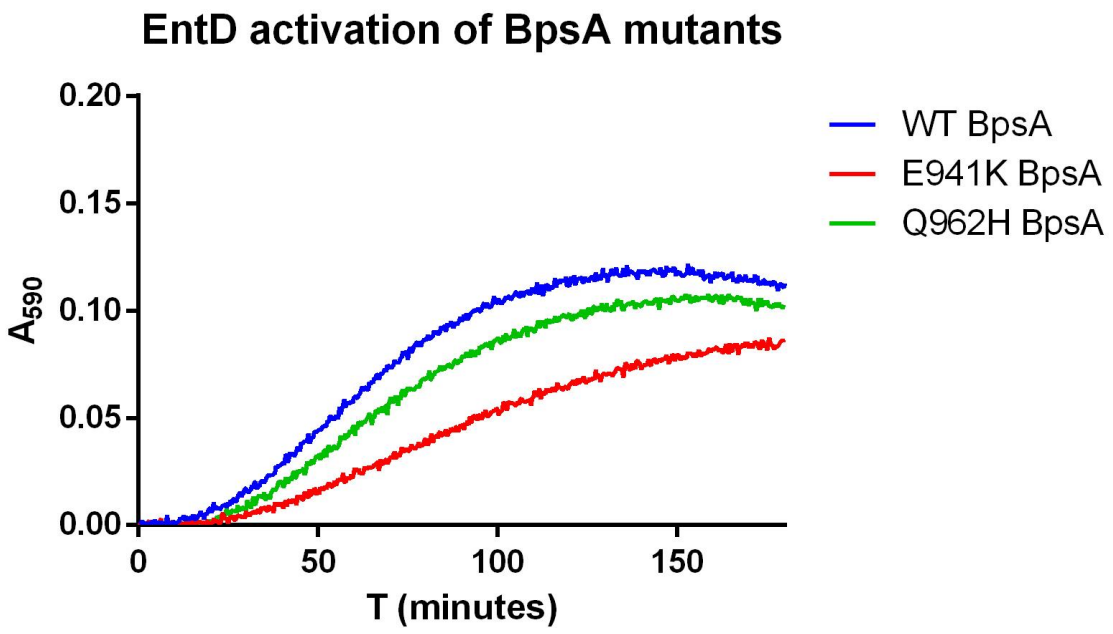

Figure 4.8: EntD activation of BpsA variants containing isolated mutations occurring frequently in evolution libraries.

Assays consisted of triplicate reactions spanning serial dilutions of CoA from 100 – 0.09765 µM in a 96-well plate. In addition to CoA, the 200 µL reactions contained 4 mM L-Gln, 1 mM ATP, 1 µM *apo*-BpsA (containing the appropriate CP), 1 µM variant PPTase. To begin the reaction, PPTase was added last and the plate shaken to mix at 900 rpm for 10s. A_{590} readings were taken at 20 s intervals for 1-2h depending on the PPTase-BpsA combination. The slope function of Microsoft Excel was used to calculate to the velocity of the reaction, and then again applied to these velocity values and the maximum values taken as maximum PPTase velocity. Prism GraphPad was then used to derive kinetic parameters.

4.4 Conclusions and Discussion

This chapter was a broad first look into the marriage of CP domain tags and BpsA, to investigate a novel method for designing tags and validating their function and to examine the poorly understood interaction between CP domains and PPTases. An initial result of this work was confirmation that the tag region of BpsA could be substituted for the tag of a foreign NRPS and function retained. The resulting function, however, was considerably impaired for both the tags investigated. As a non-quantitative tag validation system this was acceptable. It was pleasing to see that the level of indigoidine produced in the colony based *in vivo* test correlated accurately with the results observed in liquid media *in vivo* tests and in *in vitro* assays. Because this validation method is very fast and simple to implement and does not require expensive equipment I anticipate that it will be broadly applicable to laboratories interested in adopting the PPTase-tag system. There is, however, no guarantee that this system will work with other tag/PPTase pairs, especially those not belonging to NRPS-Type II PPTase systems. Further development of the system could therefore involve directed evolution of BpsA to make it more accepting of foreign tags. It is known that CP domains are highly dynamic, so beneficial mutations would probably relieve any structural impairment imposed by tag swapping, or to decrease the specificity with which PPTases are bound.

Validation of tag function confirmed the hypothesis that tags can be designed simply by excising a contiguous sequence region including the conserved serine from its cognate CP domain. This is useful, as it bypasses the necessity for large tag libraries as have been used in the past (Zhou et al., 2007). Should a tag perform particularly poorly as a PPTase substrate and require evolution to increase its utility, this strategy will have already generated a functional template for library creation which means that not every residue would need randomising, resulting in smaller library sizes and decreased screening effort.

