# Discovery, Characterisation and Engineering of Non-Ribosomal Peptide Synthetases and Phosphopantetheinyl Transferase Enzymes

By

Katherine Jane Robins

A thesis submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Biotechnology

Victoria University of Wellington

(2016)

**Abstract**

Non-ribosomal peptide synthetases (NRPSs) are multi-modular biosynthetic enzymes that are responsible for the production of many bioactive secondary metabolites produced by microorganisms. They are activated by phosphopantetheinyl transferase (PPTase) enzymes, which attach an essential prosthetic group to a specific site within a "carrier protein" (CP) domain that is an integral part of each NRPS module. Of particular importance in this work is the NRPS BpsA, which produces a blue pigment called indigoidine; but only when BpsA has first been activated by a PPTase. BpsA can be used as a reporter for PPTase activity, to identify PPTases and/or measure their activity. Several CP-substituted BpsA variants were used, in order to study and identify PPTases which may recognise different CP domains. The first part of the research described in this thesis examined the features of foreign CP interactions within BpsA that made these functional substitutions possible. Two key residues, the +4 and +24 positions relative to an invariant serine, were found to be highly important; with appropriate substitutions at these positions yielding active CP-substituted variants.

Wild type BpsA and the CP-substituted variants were then used as the basis of a screen to discover new PPTase genes, and associated natural product biosynthetic genes, from metagenomic libraries. The vast majority of bacteria that produce bioactive secondary metabolites are unable to be cultured under laboratory conditions; screening metagenomic libraries is a way to access this untapped biodiversity in order to discover new natural products. Two environmental DNA libraries were screened, and PPTase genes were identified via their ability to activate BpsA, giving rise to blue colonies in high throughput agar plate screens. This screen proved to be a powerful enrichment strategy with almost half of the novel 21 PPTase genes recovered also linked to biosynthetic gene clusters. Using the evolved CP-substituted BpsA variants (and thereby altering the PPTase recognition site) enabled a wider variety of hits to be found. This led to the hypothesis that some of the PPTases discovered via this screening method would have non-overlapping substrate specificities, a beneficial property for certain PPTase applications.

The 21 PPTase genes discovered via metagenomic screening were characterised further, using a series of assays involving BpsA to measure their activity. As is common for PPTase enzymes, there were difficulties in obtaining enough soluble protein via purification to perform a detailed analysis of each. Those that were able to be purified had much lower activity than other previously characterised PPTases, and were also not as specific for their CP substrates as they had first appeared to be. Due to these low activity levels, several other previously characterised PPTases were also studied further using the BpsA methods. All PPTases showed a relatively broad activity across a range of CP substrates.

The desire to obtain PPTases with more specific substrate specificities led to the development of a directed evolution screen to alter PPTase CP specificity. In a proof-of-principle study the *E. coli* PPTase EntD was evolved to lose activity with the BpsA CP while retaining activity with its native CP. This screen, the first of its kind to evolve PPTases for greater CP substrate specificity, was successful in recovering several improved variants. These variants had either completely abolished or vastly decreased activity for the WT BpsA CP while retaining the ability to activate the native (EntF) CP domain. The general strategy developed here can be applied to the evolution of other PPTases and CP substrates.

**Acknowledgments**

Firstly I need to acknowledge my supervisor, Dr David Ackerley, I couldn't have done this WITHout you. I cannot thank you enough for your continued support, guidance and encouragement throughout the research and writing of this thesis. Our lab is pretty much the best place to work, mainly due to your knack for filling it with such awesome people (and being an all-round excellent supervisor).

I would like to thank Victoria University and the Royal Society of New Zealand Marsden Fund for funding my studies. Thank you to the University also, for providing such a supportive learning environment, and all the technicians in SBS who keep the lab (and the incubators) running.

To everyone in the Ackerley lab past and present (and Nitros and Pyos) – a huge thank you for the support and friendship you have provided over the years. Thanks to everyone who has shared the wonderful home that is "Becky's office" with me- Elsie, Becky, Mark, Richard, Alistair, Michelle, Rory, Luke, Abby and Matt. Thanks to all my gym buddies (especially Becky!), drinking buddies and to anyone who has helped me with technology- dragging me out of the dark ages of doing sequence alignments in word. Thanks to Elsie for helping me wrestle with thesis-formatting (and teaching me about coffee), Michelle for being a referencing-wiz, and Alistair for reading my methods and always being there to chat about BpsA. Thanks also to past members of the lab- Janine and Gareth who helped me out when I was first starting out, and Jeremy to (now present lab member again) whose work inspired much of this project.

A special thank you goes to Trudi at the gym for giving me a much needed break from my studies and keeping my step game fresh.

Thanks to all my family and friends who have supported me through this journey, for trying to show an interest in what I do (enzymes?) and always being there whenever I needed a break from science. Lastly, thank you to my Mum and Dad for their love and support.

## Table of Contents

## List of Figures

## List of Tables

## List of Abbreviations

| | |
|---|---|
| A590 | absorbance at 590 nm |
| A domain | adenylation domain |
| ACP | acyl carrier protein |
| APS | ammonium persulfate |
| ArCP | aryl carrier protein |
| ATP | adenosine triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| BSA | Bovine Serum Albumin |
| bp | basepair |
| C domain | condensation domain |
| CoA | Coenzyme A |
| CP | carrier protein |
| ddH2O | double distilled water |
| DMSO | dimethylsulfoxide |
| E domain | epimerisation domain |
| eDNA | environmental DNA |
| EDTA | ethylenediaminetetraacetic acid |
| FAS | fatty acid synthase |
| GFP | green fluorescent protein |
| IPTG | isopropyl β-D-1-thiogalactoside |
| kb | kilobase |
| $k_{cat}$ | kinetic constant describing theoretical maximum rate of catalysis for a specific enzyme with a specific substrate, under the conditions tested. |
| $K_m$ | kinetic constant describing the concentration of substrate at which rate of catalysis of a specific enzyme is exactly half of the $V_{max}$, under the conditions tested. |
| KO | Knockout |
| LB | Luria broth |
| MDa | megadalton |
| NRPS | non-ribosomal peptide synthetase |
| $OD_{590}$ | optical density at 590 nm |
| $OD_{600}$ | optical density at 600 nm |
| $OD_{800}$ | optical density at 800 nm |

| | |
|---|---|
| Ox domain | oxidation domain |
| PCP | peptidyl carrier protein |
| PCR | polymerase chain reaction |
| PKS | polyketide synthase |
| Ppant | phosphopantetheine |
| PPTase | phosphopantetheine transferase |
| rpm | revolutions per minute |
| SD | standard deviation |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SEM | standard error of the mean |
| TE domain | thioesterase domain |
| TEMED | Tetramethylethylenediamine |
| TFBI | Transforming buffer I |
| TFBII | Transforming buffer II |
| Tris | Tris(hydroxymethyl)aminomethane |
| UV | ultra-violet |
| v/v | volume per volume |
| w/v | weight per volume |

## List of commonly used terms

| | |
|---|---|
| *bpsA* | Blue pigment synthetase A (gene notation) |
| BpsA | Blue pigment synthetase A (protein notation) |
| EcoBlue1 | BL21 Δ*entD* + pCDFDuet1::*bpsA* (wild-type) |
| EcoBlueP | BL21 Δ*entD* + pCDFDuet1::*bpsA* (*pvdD* carrier protein domain substitution) |
| EcoBlueE | BL21 Δ*entD* + pCDFDuet1::*bpsA* (e*ntF* carrier protein domain substitution) |
| *entD* | *E. coli* PPTase from the enterobactin biosynthesis pathway (gene notation) |
| EntD | *E. coli* PPTase from the enterobactin biosynthesis pathway (protein notation) |
| EntF | Enzyme in the enterobactin biosynthesis pathway |
| ybbr | 11 amino acid protein tag, substrate for PPTases |

# Chapter 1: Introduction

## 1.1 Nonribosomal peptide synthetases

Non-ribosomal peptide synthetases (NRPSs) are large complex enzymes, present in many bacteria and fungi, that are responsible for the synthesis of a range diverse peptides that often possess bioactive properties (Finking and Marahiel, 2004). For example, this diverse group of products includes antibiotics such as penicillin and vancomycin, the immunosuppressant cyclosporine, and the anti-cancer drugs epothilone and bleomycin. Synthesis of NRPS enzymes is a widespread feature of many microorganisms (Schwarzer et al., 2003) and it is thought that they provide a competitive advantage over rival microorganisms (Finking and Marahiel, 2004). The NRPS structure is modular, whereby each module of the enzyme is responsible for the incorporation of an amino acid monomer into the final peptide structure. The enzymes themselves are usually large and energy-intensive for the microorganisms to produce, for example the gene for cyclosporine production is 45.8 kb in size, coding for a 1.6 MDa sized protein, all to produce a product which consists of only 11 amino acid monomers (Weber et al., 1994). The fact that these enzymes represent such a huge investment of energy and resources reflects that their peptide products are often highly bioactive, i.e. provide essential offensive or defensive capabilities under certain conditions (Fischbach and Walsh, 2006).

The vast diversity of possible NRPS products stems from two key differences between ribosomal and nonribosomal synthesis; the use by NRPS enzymes of a wide variety of amino acids (beyond the usual proteinogenic 20), and the potential for these to be incorporated into the final peptide product in an unusual manner. NRPSs can incorporate over 500 unique monomers into their peptide products, with the sizes of most of these peptides ranging from 2-23 monomers (Caboche et al., 2010). Tailoring domains that are present in addition to the standard domains required for peptide product formation also help to increase the chemical diversity of the products. For example many NRPSs contain epimerisation domains which alter the stereochemistry of the corresponding amino acids from their L to D forms. Penicillin is an example of a peptide product which contains an unusual amino acid, L-α-aminoadipic acid, as well as an amino acid with altered stereochemistry, D-valine, and an unusual intramolecular cyclisation between the sidechains of the D-valine and the third amino acid monomer, L-cysteine (Felnagle et al., 2008). As a means of providing yet further diversity, NRPS enzymes can also exist as hybrids with polyketide synthetases (PKSs). PKSs are a class of enzymes that have a similar modular structure to NRPSs, but use acetate

and propionate as their monomeric building blocks as opposed to amino acids. Hybrid NRPS/PKS contain combinations of discrete NRPS and PKS modules, allowing for even greater diversity in products. For example the two anti-cancer drugs mentioned earlier (epothilone and bleomycin) are both examples of hybrid products (Schwarzer and Marahiel, 2001). Ultimately, NRPS synthesised peptides have the potential to be much more structurally diverse than traditional peptides (Finking and Marahiel, 2004).

### 1.1.1 NRPS structure and function

Structurally NRPSs are large enzymes that consist of a series of smaller enzymatic subunits known as modules, with each module responsible for the incorporation of one amino acid into the final peptide product (Schwarzer et al., 2003). Each module is in turn made up of several core catalytic domains that perform the key reactions necessary for the formation of the peptide product. These core domains are the Adenylation (A) domain, responsible for the selection and activation of a specific amino acid substrate; the peptidyl carrier protein (PCP) domain, the site of attachment responsible for transfer of the growing peptide between modules; and the condensation (C) domain, responsible for the condensation of adjacent amino acids via peptide bond formation. The final domain in the terminal module of nearly every NRPS enzyme is the thioesterase or TE domain, which is responsible for catalysing the release of the final peptide product from the larger NRPS complex (Figure 1.1). In addition to these core domains, many NRPSs contain additional tailoring domains which can result in further modifications to the product. One example is the epimerisation (E) domain which, as mentioned above, catalyses the epimerisation of bound L-amino acids into D-amino acids. Another tailoring domain is the oxidation (Ox) domain which can catalyse oxidation reactions using a bound FMN cofactor (Schwarzer et al., 2003).

When an NRPS enzyme is first synthesised, it is in an inactive (apo) form and needs to be activated into the active (holo) form in order to function. To be active, the PCP domain of the NRPS must have a phosphopantetheine (Ppant) arm attached, which acts as the point of attachment for the growing peptide product. This attachment reaction is catalysed by a phosphopantetheinyl transferase (PPTase) enzyme, which catalyses the transfer of the Ppant arm, derived from CoA, onto an invariant serine residue within the PCP (Marahiel et al., 1997). The intermediate peptide products are connected via an energy rich thioester linkage at the tip of the Ppant arm, which facilitates hydrolysis during peptide synthesis. This arm is a 20 Å long moiety, which serves as a flexible linker, allowing the transfer of the growing peptide chain along the different catalytic domains of the NRPS enzyme (Finking and Marahiel 2004).

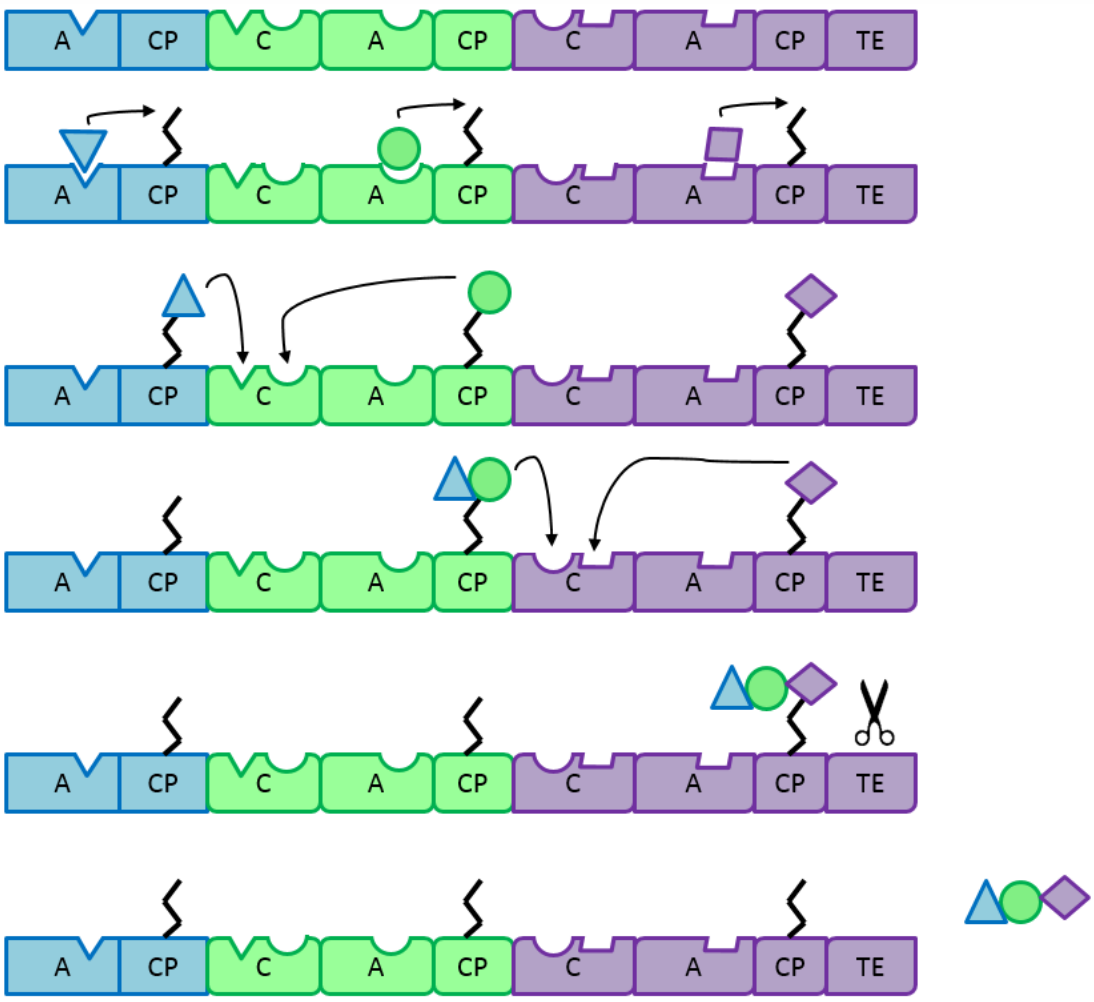



**Figure 1.1 Basic NRPS structure and function.** NRPS enzymes consist of several domains (represented in blue, green, and purple) which each govern incorporation of one amino acid ( ▲●◆ ) into the final product. First each CP must be activated by a PPTase, which adds the 4'-phosphopantetheine arm. The A domains recognise each amino acid and catalyse the attachment to the 4'-phosphopantetheine arm. C domains then catalyse peptide bond formation between adjacent amino acids, and the growing peptide product is passed along to the next module. Finally, the TE domain releases the final product. Adapted from (Calcott, 2014).