Performing competition assays between native BpsA and an isolated peptide tag proved the most quantitative method for measuring the affinity between a PPTase and a tag by avoiding the variable of how well BpsA functions with an altered CP domain. Further confirming what the previous experiments of Section 4.3.2 showed, these competition assays generated IC_{50} values demonstrating that the E12 and P12 tags act as PPTase substrates, however they do so approximately 8 to 24-fold less efficiently than the competing BpsA construct containing the entire CP domain from which the tag is derived. This result suggested considerable involvement of the CP domain scaffold in PPTase binding. Competition between the B12 tag and WT BpsA also supported this conclusion, showing a 30-fold reduction in PPTase affinity upon isolation of the tag. The observation that PcpS performs better with BpsA(E12) than BpsA(E) is interesting because it suggests that PcpS recognises the foreign BpsA CP domain scaffold with greater affinity than the foreign EntF CP domain scaffold. This interpretation would imply that for some PPTases/CP domains, the CP scaffold may, in fact, be more important for PPTase binding than the tag region.

In 2014, the structure of Sfp bound to CoA and the CP-domain of the NRPS TycC3 was solved (Figure 4.9) (Tufar et al., 2014) and mutational analyses showed that the side-chain interactions essential for reactivity of these particular proteins lay within the tag region. Disruption of the detected interactions outside the tag region had little or no effect on interaction of the two proteins. The structure of Sfp shows a pocket into which the CP positions itself, and at the heart of which the conserved serine is presented to the catalytic core of the PPTase. The results described in that article, in combination with the results of this chapter, suggest that the involvement of the CP-domain scaffold in CP-domain - PPTase binding may be related to tertiary structure rather than interaction of specific residues. Of note is that the proteins involved in this published structure are unrelated and may not be representative of the interactions at play in other NRPS-PPTase systems.

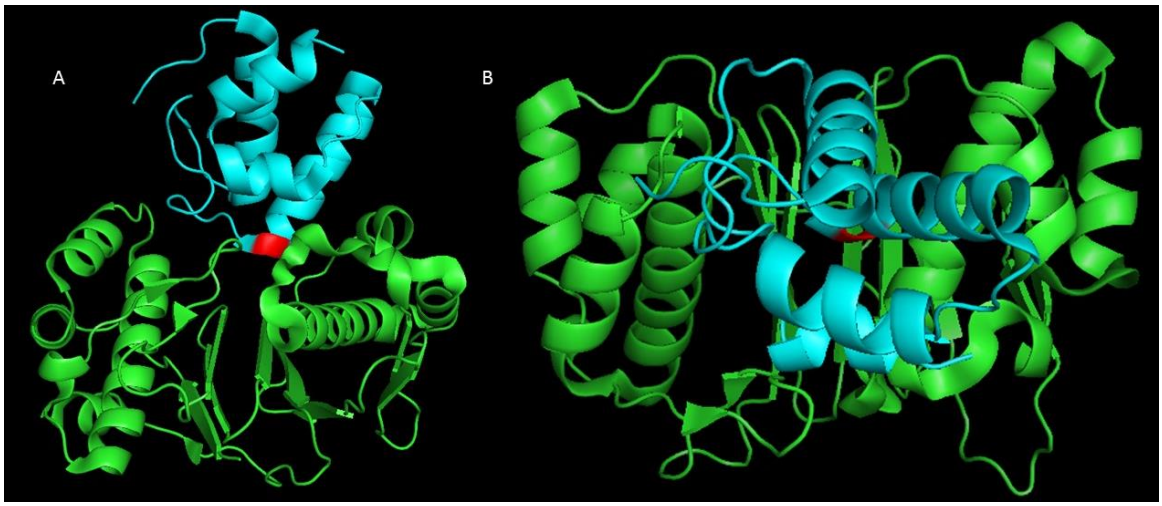

Figure 4.9: PyMOL view of Sfp (green) bound to TycC3_PCP S45A mutant (blue).

The red residue marks the location of the phosphopantetheinylated serine. A: Front-on view of interaction, B: Top-down view of interaction. A bowl-shaped pocket can be seen into which the carrier protein fits, conserved serine first. (Tufar et al., 2014).

In the experiments of this chapter, BpsA was used as a model for other NRPSs and it is important to remember that different NRPSs may have varying contributions made by their CP domain scaffold to PPTases. The results here suggest that PPTase recognition and BpsA function are two distinct characteristics, however in nature they must evolve side-by-side in a single NRPS system. The selective pressures dictating which characteristic is more important are likely to be dictated themselves by the function of the NRPS. For example, an antibiotic-producing NRPS may require rapid activation and function in response to certain stimuli and therefore have evolved both characteristics to high levels. A siderophore-producing NRPS on the other hand, may only require efficient function, and not be under such a strong pressure to be rapidly activated by a PPTase. As seen in Chapter 3, only very low levels of PPTase activity are required for siderophore production and subsequent growth in an iron-restricted environment. The outcome of this may be that NRPSs involved in strong, rapid responses may be a better place to look for tags than in NRPSs that do not have such a high requirement for PPTase activation. Future testing of a range of natural tag regions could investigate this speculation.

It remains unknown why the tagged MalE constructs displayed no inhibition of BpsA activity in competition assays that are otherwise identical to those containing the isolated peptide tags. This was an important experiment as it aimed to demonstrate the utility of the tags in a situation akin to how they could be applied in the future. It is predicted that the inability of these fusion proteins to act as PPTase substrates is an artefact of the MalE structure, which may be preventing proper presentation of the tag or formation of tag secondary structure. The highly hydrophobic nature of the candidate tags tested may have caused them to be internalised within the MalE structure. Moreover, the pH of these competition assays is tightly controlled to facilitate PPTase and BpsA function and this pH is more basic than is optimal for MalE stability. It is also possible that the structure of MalE surrounding the tag causes steric hindrance of PPTase binding. Due to time constraints, these experiments were not progressed further for this thesis. These tagged MalE vs. BpsA competition assays were intended to evaluate the function of the currently non-validated tags smaller than E12 and P12. Until this assay is optimised the properties of these tags are unknown.

Evolution of tag-swapped BpsA CP domains was employed in an effort to overcome the impairment of BpsA function imposed by tag swapping. *In vivo* screening revealed that a few mutations were sufficient to significantly increase the indigoidine production up to 4-fold in the case of BpsA(P12). Upon conception of this evolution work it was predicted that any conserved mutations would likely be involved in improving PPTase binding or CP domain function. Due to time constraints, it was not possible to fully investigate the mutations that arose in order to define their role. Isolation of two mutations, E941K and Q962H was used to show that neither mutation conferred a benefit to BpsA-EntD binding, nor do they improve WT BpsA function. Extensive future work using different combinations of these mutations and alternate BpsA CP domains in competition assays could narrow down the mechanisms leading to conservation of these mutations in the libraries. At the same as this work was being performed, Alistair Brown, a PhD student in the Ackerley lab was evolving the BpsA CP domain for activation by *Mycobacterium*

tuberculosis AcpS (a type I PPTase). Curiously, he also found high conservation of these same two mutations in his libraries. He has also shown that the mutations in isolation impaired BpsA activity with Mtb AcpS (unpublished results). The disagreement between *in vivo* and *in vitro* results in both Alistair's results and the work in this thesis hints that these mutations may be involved in a different mechanism from those previously proposed, for example tolerance to pH change between the cellular and *in vitro* environments.

It was surprising to find that E12mut1 did not possess enhanced EntD binding over E12 because the encoded mutation reached complete saturation in the library. Furthermore, this was a relatively large change to a small peptide; a shift from positively charged lysine to nonpolar isoleucine. Almost all functionally tested tags contain a nonpolar residue in this position, despite the fact that there is high polarity variance in this position in wild-type CP domains (Beer et al., 2014). As with the two conserved mutations outside the tag region, K978I appears to not confer an advantage on its own, suggesting it is part of a synergistic mechanism with other mutations, or provides an advantage to a different mechanism altogether. It has been reported that this residue in the A1 peptide undergoes little chemical shift when bound by AcpS (Zhou et al., 2008). Interestingly, the same lab have hypothesised that this residue may be partly responsible for the relatively specific interaction of AcpS and the ACP (Zhou et al., 2007). This glutamic acid residue also occurs in this position in other ACPs and therefore may be more generally involved in type I PPTase – ACP binding, rather than just in the case of AcpS.

Chapter 5: Future Directions & Concluding Remarks

The findings of this project have added to the collective knowledge of PPTases and their substrates via manipulation of their structures. Whilst the evolution work performed in Chapter 3 was not successful in generating the desired substrate-selective *entD* variant, it nevertheless provided valuable information regarding screen design and the sensitivity of BpsA as a biological reporter of PPTase activity. These findings may prove of critical importance in the design of future strategies that seek to achieve the same outcome, as well as other experiments that utilise BpsA activity as a measure of any parameter of PPTase activity (a common strategy in our laboratory). Building on the foundation of knowledge presented here, possible further approaches to generate substrate-selective PPTases are considered in the following section.

In Chapter 4, a simplistic method for the design and validation of two novel tags (E12 and P12) from natural CP domains was demonstrated, showing that an extensive library screening approach is not necessary to obtain effective PPTase-tag partnerships. It was also demonstrated that BpsA could provide an effective tag validation platform. The competition assays that were used to assess tag function revealed that the tertiary structure of the CP domain is an important factor modulating the interaction of a tag with a PPTase, and that removal from the CP domain context can yield a tag with impaired affinity for the PPTases tested here. Methods for further improvement of the E12 and P12 tags, as well as development of novel tags are discussed in Section 5.2.

5.1 PPTase evolution

In a number of ways this project aimed to build on the work of previous researchers and to use existing protocols to generate results, however, on this journey, as is often the case, more new questions arose than were answered. This was especially the case for the evolution of *entD* in which multiple flaws of the evolution strategy were uncovered, and when attempts were made to overcome these flaws, none could rescue the strategy and provide a compelling case for continuation of its use. Instead, two further strategies have been envisaged to possibly provide a reliable system for identifying substrate-selective PPTases from directed evolution libraries.