#### 1.1.1.1 Carrier protein domains

The domain of most relevance to the work described in this thesis is the PCP domain (sometimes referred to as just a PCP, or as a Thiolation or T domain), which will be described in more detail here. The PCP domain in NRPSs serves as the specific site of attachment for the Ppant arm (covalently attached to an invariant serine residue by a PPTase), which in turn serves as the point of attachment for the intermediate products during the synthesis process, transferring the product between the different catalytic domains. After a PCP has been activated by a PPTase it transforms from its inactive, or apo-enzyme state to the active, or holo-enzyme state. PCPs are one form of carrier protein (CP), so named because they specifically carry amino acid or peptidyl

substrates. The wider CP family also comprises aryl carrier proteins (ArCPs), which carry aryl acid substrates; and acyl carrier proteins (ACPs), which carry acetate, propionate or malonate monomers. ArCPs are a subset of the PCP domains, which are also found in NRPSs and are usually involved in siderophore biosynthesis (Crosa and Walsh, 2002). Siderophores are another common NRPS product, molecules for iron scavenging that are often key virulence factors in infection. In contrast, ACPs are found in PKSs as well as fatty acid synthases (FAS). CP domains are further defined by whether they exist as part of a larger polypeptide (type I) or are themselves standalone proteins (type II). Most of the PCPs in NRPSs appear to be type I and are part of the larger NRPS unit, but several type II PCPs have also been found (Du and Shen, 1999; Lohman et al., 2014)

CP domains are typically small protein domains (75-90 bp) that contain a small, identifying region of highly conserved sequence, consisting of the invariant serine residue that is the site of Ppant attachment and the amino acids immediately surrounding it, GX(D/H)S(L/I)(D/K), (Mercer and Burkart, 2007; Mofid et al., 2002). The first PCP to have its structure solved, using NMR, was TycC3 from the tyrocidine synthetase NRPS in *Bacillus* brevis (Weber et al., 2000). This structure appeared as a four helix bundle, a very similar structure to that of previously solved ACP domains, with three longer helices that are parallel to each other and a shorter forth helix (Crosby and Crump, 2012). All CPs to date have this conserved 4 helix structure (Finking and Marahiel, 2004). The helices are joined by loop regions, and the invariant serine is on the loop between helix I and II, near the start of helix II. In particular helix II of the CP has been found to be important for interactions between PPTases and CP domains (Mofid et al., 2002; Parris et al., 2000; Tufar et al., 2014). One of the main differences between ACP and PCP domains is the overall charge of the protein, with PCPs being neutral and ACPs negatively charged (Weber et al., 2000). The CP domains will be discussed further in Chapter 3.

### 1.1.2 New natural products via NRPS engineering or discovery

The highly bioactive nature of many of the NRPS peptides means there is great interest in discovering new natural products, through either the engineering or discovery of NRPS genes. The modular structure of NRPSs suggests it should be possible to perform a straightforward manipulation of the order and number of modules to create novel peptides (Mootz et al., 2002a). This strategy, while promising, usually results in vastly decreased yields (Calcott and Ackerley, 2014; Sieber and Marahiel, 2005). A greater understanding of the interactions between different domains and modules is required in order to improve the outcome of these manipulation experiments. In particular, the interactions between CP domains and downstream TE domains are a primary focus of the research described in Chapter 3. Another approach is the discovery of

entirely new NRPS genes that have not been found before, this is discussed further in section 1.3.2 and is the subject of the work presented in Chapter 4.

## 1.2 Phosphopantetheinyl transferase enzymes

The PPTase superfamily of enzymes was first described by Lambalot et al. (1996), and more recently has been the subject of an extensive review (Beld et al., 2013). PPTases catalyse the transfer of the 4'-phosphopantetheinyl arm of CoA onto a CP domain, thereby activating the CP domain (Figure 1.2). These CP domains may belong to NRPS, PKS, or FAS enzymes, and all require this activation by a PPTase in order to function. PPTases are essential enzymes found in all living cells, from microorganisms to mammals, and are divided into different subtypes (types I-III) defined by differing sequences, physical characteristics and activities (Copp and Neilan, 2006; Joshi et al., 2003). All organisms contain at least one PPTase, necessary to activate the essential FAS of primary metabolism, and often contain at least one additional PPTase, specialised for activation of the NRPS and PKS enzymes of secondary metabolism. PPTases have low sequence homology, though they do contain several conserved motifs (Lambalot et al., 1996; Sanchez et al., 2001; Walsh et al., 1997). Recently the signature sequence of bacterial PPTases has been extended and redefined as (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 x is any amino acid, h is any hydrophobic amino acid and n is 42-48 for type I PPTases and 38-41 for type II PPTases. There is also another less well conserved region within this middle stretch, (L/I/V/M/F)$(x)_3$(V/I/Y/F)h$(x)_3$E$(x)_3$h, which contains a highly conserved glutamine residue (Asghar et al., 2011).

PPTases themselves are used as tools in many biotechnological applications, due to their ability to transfer a host of molecules from CoA analogues onto CP domain substrates (Figure 1.2).

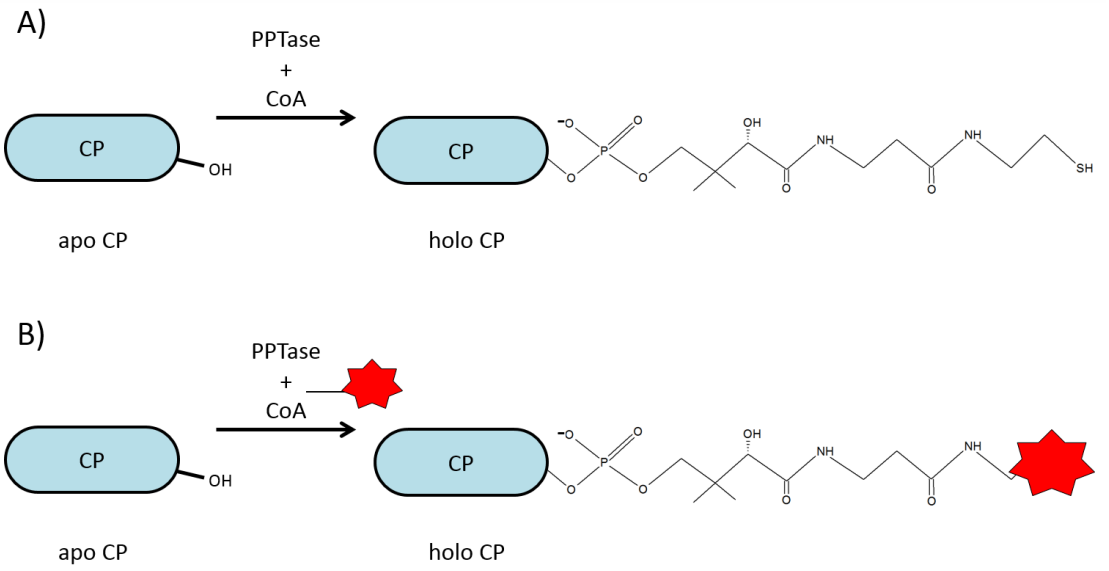



**Figure 1.2 PPTases function.** A) PPTases catalyse the reaction whereby the 4'-phosphopantetheine arm of CoA is transferred onto the sidechain of an invariant serine residue within a CP. B) Many PPTases are also capable of utilising modified CoA analogues that have various chemical moieties covalently in place of the free sulfhydryl group on the phosphopantetheine arm.

### 1.2.1 PPTase subtypes

PPTase enzymes are classified into three subtypes, depending on their sequence, structure and activity. Type I PPTases are typified by the *E. coli* PPTase AcpS, and are usually involved in pathways of primary metabolism, and as a result are often essential. Type II PPTases are twice the size of type I, are typified by the *Bacillus subtilis* PPTase Sfp, and are usually involved in pathways of secondary metabolism. Type III PPTases are the least well studied of the subtypes, and are generally only found in fungi. These PPTases are integrally associated with their larger target enzymes, for instance in *Saccharomyces cerevisiae* a type III PPTase domain was identified as part of the FAS enzyme complex (Fichtlscherer et al., 2000). These integrated type III PPTases have a similar sequence to type I PPTases, and will not be discussed further as they are outside the scope of this thesis.

#### 1.2.1.1 Type I PPTases

Type I, or AcpS-like PPTases, are found in almost all organisms, and activate the ACP domains of FAS and PKS enzymes. AcpS from *E. coli* was one of the first PPTase enzymes to be discovered (Elovson and Vagelos, 1968) and is one of the best characterised (Gehring et al., 1997; Lambalot and Walsh, 1995). Type I PPTases are the smallest enzymes in the family at approximately 120

amino acids, and form a homotrimer (Parris et al., 2000). They are usually more specific for ACP domains, and generally do not exhibit substantial levels of activity with PCP domains.

#### 1.2.1.2 Type II PPTases

Type II, or Sfp-like PPTases primarily activate the PCP or ArCP domains found in NRPSs. These PPTases are around twice the size (at approximately 240 amino acids) of type I enzymes, and have a pseudo-homodimer structure (a single peptide chain that folds into two symmetrical units resembling two separate domains which have formed a dimer). In addition to Sfp, the first type II PPTases described were Gsp from *Bacillus brevis* and EntD from *E.* coli (Lambalot et al., 1996). Sfp, whose native substrates are the NRPS responsible for surfactin production, is in particular known for its broad substrate specificity, and has been shown to activate many PCP domains from foreign NRPSs as well as ACP domains, although it has a much lower efficiency with ACP domains (Lambalot et al., 1996; Nakano et al., 1992). Due to the broad specificity of Sfp, it has been widely used for the heterologous expression of NRPS and PKS enzymes in *E. coli* (Kealey et al., 1998; Pfeifer et al., 2001). While *E. coli* is very amenable to genetic manipulation, its own PPTases are relatively specific for endogenous substrates, hence a need for broadly specific PPTases like Sfp to activate foreign enzymes. The *E. coli* type II PPTase, EntD is a good example of a class II enzyme that is more specific for its cognate substrates, the NRPSs responsible for production of the siderophore enterobactin. EntD shows little crossover in terms of substrates with AcpS (Flugel et al., 2000; Lambalot et al., 1996), which are involved in FAS and primary metabolism, and also relatively poor activation of other bacterial NRPS enzymes when these are expressed in *E. coli* (Caffrey et al., 1991; Pfeifer et al., 1995; Roberts et al., 1993).

An unusual type II PPTase that is relevant to this thesis is PcpS from *Pseudomonas aeruginosa*, which is the only PPTase present in *P. aeruginosa*, and as such is responsible for modifying both ACP and PCP domains present in this organism. Though PcpS is clearly a type II PPTase, it shows a higher activity with its endogenous ACP domains compared to its endogenous PCP domains, and hence substantially higher activity with ACP domains than is generally seen with other type II PPTases. This may emphasise the essential role of PcpS in primary metabolism (Barekzi et al., 2004). Because of this unusual property PcpS is sometimes referred to as a subtype of type the type II class (Finking et al., 2002).

### 1.2.2 PPTase structure

A number of PPTases have had their structures described. Many of these structures have been solved in the presence of various ligands, e.g. CoA, their native CP substrate, or the essential $Mg^{2+}$

ion. Only four type II PPTases have had their crystal structure solved (with two of these being very recent (Jung et al., 2014; Vickery et al., 2014)), the majority of solved structures have been for type I PPTases. Despite low sequence homology being shared among the PPTases, the basic structure is very similar, with each subunit of the type I PPTases and each half of the pseudohomodimeric type II PPTases having the same basic fold structure.

#### 1.2.2.1 Structure of type I PPTases

Type I PPTases are the most well structurally characterised PPTases to date. AcpS from *B. subtilis* and AcpS from *Streptococcus pneumoniae* were the first type I PPTases crystallised, and these structures showed that AcpS family members exist in a trimeric configuration (Chirgadze et al., 2000; Parris et al., 2000). Each subunit is about 15 kDa, with an α/β fold, and these come together to form a tightly packed trimer. At each interface between the subunits an active site is formed, resulting in three active sites per enzyme. Site directed mutagenesis was used to confirm the importance of this secondary structure, as when AcpS was unable to trimerise it displayed no activity. The structure of the *B. subtilis* AcpS in complex with CoA showed that the interactions between the 4'-phosphopantetheine arm and the protein are weak (with the Ppant resting in a shallow hydrophobic pocket on the protein's surface), and that the 4'-phosphopantetheine arm extends away from the base of the CoA molecule. Interactions between AcpS and ACP were observed between helix I of AcpS and helix II of ACP, these interactions being mostly hydrophilic in nature (Parris et al., 2000).

#### 1.2.2.2 Structure of type II PPTases

Sfp was the first PPTase from any family to have its crystal structure solved (Reuter et al., 1999). The structure of Sfp was found to be unique, with a novel α/β fold and a pseudohomodimeric structure, with the N terminal and C terminal halves of the protein separated by a short flexible polypeptide loop. Each half of the protein is very similar to one of the type I PPTase monomers. Because of this structure it is thought that type II PPTases may have evolved from an ancient duplication event (Mootz et al., 2001). The C terminal half of Sfp more closely resembles an AcpS monomer than does the N-terminal half, as it contains the three residues that coordinate the $Mg^{2+}$ ion (Asp107, Glu109 and Glu151 in Sfp). The C-terminus does however also contain an extra loop relative to AcpS, thought to confer stability to the pseudodimer (Beld et al., 2013). This crystal structure was solved in the presence of CoA, which bound in the pocket formed between the N-terminal and C-terminal halves. The pantetheine arm appeared to be quite flexible in the crystal structure, and did not appear to have any strong interactions with the Sfp enzyme. This was

thought to explain the earlier observation that Sfp had shown promiscuity towards a number of different CoA analogues that contained modified phosphopantetheine arms (Quadri et al., 1998). A recently solved structure of the Sfp/PCP complex showed the two proteins associated via a hydrophobic interaction between helix II of the PCP and the first and third helices in the C-terminal domain of Sfp. The binding of the PCP caused a shift in the C-terminal domain of Sfp, opening up a cavity between the two domains (Tufar et al., 2014).

The second type II PPTase to be crystallised, the human PPTase AASDHPPT (Bunkoczi et al., 2007) has some notable differences from other type II PPTases that are known, primarily that it is a lot larger (almost 100 residues longer), with a larger "tail" wrapping around the back of the PPTase. The CoA molecule and ACP domain were found to bind in a similar manner to Sfp, in an active site located between the N- and C-terminal domains.

More recently two groups have solved the crystal structure of the *Mycobacterium tuberculosis* PPTase, PptT (Jung et al., 2014; Vickery et al., 2014), with Vickery et al. also solving another mycobacterial type II PPTase, MuPPT from *M. ulcerans*. These mycobacterial PPTases were found to have similar folding and overall structure to Sfp and AASDHPPT but did display some notable differences in the active site, most notably in the binding of CoA. In both of the mycobacterial PPTases, the pantetheine arm bound to the PPTase in a deep hydrophobic pocket, forming a close bond with the PPTase. The adenine base of CoA is also found in a deep hydrophobic pocket whereas in both Sfp and AASDHPPT this pocket was a lot shallower. This may explain the observation that unlike many other PPTases, PptT is poorly able to accommodate modified CoA substrates (Leblanc et al., 2012).

### 1.2.3 Methods for assessing PPTase activity

PPTases are generally classified by their ability to activate different CP substrates. *In vivo* complementation studies can be performed for confirmation of a gene's identity as a PPTase (Borchert et al., 1994; Mootz et al., 2002b; Nakano et al., 1992). The ability of PPTases to activate various CP domains can also be tested *in vitro,* although traditional methods can be laborious and time consuming and generally are only able to be performed on a low throughput scale.

PPTases can be qualitatively assessed to determine whether they are capable of activating a given CP substrate, by using various assays to look at the amount of the CP that is phosphopantetheinylated within a set period of time. The PPTase and CP are expressed in the same cell, or in an in vitro mixture, and the CP is isolated so that the ratio of phosphopantetheinylated to unphosphopantetheinylated can be determined, either through a

radiometric assay, HPLC or by using mass spectroscopy (Copp et al., 2007; Lambalot and Walsh, 1997; Liu et al., 2005). Full kinetics are more time intensive and technically challenging. Many PPTases are also very difficult to express as soluble proteins. Due to the laborious methods for characterisation, only seven PPTases have been kinetically characterised by these means (Beld et al., 2013). These are *E. coli* AcpS, Sfp, PcpS, AASDHPPT (the human PPTase), Lys5, Svp and FdmW. A new approach to assess the kinetic parameters of PPTases was developed in our lab (Owen, 2010). PPTase activity is measured via their ability to activate BpsA (described in section 1.4), an NRPS which synthesises a coloured product, indigoidine. This method was used to assess three PPTases- Sfp, PcpS and PP1183 (Owen et al., 2011). This approach will be discussed further in Chapter 5.

## 1.3 Biotechnological applications of PPTases

The flexibility of the PPTase catalysed reaction - the transfer of the Ppant arm of CoA onto a CP domain - can be exploited for use in many different biotechnological applications. This utility largely stems from the promiscuity of PPTases towards different substrates, in particular their ability to accommodate a wide variety of CoA analogues. This property is thought to be explained by the mode of CoA binding seen in both Sfp and AcpS type PPTase crystal structures, where the phosphate groups of CoA provide the main interaction with the PPTase active site and the $Mg^{2+}$ cofactor, while the terminal thiol end of CoA extends out of the binding pocket in the case of Sfp, or rests in a shallow pocket with weak force interactions in the case of AcpS (Parris et al., 2000; Reuter et al., 1999). This has meant the pantetheine end of CoA can be modified dramatically with little effect on overall catalytic activity (Quadri et al., 1998). The range of different molecules that have been incorporated into these CoA analogues includes peptides, fluorophores, carbohydrates, biotin, quantum dots, and DNA (George et al., 2004; La Clair et al., 2004; Maillard et al., 2011; Meier and Burkart, 2009; Meier et al., 2006; Sieber et al., 2003; Sunbul et al., 2008). The fact that PPTases can transfer any of these molecules onto any CP domain-like substrate makes them versatile in a wide range of applications including site-specific protein labelling and protein immobilisation.