The first of these strategies is still based on an *in vivo* screen that utilises visualisation of colonies grown on agar plates with BpsA as a reporter of PPTase activity. However, this strategy differs from that in Chapter 3 by using *P. aeruginosa* as the first-tier screening strain and BpsA as a positive-selection tool, rather than a negative-selection tool (i.e. active indigoidine synthesis is a desirable outcome, rather than loss of synthesis, which may occur via mutations in BpsA rather than the target PPTase). My proposal is as follows: that the *entD* epPCR library be expressed from plasmid pSX (Owen, 2010) in *P. aeruginosa* PAO391 transformed with *bpsA(E)* on plasmid pSW196 (Baynham et al., 2006) (these plasmids selected because each can be expressed by both *E. coli* and *P. aeruginosa* which simplifies cloning and further screening). *P. aeruginosa* PAO391 is a modified strain of PAO1 in which the *pcpS* gene encoding the sole PPTase of *P. aeruginosa* is knocked out. Because *pcpS* is the sole PPTase, this strain must be complemented with another PPTase gene that can rescue PPTase-reliant primary metabolic processes (e.g. fatty acid synthesis). To fulfil this requirement, *E. coli acpS* has been chromosomally integrated. *acpS* was selected because it can replace the primary metabolic role of *pcpS,* but it does not activate the NRPS enzymes involved in synthesis of the *P. aeruginosa* siderophore pyoverdine, nor does it activate BpsA (meaning it does not interfere with screening). This library of transformants should then be plated onto indigoidine

production media containing a low concentration of iron chelator that is sufficient to promote synthesis of pyoverdine if the expressed EntD variant is still able to activate the CP domains of pyoverdine NRPS enzymes, but no so much that pyoverdine is essential to viability. Once colonies have formed, *bpsA(E)* and variant *entD* expression should be induced. Because pyoverdine is yellow-green and fluorescent it might be possible to use brief UV exposure to visually select colonies that have produced indigoidine but lack fluorescence due to pyoverdine loss. This strategy is essentially an inversion of the strategy used in Chapter 3 that will still screen for *entD* variants that have exclusive substrate-selectivity for the EntF CP domain, but this time there is a prevailing selection for, rather than against, the pigment that is shut down. Moreover, fluorescence screens are generally more sensitive than colorimetric screens (Mentel et al., 1996; Shimony et al., 2008; Uyama et al., 1988) and hence retention of indigoidine production coupled with loss of pyoverdine synthesis should indicate a strong preference for the EntF CP domain over the pyoverdine NRPS CP domains. However, considering the lessons learned during this project, I would not expect this alternative strategy to have a high chance of working efficiently; specifically, I would expect to see loss of pyoverdine fluorescence on plates, but indigoidine production resulting from *entD* variant activity with BpsA(P) in in a second-tier screen. Furthermore, the subsequent second-tier screen would still introduce the issue of false-positive reporting observed during this project (Section 3.2.2.2). Despite these doubts, work on this alternative strategy was begun, however the plasmid pSW196 proved difficult to work with owing to very low transformation efficiency into *E. coli* which meant cloning *bpsA(E)* into it could not be achieved. With more time and larger DNA quantities in the cloning stages, this construct could be generated and this strategy could be tested.

The second strategy envisioned is an *in vitro* expression approach that requires techniques and equipment not present in our laboratory, but rather is based on methods developed by the Levy laboratory at the Albert Einstein College of Medicine (Gianella et al., 2016). They have developed a system for *in vitro* compartmentalisation (IVC) of genes that catalyse ligation reactions, together

with the substrates for that reaction. One substrate is affixed to a nanoparticle bead, together with up to 4,000 affixed gene variants per bead. Each bead is physically compartmentalised by a water-in-oil emulsion. The second substrate of the target enzyme is free-floating in this emulsion and is fluorescently labelled. Then, a cell-free transcription-translation mixture such as the PURE system (a mixture of the recombinant proteins and precursors required for *E. coli* transcription and translation, (Shimizu et al., 2001)) is then incubated with the beads to synthesise the protein library from the affixed gene variants. The resulting enzyme variants can then perform their function (if active), ligating the fluorescent substrate to the bead-tethered substrate. The emulsion is then broken and the beads can be subjected to FACS treatment, and beads containing even a single active variant can be selected based on fluorescence. PCR is then employed to amplify the DNA associated with the selected beads and separation of gene variants followed by further rounds of mutagenesis can be performed. As the evolution scheme progresses through rounds of mutagenesis, the ratio of gene variants/bead is lowered such that individual improved variants are more directly selected via the FACS treatment. I envisage that this general system could be adapted for the evolution of substrate-selective PPTases. In this strategy, first the EntF CP domain would be tethered to beads together with a library of *entD* gene variants in the presence of fluorescently labelled CoA. Beads expressing variants active with the EntF CP domain would be selected based on strength of fluorescence (reporting retention of EntF reactivity), then the library DNA (which is also compartmentalised upon the bead) could be amplified and tethered to new beads that are this time coupled to the PvdD CP domain. Selection of this round would be for beads lacking fluorescence, meaning they had lost reactivity with the foreign substrate. This two-phase selection strategy ought to yield *entD* variants without cross-reactivity. Candidate variants identified via this method could then be investigated individually *in vitro* to confirm their properties and determine their kinetic parameters with each substrate. I believe the high sensitivity of this *in vitro* approach, coupled with its direct reporting of binding events would address the issues inherent to the *in vivo* approach used in Chapter 3. The possibility for

collaboration with the Levy laboratory has arisen and we hope to further discuss and develop this idea with their team and hopefully conduct the experiment described here. A further advantage of this type of strategy is that it will be applicable to other PPTase-CP domain pairs, and therefore the evolution of *pcpS* which was intended as the second aim of this project might be achievable using a slight adaptation (exchange of CP domain substrates) of this method. Alternatively, this strategy could be used to evolve a specific and non-overlapping EntD-tag and AcpS-tag pairing to improve on the Sfp-S6 and AcpS-A1 combination previously developed (Zhou et al., 2007). Starting with a type I and type II PPTase pairing may prove more amenable to elimination of cross-reactivity with each other's tag substrates.

An exciting and novel avenue of research could also be facilitated by this IVC method, that is, directed evolution of PPTases to gain activity with proteins unrelated to NRPS/PKS/FAS enzymes. More specifically, it may enable us to generate PPTase variants that can recognise, bind and pantetheinylate serine residues accessible on different cell-surface or extra-cellular proteins. This repurposing has not been tried before and could, in combination with different CoA analogues, lead to useful applications, for example the physical blockage of efflux pumps, or the *in vivo* capture or down-regulation of messenger proteins such as growth factors. We foresee applications such as this to be potentially relevant to many fields of research. Because this concept has not been tried before, it isn't known what structural motifs might indicate a good candidate substrate to evolve PPTase specificity towards. A good starting point might be to look for regions with "tag-like" sequences similar to those found in CP domains., i.e. an alpha helix comprising a hydrophobic stretch near a serine residue. Another approach could be to look for the sequence motif (D/H)**S**L and to evolve towards recognition of sites like this, although evidence does suggest that the hydrophobic stretch following the conserved serine is critically important (Tufar et al., 2014; Zhou et al., 2008) and therefore likely to be a better motif to search for. The results of Section 4.3.3 (where tagged maltose binding protein presumed to have an inwards facing serine could not compete against BpsA for PPTase activity) suggest that a target motif would have to be

presented on the surface of the protein of interest and in particular that the serine would have to be jutting outwards to be accessible by the PPTase.

The collective results presented in Chapter 3 suggest that the PPTase-CP domain interaction is a delicate one in which minor alterations can make a substantial impact. Although this work did not achieve its primary goal of creating highly active and selective PPTase variants, it did nevertheless cast some light on how sensitive a screen must be to differentiate between active and inactive PPTases. This realisation will inform work to develop and implement new screening strategies and these may one day allow the evolution of not only substrate-selective PPTase activity, but potentially also repurposed PPTase activity towards foreign proteins. Moreover, use of these novel strategies might permit Aim 2 (to generate a *pcpS* variant unable to modify the EntF CP domain but retaining a high level of activity with its native substrate) of this project to be pursued alongside further *entD* evolution to generate the non-overlapping pair of PPTases conceptualised at the beginning of this project.

5.2 Tag development

The efforts detailed in Chapter 4 to design and validate novel peptide tags as PPTase substrates achieved greater success than those described in Chapter 3. This tag development work was explorative and has laid the groundwork for the generation of more tags originating from other NRPS enzymes, as well as a method for testing how well they perform as PPTase substrates. It was interesting to find that the isolated tags performed substantially worse than complete CP domain-swapped BpsA constructs. While the competition assays that were implemented gave a good general idea of how well the tags perform as PPTase substrates, the assays do not conclusively describe how much worse they perform in comparison to an entire CP domain because the tags are held in the foreign context of BpsA, and the extent to which insertion of a novel tag sequence into BpsA may have impaired indigoidine synthesis is unknown. To further investigate a tag's reliance on the surrounding CP domain structure for PPTase binding, I would run competition assays using purified CP domains to inhibit the CP domain-swapped BpsA constructs, then compare this level of inhibition to the inhibition caused by isolated peptide tags. Comparison of the IC_{50} values for each of these assays would give a more precise report of the importance of the CP domain scaffold for PPTase binding.

Although it was unclear why the tag-fused MalE constructs showed no ability to compete with BpsA activity for the available CoA, some of the proposed reasons (serine inaccessibility, selection of fusion protein; Section 4.4) could readily be tested. The primary test to perform would be to fuse the tags to MalE in the opposite orientation. In the research described in Section 4.3.4, the primers used to make these fusion constructs positioned the tag on the C terminus of MalE with the conserved serine separated from the native MalE sequence only by a Gly-Gly linker and the first two residues of the tag. If it is indeed steric hindrance that was responsible for the lack of inhibition, then inverting the tag to present the conserved serine at the non-MalE end of the tag

may overcome this. The published structure of Sfp with TycC3_PCP bound (Tufar et al., 2014) (see also Figure 4.9) suggests that this may be important because the conserved serine is positioned well within the binding pocket of Sfp. This tag-inversion modification would be simple to achieve using the existing pET28a(+)::*malE* construct from this study, simply requiring new primers encoding the inverted tags. Another approach to testing foreign protein-tag fusion as a model for protein labelling would be fuse the tag to a different protein as a contingency plan in case there is a structural artefact of maltose-binding protein that causes misfolding of the tag. When selecting this protein, it might be valuable to use one that is expressed on the cell surface so that if it is indeed a substrate for a PPTase, it could be further tested as a proof-of-concept target for site specific labelling with fluorescent CoA in a whole-cell context.