### 1.3.1 Heterologous expression of NRPS and PKS enzymes

One of the first biotechnological applications of PPTases utilised the fact that some PPTases are able to activate a broad range of CP substrates. In order to characterise or manipulate NRPS or PKS enzymes, their genes are often heterologously expressed in a host strain such as *E. coli.* However the native PPTases of *E. coli* do not have a sufficiently broad activity to routinely activate

foreign NRPS/PKS, leading to the expression of inactive apo-enzymes (Caffrey et al., 1991; Pfeifer et al., 1995; Roberts et al., 1993). Co-expression of a PPTase that has a broad substrate range such as Gsp (Ku et al., 1997) or Sfp (Kealey et al., 1998) can resolve this problem. Despite the incredibly broad substrate range of Sfp, there have been cases documented where it has been unable to activate or fully activate all heterologous CPs (Allen and Bartlett, 2002; Gokhale et al., 1999), indicating it may yet be beneficial to test multiple broad range PPTases or even to find new PPTases that possess a broader substrate specificity.

### 1.3.2 PPTases in natural product discovery

Natural products account for a large number of drugs currently on the market. Microorganisms are the source of many of these natural products, which are often secondary metabolites synthesised by NRPS or PKS enzymes (Li and Vederas, 2009; Osbourn, 2010). The fact that these have evolved to have a biological activity is likely the reason that the "hit rate" for compounds that go on to become therapeutic drugs is orders of magnitude higher for libraries of natural product compounds compared to chemically synthesised compounds (Berdy, 2005). For practical reasons, the majority of the microbially sourced drugs in clinical use today have been discovered from cultivable microbes, yet these are estimated to make up less than 1% of total microbial diversity in existence (Keller and Zengler, 2004; Rappé and Giovannoni, 2003; Torsvik et al., 2002). Efforts to tap into this vast resource of uncultivable microorganisms, by isolating DNA from environments known to be rich in microbial diversity and mining it for genes of interest, are known as metagenomic screening (Daniel, 2004).

The idea of utilising the activity of PPTases in this effort was first described by Yin et al. (2007) who used Sfp to identify CP domains from a library of genome fragments, with the aim of connecting these to the wider NRPS or PKS genes they were associated with. Phage display was used to express protein fragments on phage surfaces, with the gene for the protein of interest contained within the phage. If the displayed protein fragment was recognised by the PPTase Sfp it was then labelled with a CoA biotin analogue, enabling subsequent recovery of the phage particle via streptavidin binding. The effectiveness of this method was demonstrated using a libraries created from *B. subtilis* or *Myxococcus xanthus* DNA. This approach proved to be an effective enrichment method for CP-containing NRPS and PKS genes, finding almost half of the CPs present in *B. subtilis* and almost 20% of the CPs present in *Myxococcus xanthus.* However, the fact that the broad activity PPTase Sfp could only recover half of the CPs in its own genome, and even fewer in a foreign genome, indicates that the approach may work better with an even broader activity PPTase, or using a panel of different PPTases (Yin et al., 2007). The same group then repeated this

experiment with metagenomic DNA, and found it enriched for CP proteins in this metagenomic sample, one of which appeared to be part of a larger NRPS cluster (Zhang et al., 2009).

An interesting side effect of this approach was that it was not only CP domains that were recovered, but also a number of other proteins that turned out to be capable of being recognised by Sfp and labelled with biotin. These proteins appeared to have no relation to CPs but did all bear the recognition sequence DSL (or in some cases the alternative ESL), where S is the serine for the point of phosphopantetheine attachment (Yin et al., 2007; Zhang et al., 2009). The sequence of one of these proteins, YbbR was used to determine a minimal sequence of 11 residues that is capable of being modified by a PPTase, called ybbR or the ybbR tag (Yin et al., 2005). This has been utilised as an alternative to full CP domains in many PPTase labelling applications, as covered in the next section.

Complementary to searching a metagenomic DNA sample to recover CP domains, another approach is to screen for PPTase genes. PPTase genes are often found in the same gene cluster as the biosynthetic genes that they activate, so the discovery of these genes is likely to lead to the discovering of associated NRPS or PKS gene clusters. This was first proposed as an approach by (Mootz et al., 2002b), where it was suggested unknown PPTase genes could be found through complementation screening with Lys5, a PPTase from *Saccharomyces cerevisiae* that is essential under conditions of lysine starvation. A novel approach for the discovery of PPTase genes and associated biosynthetic gene clusters will be the subject of the work presented in Chapter 4 of this thesis.

### 1.3.3 The development of PPTase-mediated site specific protein labelling

CP domains or CP domain derivatives (like the ybbR tag) can be attached to a protein of interest and subsequently serve as substrates for a PPTase, which are then able to label this tag with any molecule present on a CoA analogue. For site specific labelling of target proteins, these labels can be fluorophores or quantum dots, which then enable visualisation of proteins or potentially FRET measurements. This technique is used in the study of proteins and their interactions, allowing visualisation of protein movements and interactions with other proteins and molecules, facilitating the elucidation of protein function (Sunbul and Yin, 2009).

One of the first studies to demonstrate this technique tagged two cell surface receptors with ACP tags, α-agglutinin receptor (Aga2p) on yeast cells and human G protein-coupled receptor neurokinin 1 ($NK_1$) in mammalian cells. AcpS was then used to label these with both fluorophore

and biotin CoA analogues. Highly specific labelling was seen in this *in vivo* system (George et al., 2004). Around the same time, another group created a series of fusion proteins, combining one of two different PCP domains with one of three other proteins, enhanced green fluorescent protein (EGFP), maltose binding protein (MBP) and glutathione-S-transferase (GST). The PCP tags did not appear to affect any of the functions of these proteins. The tags were then successfully labelled with a biotin CoA analogue, confirmed by ELISAs and western blotting (Yin et al., 2004).

The potential to label proteins with CP tags was then subsequently improved upon with the serendipitous discovery of the ybbR tag, as noted in the previous section. The ybbR tag was shown to be able to be fused to other proteins, on either the C- or N- terminus, as well as on internal flexible loops. This tag also retained the ability to be labelled with multiple CoA analogues, demonstrated using several fluorescent tags as well as biotin (Yin et al., 2005). This technology was then further improved by a different application of the phage display method, seeking to deliberately identify small peptide tags capable of recognition and labelling by a PPTase (Zhou et al., 2007a). A library consisting of peptides that were 12 residues long, all being combinatorial variants of the core ybbR sequence, was created and then divided in two. The two libraries were then screened in parallel for peptides that could be recognised by either Sfp or AcpS. This resulted in the discovery of two separate peptides, S6 and A1, which were able to be specifically labelled by Sfp and AcpS respectively. Sfp had a 442-fold higher $k_{cat}/K_m$ with S6 than did AcpS, whereas AcpS had a 30-fold higher $k_{cat}/K_m$ with A1 than Sfp. It was shown that these tags could also be added to cell surface proteins without affecting their functions, and that they could then specifically be separately labelled with fluorescent CoA analogues. The A1 tag was then improved even further still, shortening it to only 8 amino acids (Zhou et al., 2008).

One limitation of the technology is that CoA analogues are unable to pass through the cell membrane, so at this stage are primarily used for labelling of cell surface proteins. This can, however, also be an advantage, as cell surface proteins are then able to be labelled and distinguished with no background fluorescence inside the cell. One study labelled an odorant receptor (a G protein-coupled receptor) using an ACP tag on the N-terminus of the receptor, labelled by a fluorophore CoA analogue added by AcpS, thereby labelling the receptors on the cell surface. Some receptors were also labelled with a GFP protein tag, fused to the intracellular C-terminus of the protein, which labelled all receptors including those internal to the cell. This enabled a visualisation and comparison of cell surface *vs.* total receptor population at any one time, and allowed the tracking of single molecules, following their movement through the membrane as well as their internalisation (Jacquier et al., 2006).

#### 1.3.3.1 Recent examples of PPTases being applied for site specific protein labelling

The past two years have seen an increase in the use of PPTases for site specific labelling of proteins. Due to the limitation of CoA analogues not being able to pass through the cell membrane, the technique is still mainly used for the labelling of cell surface proteins, although there have been some instance of *in vitro* labelling as well.