The IVC method for screening peptide libraries (Gianella et al., 2016) outlined in Section 5.1 might also be adaptable for directed evolution of tags, although with a few modifications. In this case, because the library to be screened does not encode an enzyme that performs a ligation reaction, this library must be fused to a fluorescent protein, such as GFP, and its functional translated products captured by the beads. To fulfil this, CoA could be tethered to the beads via its thiol group (as performed by (Wong et al., 2008)) with a relevant PPTase free-floating in the emulsion. Transcription and translation of the tag library would be followed by PPTase ligation of tag variants to the tethered CoA when possible, and the fluorescence of GFP retained upon that bead; selectable by FACS. This screening method allows interrogation of libraries containing as many as 10^{12} variants, making complete randomisation of nearly all the tag residues either side of the conserved serine a feasible approach. This method would be a significant divergence from the *in vivo* screen used in Chapter 4 which was presented as a rapid and simple approach, and would only be usable by those with access to the elaborate equipment required. Despite this, the sheer screening power the IVC system provides would justify its use to generate tag-PPTase pairs that could then be used by others.

If future endeavours prove successful in generating substrate-selective PPTases, it would be both interesting and useful to evolve a pair of non-overlapping PPTases for use with the E12 and P12 tags from this project, or novel tags which might arise from further directed evolution in the future. Such an experiment might successfully unify the two main aims of this project to yield a single, applicable biotechnological tool.

In conclusion, the tag-related work performed in this chapter added to the current knowledge of these PPTase substrates through proof-of-principle experiments demonstrating a novel method for their design and validation. These developments alone are useful and it is anticipated that they may one day allow the realisation of the full value of the work of this thesis with the advent of substrate selective PPTases to be paired with tags generated by this method.

Bibliography

Asghar, A. H., Shastri, S., Dave, E., Wowk, I., Agnoli, K., Cook, A. M., & Thomas, M. S. (2011). The pobA gene of Burkholderia cenocepacia encodes a group I Sfp-type phosphopantetheinyltransferase required for biosynthesis of the siderophores ornibactin and pyochelin. *Microbiology, 157*(2), 349-361.

Barb, A. W., Cort, J. R., Seetharaman, J., Lew, S., Lee, H. W., Acton, T., . . . Montelione, G. T. (2011). Structures of domains I and IV from YbbR are representative of a widely distributed protein family. *Protein Science, 20*(2), 396-405.

Baynham, P. J., Ramsey, D. M., Gvozdyev, B. V., Cordonnier, E. M., & Wozniak, D. J. (2006). The Pseudomonas aeruginosa ribbon-helix-helix DNA-binding protein AlgZ (AmrZ) controls twitching motility and biogenesis of type IV pili. *Journal of bacteriology, 188*(1), 132-140.

Beer, R., Herbst, K., Ignatiadis, N., Kats, I., Adlung, L., Meyer, H., . . . Kurzawa, N. (2014). Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Molecular BioSystems, 10*(7), 1709-1718.

Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P., & Burkart, M. D. (2014). The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Natural product reports, 31*(1), 61-108.

Belshaw, P. J., Walsh, C. T., & Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science, 284*(5413), 486-489.

Burkart, M. D., Kosa, N. M., & Haushalter, R. W. (2015). Reversible Chemoenzymatic Labelling of Native and Fusion Carrier Motifs: US Patent 20,150,253,335.

Callegari, A., Luin, S., Marchetti, L., Duci, A., Cattaneo, A., & Beltram, F. (2012). Single particle tracking of acyl carrier protein (ACP)-tagged TrkA receptors in PC12nnr5 cells. *Journal of neuroscience methods, 204*(1), 82-86.

Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C., & Guilhot, C. (2006). The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria. *Proceedings of the National Academy of Sciences, 103*(22), 8511-8516.

Chen, D., Wu, R., Bryan, T. L., & Dunaway-Mariano, D. (2009). In Vitro Kinetic Analysis of Substrate Specificity in Enterobactin Biosynthetic Lower Pathway Enzymes Provides Insight into the Biochemical Function of the Hot Dog-Fold Thioesterase EntH†. *Biochemistry, 48*(3), 511-513.

Clarke, K. M., Mercer, A. C., La Clair, J. J., & Burkart, M. D. (2005). In vivo reporter labeling of proteins via metabolic delivery of coenzyme A analogues. *Journal of the American Chemical Society, 127*(32), 11234-11235.

Elhussein, S. A., Miernyk, J. A., & Ohlrogge, J. B. (1988). Plant holo-(acyl carrier protein) synthase. *Biochem. J, 252*, 39-45.

Elovson, J., & Vagelos, P. R. (1968). Acyl carrier protein X. Acyl carrier protein synthetase. *Journal of Biological Chemistry, 243*(13), 3603-3611.

Finking, R., & Marahiel, M. A. (2004). Biosynthesis of nonribosomal peptides 1. *Annu. Rev. Microbiol., 58*, 453-488.

Finking, R., Solsbacher, J., Konz, D., Schobert, M., Schäfer, A., Jahn, D., & Marahiel, M. A. (2002). Characterization of a new type of phosphopantetheinyl transferase for

fatty acid and siderophore synthesis in Pseudomonas aeruginosa. *Journal of Biological Chemistry, 277*(52), 50293-50302.

Gehring, A. M., Bradley, K. A., & Walsh, C. T. (1997). Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2, 3-dihydroxybenzoate. *Biochemistry, 36*(28), 8495-8503.

Gehring, A. M., Lambalot, R. H., Vogel, K. W., Drueckhammer, D. G., & Walsh, C. T. (1997). Ability of Streptomyces spp. aryl carrier proteins and coenzyme A analogs to serve as substrates in vitro for E. coli holo-ACP synthase. *Chemistry & biology, 4*(1), 17-24.

Gianella, P., Snapp, E. L., & Levy, M. (2016). An in vitro compartmentalization-based method for the selection of bond-forming enzymes from large libraries. *Biotechnology and bioengineering*.

Gillam, E. M., Copp, J. N., & Ackerley, D. F. (2014). Directed evolution library creation. *Methods in Molecular Biology*.

Green, M. R., & Sambrook, J. (2012). *Molecular cloning : a laboratory manual*. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Grossman, T. H., Tuckman, M., Ellestad, S., & Osburne, M. (1993). Isolation and characterization of Bacillus subtilis genes involved in siderophore biosynthesis: relationship between B. subtilis sfpo and Escherichia coli entD genes. *Journal of bacteriology, 175*(19), 6203-6211.

Johansson, P., Mulinacci, B., Koestler, C., Vollrath, R., Oesterhelt, D., & Grininger, M. (2009). Multimeric options for the auto-activation of the Saccharomyces cerevisiae FAS type I megasynthase. *Structure, 17*(8), 1063-1074.

Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C., & Barr, P. J. (1998). Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proceedings of the National Academy of Sciences, 95*(2), 505-509.

Kolb, H. C., Finn, M., & Sharpless, K. B. (2001). Click chemistry: diverse chemical function from a few good reactions. *Angewandte Chemie International Edition, 40*(11), 2004-2021.

Kosa, N. M., Pham, K. M., & Burkart, M. D. (2014). Chemoenzymatic exchange of phosphopantetheine on protein and peptide. *Chemical Science, 5*(3), 1179-1186.

Ku, J., Mirmira, R. G., Liu, L., & Santi, D. V. (1997). Expression of a functional non-ribosomal peptide synthetase module in Escherichia coli by coexpression with a phosphopantetheinyl transferase. *Chemistry & biology, 4*(3), 203-207.

La Clair, J. J., Foley, T. L., Schegg, T. R., Regan, C. M., & Burkart, M. D. (2004). Manipulation of carrier proteins in antibiotic biosynthesis. *Chemistry & biology, 11*(2), 195-201.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature, 227*(5259), 680-685.

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., . . . Walsh, C. T. (1996). A new enzyme superfamily—the phosphopantetheinyl transferases. *Chemistry & biology, 3*(11), 923-936.

Linne, U., Doekel, S., & Marahiel, M. A. (2001). Portability of epimerization domain and role of peptidyl carrier protein on epimerization activity in nonribosomal peptide synthetases. *Biochemistry, 40*(51), 15824-15834.

Louden, B. C., Haarmann, D., & Lynne, A. M. (2011). Use of blue agar CAS assay for siderophore detection. *Journal of microbiology & biology education: JMBE, 12*(1), 51.

Marchetti, L., De Nadai, T., Bonsignore, F., Calvello, M., Signore, G., Viegi, A., . . . Cattaneo, A. (2014). Site-specific labeling of neurotrophins and their receptors via short and versatile peptide tags.

Mentel, R., Matthes, E., Janta-Lipinski, M., & Wegner, U. (1996). Fluorescent focus reduction assay for the screening of antiadenoviral agents. *Journal of virological methods, 59*(1-2), 99-104.

Mercer, A. C., & Burkart, M. D. (2007). The ubiquitous carrier protein—a window to metabolite biosynthesis. *Natural product reports, 24*(4), 750-773.

Meyer, J.-M., Stintzi, A., De Vos, D., Cornelis, P., Tappe, R., Taraz, K., & Budzikiewicz, H. (1997). Use of siderophores to type pseudomonads: the three Pseudomonas aeruginosa pyoverdine systems. *Microbiology, 143*(1), 35-43.