Various chemokine receptors have been labelled using PPTases. The chemokine receptor CXCR7 was tagged with PCP, then labelled with a fluorescent CoA analogue using Sfp, confirming the expression of this receptor on human B cells. It was also used to differentiate between a full length and truncated form of the receptor, showing that the truncated version could not be internalised (Humpert et al., 2012). In another instance, the chemokine receptor CXCR4 was labelled using a PCP tag and a fluorescent CoA, and time-lapse video microscopy was used to track receptors in real time, showing how the receptors reacted to different stimulants. Cells were also treated with viral proteins (from the Herpes simplex virus), which were found to affect chemokine function (Martinez-Martin et al., 2015). Another group labelled several different chemokine molecules with the S6 tag, and used fluorophores to visualise the interactions of the receptors, demonstrating how this varied for different combinations of chemokine and chemokine receptor (Kawamura et al., 2014).

PPTases have also been used in the study of neurotrophin nerve growth factor (NGF) receptors, which are involved in the development and survival of neurons, as well as neuronal plasticity. The first study used an ACP tag on one of these receptors, tropomyosin receptor kinase A (TrkA), and labelled this with quantum dots and Sfp. The labelled receptor had identical properties to an unmodified protein, and the quantum dots enabled study of protein dynamics and tracking of single molecules on living cells (Callegari et al., 2012). This same technique was subsequently used by the same group, to look at how TrkA responded to different ligands, again using quantum dots to allow long-term tracking of single molecules. This allowed determination of the oligomeric state of the receptor, as well as showing that different ligands resulted in different patterns of receptor movement within the membrane (Marchetti et al., 2013). More recently still, this group has used a dual labelling approach, with the A1 tag on the TrkA and the S6 tag on another NGF receptor, p75NTR (Marchetti et al., 2014). NGF itself was also tagged with A4 (the 8 amino acid residue AcpS specific tag (Zhou et al., 2008)). The tagged and biotinylated NGF was found to be fully functional and still able to interact with receptors. However, it was noted that AcpS had a lower efficiency of labelling than Sfp with their respective tags, and some non-specific labelling was seen.

Nonetheless, it was concluded this type of dual labelling will allow for further study of the interactions between NGF and its receptors, as tracked on living cells. An important advantage of the PPTase system over traditional methods of labelling for these receptors was the stoichiometry of the reaction; previously neurotrophins had been labelled by mainly chemical methods, which give inconsistently labelled proteins (Marchetti et al., 2014).

PPTase mediated site specific labelling has also been used in the study of HIV. The HIV envelope protein, Env, mediates viral entry into human cells. Env was dual labelled with both the A1 tag and a Q3 tag (used for transglutaminase labelling), at two separate locations within the protein. This allowed for labelling with two different fluorophores, a donor and acceptor fluorophore, thereby enabling FRET measurements. The dynamics of the protein were tracked in real time, tracking the different conformations of the protein as it interacted with the human CD4 receptor, including examination of the effects of adding an antibody that decreased the virus's ability to infect cells. HIV Env is an target for vaccines, and it was argued that the information collected from this real-time visualisation could facilitate development of HIV vaccines or other targeted therapies (Munro et al., 2014a; Pancera et al., 2014). The Gag proteins from HIV, which drive virus assembly, have also been studied using this method, in this instance being dual labelled with the S6 and A1 tags and using two different fluorophores for FRET analysis. This enabled the Gag reaction to be monitored *in vitro*, revealing the conformational state of the protein before and during virus assembly (Munro et al., 2014b).

These and a few other examples of other recent uses of PPTases for site-specific labelling are summarised in Table 1.1

**Table 1.1 Recent examples of proteins labelled by PPTase mediate site specific labelling**

| Target Protein | PPTase/tag | Notes | Reference |
|---|---|---|---|
| TrkA | Sfp/ACP | Single molecule tracking over long time periods | (Callegari et al., 2012) |
| CXCR7 | Sfp/PCP | Confirmed expression of receptor on human cells, differentiated from truncated receptor | (Humpert et al., 2012) |
| EGFR | Sfp/ACP | Studied conformational states and ligand induced structural changes | (Ziomkiewicz et al., 2013) |
| TrkA | Sfp/ACP | Response to different ligands via single molecule tracking | (Marchetti et al., 2013) |
| IgE and IgE receptor | Sfp/S6 | Dual labelling with a different, chemical, method with FRET measurement between IgE and its receptor | (Grünewald et al., 2014) |
| Interferon | Sfp/ybbR | Visualisation of the targeting of interferon to specific cells | (Garcin et al., 2014) |
| Histone methyltransferase | Sfp/ybbR | Elucidation of dimeric protein structure | (Davidovich et al., 2014) |
| Synaptotagmin | Sfp/S6 | Interaction of protein domains with the plasmid membrane | (Vasquez et al., 2014) |
| CXCL12, CCL2, CCL21 | Sfp/S6 | Visualisation of reactions between multiple chemokine/receptor combinations | (Kawamura et al., 2014) |
| TrkA<br>P75NTR<br>NGF | AcpS/A1<br>Sfp/S6<br>AcpS/A4 | Multiple labelling of different receptors and ligands | (Marchetti et al., 2014) |
| Gag | Sfp/S6<br>AcpS/A1 | FRET analysis of viral Gag assembly *in vitro* | (Munro et al., 2014b) |
| Env | AcpS/A1 | Also labelled with transglutaminase tag for dual labelling and FRET measurements | (Munro et al., 2014a)<br>(Pancera et al., 2014) |
| Interferon | Sfp/ybbR | Single molecule tracking and receptor assembly | (Wilmes et al., 2015) |
| CXCR4 | Sfp/PCP | Receptors tracked in real time, in response to multiple stimulants | (Martinez-Martin et al., 2015) |
| EGFR | Sfp/ACP | Dynamics of mutant EGFR proteins implicated in non-small cell lung carcinoma | (Valley et al., 2015) |
| Insulin receptors | Sfp/S6 | Interaction between protein domains | (Cabail et al., 2015) |

#### 1.3.3.2 Immobilisation

As a subset of site specific labelling applications, immobilisation occurs when proteins are labelled with a molecule (e.g. biotin) that can be used to immobilise said protein onto a solid surface. Immobilisation is an important step in applications such as drug screening or medical diagnostics (Ebisu et al., 2009). Early protein immobilisation for these applications involved non-specific binding, which was problematic as a very pure protein was required to begin with and it resulted in heterogeneous display of the protein. More recently site-specific covalent protein immobilisation has been explored. In one of the first instances of using PPTases to immobilise

enzymes, Sfp was used to attach a biotin-CoA analogue to a PCP domain that had been fused to a target protein. These biotin-labelled fusions could then be immobilised onto avidin glass slides (Yin et al., 2004). The proteins of interest were able to be isolated from cell lysates and did not lose any activity in the process. This method of immobilisation required two steps, the labelling of the protein of interest and subsequent attachment to a solid support. In a study by Wong et al (2008), a one-step immobilisation was performed. In this case CoA was already covalently attached to a solid support matrix and through the action of Sfp a recombinant protein (containing the ybbR tag) was directly immobilised on the surface. It was shown that proteins could also be immobilised directly from cell lysate (Wong et al., 2008).

Recently these immobilisation techniques have been applied to molecular force assays, which measure the strength of the interactions either between protein molecules or between protein and DNA. This usually requires the protein in question to be immobilised to a solid support. YbbR fusion proteins for enzymes involved in cellulose degradation were immobilised onto solid surfaces, via CoA attached to a surface, in order to study the forces withstood by these enzymes, which has implications in the field of biofuels (Schoeler et al., 2014). Molecular force assays between protein and DNA (as of might be of interest to understand gene expression regulatory mechanisms) were performed again by immobilising ybbR fusion proteins onto CoA coated surfaces. Multiple proteins can be immobilised to the same surface, for high throughput parallel screening (Limmer et al., 2014). Otten et al. (2014) performed a high throughput ybbR mediated immobilisation of multiple proteins, to perform single molecule force assays to characterise the mechanical properties of multiple proteins.

#### 1.3.3.3 PPTases in the formation of hydrogels

PPTases can be used in the formation of bioactive hydrogels. In a proof-of-principle study, ACP dimers were used as the base and point of attachment for CoA-polyethylene glycol (PEG) molecules. Mixing the ACP dimers and CoA-PEG in the presence of Sfp caused a gel to form. Many chemical methods of gel formation lack selectivity and can negatively affect cells and other biomolecules. PPTases avoid these problems as their labelling is specific and controllable, and cells and other biomolecules were successfully incorporated into the PPTase catalysed gels with little loss in activity (Mosiewicz et al., 2009).