Mootz, H. D., Finking, R., & Marahiel, M. A. (2001). 4′-Phosphopantetheine transfer in primary and secondary metabolism of Bacillus subtilis. *Journal of Biological Chemistry, 276*(40), 37289-37298.

Mosiewicz, K. A., Johnsson, K., & Lutolf, M. P. (2010). Phosphopantetheinyl transferase-catalyzed formation of bioactive hydrogels for tissue engineering. *Journal of the American Chemical Society, 132*(17), 5972-5974.

Muir, T. W., Sondhi, D., & Cole, P. A. (1998). Expressed protein ligation: a general method for protein engineering. *Proceedings of the National Academy of Sciences, 95*(12), 6705-6710.

Owen, J. (2010). Characterisation, manipulation and directed evolution of nonribosomal peptide synthetases. *Victoria University of Wellington, School of Biological Sciences.*

Owen, J., Calcott, M. J., Robins, K. J., & Ackerley, D. F. (2016). Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chemical Biology, 23*(11), 1395-1406.

Owen, J., Copp, J., & Ackerley, D. (2011). Rapid and flexible biochemical assays for evaluating 4prime-phosphopantetheinyl transferase activity. *Biochemical Journal, 436*(3), 709-717.

Owen, J., Robins, K., Parachin, N., & Ackerley, D. (2012). A functional screen for recovery of 4′-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environmental microbiology, 14*(5), 1198-1209.

Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., . . . Somers, W. S. (2000). Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure, 8*(8), 883-895.

Paul, S., Ishida, H., Nguyen, L. T., Liu, Z., & Vogel, H. J. (2017). Structural and dynamic characterization of a freestanding acyl carrier protein involved in the biosynthesis of cyclic lipopeptide antibiotics. *Protein Science*.

Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P., & Walsh, C. T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry, 37*(6), 1585-1595.

Raymond, K. N., Dertz, E. A., & Kim, S. S. (2003). Enterobactin: an archetype for microbial iron transport. *Proceedings of the National Academy of Sciences, 100*(7), 3584-3588.

Robins, K. (2016). Discovery, Characterisation and Engineering of Non-Ribosomal Peptide Synthetases and Phosphopantetheinyl Transferase Enzymes.

Schwyn, B., & Neilands, J. (1987). Universal chemical assay for the detection and determination of siderophores. *Analytical biochemistry, 160*(1), 47-56.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature biotechnology, 19*(8), 751-755.

Shimony, O., & Jaffe, C. L. (2008). Rapid fluorescent assay for screening drugs on Leishmania amastigotes. *Journal of microbiological methods, 75*(2), 196-200.

Sunbul, M., & Yin, J. (2009). Site specific protein labeling by enzymatic posttranslational modification. *Organic & biomolecular chemistry, 7*(17), 3361-3371.

Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., & Sugiyama, M. (2007). Cloning and characterization of a Streptomyces single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *Journal of Biological Chemistry, 282*(12), 9073-9081.

Tufar, P., Rahighi, S., Kraas, F. I., Kirchner, D. K., Löhr, F., Henrich, E., . . . Marahiel, M. A. (2014). Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chemistry & biology, 21*(4), 552-562.

Uyama, O., Okamura, N., Yanase, M., Narita, M., Kawabata, K., & Sugita, M. (1988). Quantitative evaluation of vascular permeability in the gerbil brain after transient ischemia using Evans blue fluorescence. *Journal of Cerebral Blood Flow & Metabolism, 8*(2), 282-284.

Vivero-Pol, L., George, N., Krumm, H., Johnsson, K., & Johnsson, N. (2005). Multicolor imaging of cell surface proteins. *Journal of the American Chemical Society, 127*(37), 12770-12771.

Walsh, C. T., Liu, J., Rusnak, F., & Sakaitani, M. (1990). Molecular studies on enzymes in chorismate metabolism and the enterobactin biosynthetic pathway. *Chemical Reviews, 90*(7), 1105-1129.

Wong, L. S., Thirlway, J., & Micklefield, J. (2008). Direct site-selective covalent protein immobilization catalyzed by a phosphopantetheinyl transferase. *Journal of the American Chemical Society, 130*(37), 12456-12464.

Yin, J., Liu, F., Li, X., & Walsh, C. T. (2004). Labeling proteins with small molecules by site-specific posttranslational modification. *Journal of the American Chemical Society, 126*(25), 7754-7755.

Yin, J., Straight, P. D., McLoughlin, S. M., Zhou, Z., Lin, A. J., Golan, D. E., . . . Walsh, C. T. (2005). Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proceedings of the National Academy of Sciences of the United States of America, 102*(44), 15815-15820.

Zhou, Z., Cironi, P., Lin, A. J., Xu, Y., Hrvatin, S., Golan, D. E., . . . Yin, J. (2007). Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS chemical biology, 2*(5), 337-346.

Zhou, Z., Koglin, A., Wang, Y., McMahon, A. P., & Walsh, C. T. (2008). An eight residue fragment of an acyl carrier protein suffices for post-translational introduction of

fluorescent pantetheinyl arms in protein modification in vitro and in vivo. *Journal of the American Chemical Society, 130*(30), 9925-9930.