### 1.3.4 Discovery of PPTase inhibitors

The essential nature of PPTases in many pathogenic species has led to PPTases becoming a potential target for inhibitors, with the aim being to recover inhibitors that can be used as

antibiotics. PcpS, the only PPTase in the pathogenic species *Pseudomonas aeruginosa*, was identified as a potential antibiotic target as it is the only PPTase present in the organism so it is essential for primary metabolism, as well as playing a key role in synthesis of a number of virulence factors (including the siderophores pyoverdine and pyochelin) (Finking et al., 2002). The PPTase in *M. tuberculosis*, PptT, has also been identified as a potential antibiotic target (Chalut et al., 2006; Leblanc et al., 2012), and is required for growth and persistence of *M. tuberculosis* in a mouse model (Leblanc et al., 2012). Other PPTases have also been shown to be involved in the virulence of a certain pathogenic *E. coli* strain, which contains the PPTase ClbA, required for production of the genotoxin colibactin. At least one PPTase from this pathogenic *E. coli*, either ClbA or EntD, needs to be present for the survival of this organism in a mouse model; a knockout of both PPTases had its virulence abolished (Martin et al., 2013). Owing to the promise of PPTases as an antibiotic target, several screening methods have been developed to find new inhibitors. Screening for these inhibitors has also utilised the activity of PPTases to accept fluorescent CoA analogues, using fluorescence based assays to determine the effect of any potential PPTase inhibitors (Duckworth and Aldrich, 2010; Foley and Burkart, 2009; Yasgar et al., 2010).

## 1.4 Blue pigment synthetase A as a model NRPS

One NRPS of particular relevance to this project is Blue pigment synthetase A (BpsA) from *Streptomyces lavendulae*. BpsA is a single module NRPS which synthesises the blue pigment indigoidine (Takahashi et al., 2007). The structure of BpsA is unusual in that it consists of a single module, which has an oxidation domain located in the middle of the A domain (Fig 1.3 A). This single module catalyses a reaction wherein two molecules of L-glutamine are individually cyclised and then dimerised to form a molecule of indigoidine (Fig 1.3 B). This pigment is easily visualised on either solid (Fig 1.3 C) or in liquid (Fig 1.3 D) media. The advantages of working with BpsA are numerous. The *bpsA* gene is small (for an NRPS) at 3800 bp, making it relatively easy to work with and genetically manipulate. Perhaps in part due to this small size the protein can be purified relatively easily, again unusual for an NRPS. This allows BpsA to be employed in both *in vivo* and *in vitro* colourimetric assays. Like all NRPS, BpsA first needs to be activated by a PPTase (Fig 1.3 A), and the native PPTases of *E. coli* do not activate BpsA (Owen, 2010; Takahashi et al., 2007) (Fig 1.3 D). The exact role of indigoidine in its natural setting is not known, though it has been shown to have some antimicrobial and antioxidant properties (Cude et al., 2012; Reverchon et al., 2002).

### 1.4.1 Applications of BpsA

BpsA has been utilised for several different flexible and elegant applications within our lab, based primarily on the ability to measure BpsA function via indigoidine production. This can provide a measure of the function of BpsA itself, for example when *bpsA* has been manipulated via substitution of foreign CP domains. The activity of such substitutions can subsequently be improved via directed evolution by screening for the production of indigoidine (Owen, 2010). Directed evolution is a process whereby mutations are introduced into a gene and new or altered properties are then screened for and selected, thereby allowing the discovery of new and improved variants. Indigoidine production can be easily screened for in a high throughput manner meaning that this system is adaptable to a range of directed evolution strategies. In Chapter 3, the interactions between the different domains of BpsA were investigated via directed evolution of CP-substituted variants, in order to highlight important residues.

BpsA can also be used to measure PPTase activity. The native PPTases in *E. coli* do not recognise BpsA, enabling purification of BpsA in its inactive, or apo, form. This apo-BpsA can then either be co-expressed with a PPTase *in vivo* or incubated with a PPTase *in vitro* and indigoidine synthesis used as a measure of activation of BpsA by that PPTase. This approach was utilised to discover PPTases from metagenomic DNA (investigated further in Chapter 4), as well as to characterise the kinetic activity of various PPTases (relevant to work described in Chapter 5) (Owen, 2010; Owen et al., 2011). Another application of BpsA is its use in assays to discover PPTase inhibitors, where potential inhibitors are identified and validated through their ability to block indigoidine production (Owen et al., 2011).

More recently we found that the amount of indigoidine produced in the BpsA catalysed reaction is proportional to the amount of L-glutamine present, and developed an assay for the quantification of L-glutamine, resulting in a PCT patent filing (Ackerley et al., 2015). I contributed to the development of this assay, which while not directly related to the research presented in this thesis is summarised in Appendix 1.

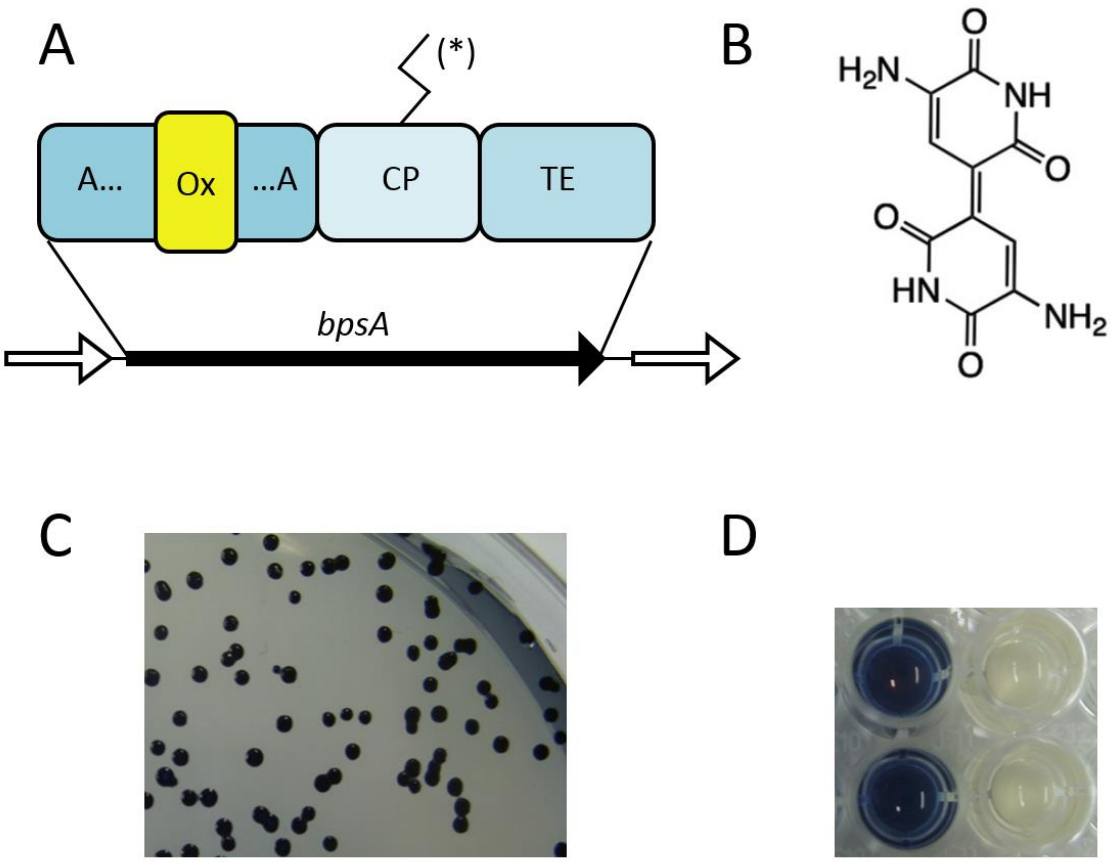



**Figure 1.3 The NRPS BpsA and its product, indigoidine.** A) Gene and protein domain structure of the NRPS BpsA, with (*) indicating that the Ppant arm is required for activation and is added by a PPTase; B) structure of indigoidine; C) indigoidine synthesised by colonies expressing both *bpsA* and an activating PPTase on agar plates containing L-glutamine; D) indigoidine production in L-glutamine containing liquid cultures by *E. coli* expressing *bpsA* and either co-expressing an activating PPTase (Sfp from *B. subtilis*; left) or expressing *bpsA* in the absence of an activating PPTase (right).

## 1.5 Aims of this study

The first aim of this thesis was to use BpsA to elucidate some of the mechanisms involved in CP interactions with the CPs downstream domain partner, enabling understanding of key factors that contribute to successful CP substitutions. Several foreign CP domains were swapped into BpsA to further investigate important residue positions that had been previously identified by Dr Jeremy Owen as playing an important role in determining the functionality of domain substituted NRPS constructs. In addition to this, site directed mutagenesis was applied to fully investigate the contributions of two particular residues within the CP domain. This work was the subject of Chapter 3.

The second aim of this thesis was to use WT BpsA, together with two BpsA variants possessing alternative CP domains, and use these to screen two metagenomic libraries for the presence of novel PPTase and secondary metabolite producing genes. This work was the subject of Chapter 4.

Following the successful discovery of new PPTase genes as part of the work in Chapter 4, the next aim was to characterise these PPTases further, using a series of BpsA based assays to define their activity and substrate specificity. When these were found to be of generally low activity a number of other, already known, PPTase enzymes were selected for additional characterisation using these same assays. This work is presented in Chapter 5.

The final aim of this thesis was to develop a directed evolution strategy in order to alter PPTase substrate specificity, using the PPTase EntD in a proof of principle screening method. The aim here was to lose activity with one CP substrate while retaining it with another, in order to engineer a PPTase that had a narrow substrate specificity. Different CP-substituted variants of BpsA were again used to achieve this, based on the ability to monitor either retention or loss of indigoidine production by different PPTase variants.

# Chapter 2: Material and Methods

## 2.1 General chemical reagents, enzymes and media

Chemicals, reagents and media used in this study were obtained from Sigma-Aldrich (St Louis, MO, USA) or Thermo Fisher Scientific (Waltham, MA, USA), unless otherwise stated. L-glutamine was purchased from iHerb (Perris, CA, USA). Restriction enzymes were purchased from New England Biolabs (NEB; Ipswich, MA, USA). IPTG (isopropyl β-D-thiogalactoside) and BioMix™ Red Polymerase Mastermix were supplied by Bioline (London, UK). Phusion™ high-fidelity DNA polymerase was supplied by Finnzymes (Espoo, Finland). T4 DNA ligase was supplied by either Bioline or Invitrogen (Carlsbad, CA, USA). Mutazyme was obtained from Agilent Technologies (Santa Clara, CA, USA).

## 2.2 Bacterial strains and plasmids

### 2.2.1 Bacterial Strains

All bacterial strains used in this study are given in Table 2.1.

**Table 2.1 *E. coli* strains used in this study**

| Strain | Relevant characteristics | Source |
|---|---|---|
| DH5α λpir | $F^-$ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 *Δ*(*lacZYA-argF*)U169, hsdR17($r_{K-}$ $m_{K+}$), λ(pir) | Invitrogen |
| BL21 | $F^-$ ompT gal dcm lon $hsdS_B(r_B^-\ m_B^-)$ λ(DE3) | Novagen |
| BL21 ΔentD | BL21 Δ*entD* | This study |
| EcoBlue1 | BL21 Δ*entD* + pCDFDuet1::*bpsA* (wild-type) | This study |
| EcoBlueP | BL21 Δ*entD* + pCDFDuet1::*bpsA* (*pvdD* carrier protein domain substitution) | This study |
| EcoBlueE | BL21 Δ*entD* + pCDFDuet1::*bpsA* (e*ntF* carrier protein domain substitution) | This study |

### 2.2.2 Plasmids

All plasmids used in this study are given in Table 2.2.

**Table 2.2 Plasmids used in this study**

| Plasmid | Relevant characteristics | Source |
|---|---|---|
| pET28a(+) | $Kan^R$, expression vector for His-tagged enzyme purification, T7 promoter, ColE1 origin of replication | Novagen |
| pCDFDuet1 | *lacI*$^q$*,*T7prom*,* spec$^R$*,* CDFori | Novagen |
| pETDuet1 | *lacI*$^q$, T7 promoter, amp$^R$, ColE1 origin of replication | Novagen |
| pRSETB | *lacI*$^q$, T7 promoter, amp$^R$, ColE1 origin of replication | Invitrogen |
| pLBPSA3 | pCDFDuet based staging vector allowing introduction of foreign CP domains into *bpsA* with silent restriction sites | (Owen, 2010) |
| pBAD | Amp$^R$, araBAD promoter, *ara*C, pBR322 origin of replication | Invitrogen |
| pKD4 | Amp$^R$, Kan$^R$, Pir-origin of replication, kan$^R$ gene template for KO-PCR | (Datsenko and Wanner, 2000) |
| pKD46 | Amp$^R$, Heat sensitive origin of replication, recombinase expression vector, ara operon promoter | (Datsenko and Wanner, 2000) |
| pCP20 | Amp$^R$, heat sensitive origin of replication, encodes FLP recombinase | (Datsenko and Wanner, 2000) |

## 2.3 Oligonucleotide primers

Oligonucleotide primers used in this study were synthesised by Integrated DNA Technologies (IDT; Carlsbad, CA, USA). Primers were supplied lyophilised and on receipt were made up to 100 µM stock solutions with 1x TE buffer, pH 8.0 (10 mM Tris-Cl pH 8.0, 0.1 mM EDTA). Working stocks were diluted to 10 µM in filter sterilised autoclaved dd$H_2O$ and both working and stock solutions were stored at -20 °C. The names and sequences of primers used in this study are listed in Table 2.3, with restriction sites underlined.

**Table 2.3 Oligonucleotide primers used in this study**

| Primer Name | Sequence (5' → 3') |
|---|---|
| PPTase specific primers | |
| Ps1448a_PPT_fwd | CTAGCATATGCTGGTCAGTGG |
| Ps1448a_PPT_rev | GTACGTCGACTCAGCCTTCG |
| entD_pBAD_fwd | AGCTCTCGAGATGGTCGATATG |
| entD_pBAD_rev | AGCTAAGCTTTTAATCGTGTTGG |
| entD_pET_fw | AGCTCATATGATGGTCGATAT |
| entD_pET_rev | AGCTGTCGACTTAATCGTGTT |
| Swhit1_fwd | GTACCATATGGATTGTCTCGC |
| Swhit1_rev | GTACAAGCTTTCACAGCTCGGT |
| Swhit2_fwd | GTACCATATGAAGCCATGGTC |
| Swhit2_rev | GTACCTCGAGCTACCTCGCTG |
| Swhit3_fwd | GTACCATATGGACAGCAAATGG |
| Swhit3_rev | GTACGTCGACTCAGTCCCCGCGAA |
| Swhit4_fwd1 | GTACCATATGCGTTCGCCGAGT |
| Swhit4_fwd2 | GTACCATATGGGCTGCTCCGTA |
| Swhit4_rev | GTACGTCGACCTACGAACTCGA |

| | |
|---|---|
| Swhit5_fw | GTACCATATGACCGGCATTGAA |
| Swhit5_rev | GTACGTCGACTCAAGGGCAGTG |
| Swhit6_fwd | GTACCATATGAGCGGAGGC |
| Swhit6_rev | GTACAAGCTTCTACTTGATCCG |
| Swhit7_fwd1 | GTACCATATGGCGCTCCCGGTC |
| Swhit7_fwd2 | GTACCATATGATGTTGCATGTCGGG |
| Swhit7_rev | GTACCTCGAGTCATTTGGTTGA |
| NZhit1_fwd | GTACCATATGGAGGAGCTGCTT |
| NZhit1_rev | GTACGTCGACTCAGGCGCCG |
| NZhit2_fwd | GTACCATATGAAGCCATGGTC |
| NZhit2_rev | GTACCTCGAGCTACCTCGCTGA |
| NZhit3_fwd1 | GTACCATATGGACTACCAAGC |
| NZhit3_fwd2 | GTACCATATGGCCCGGTTGG |
| NZhit3_rev | GTACCTCGAGCTAGTGCACC |
| NZhit4_fwd | GTACCATATGCACGACGAGGAT |
| NZhit4_rev | GTACCTCGAGTCACGCGCC |
| NZhit5_fwd | GTACCATATGGCAGATCTCAAT |
| NZhit5_rev | GTACGTCGACTCAGGGCAT |
| NZhit6_fwd | GTACCATATGAGCGCGGCAC |
| NZhit6_rev | GTACCTCGAGTCAGGCCCA |
| NZhit7_fwd | GTACCATATGACCCGCTATC |
| NZhit7_rev | GTACGTCGACTCAATCAGTC |
| NZhit8_fwd | GTACCATATGCCTGTCGTAT |
| NZhit8_rev | GTACGAATTCTTAGGTCCATG |
| NZhit9_fwd | GTACCATATGATCACCTCGC |
| NZhit9_rev | GTACGTCGACTCACTCTATTAG |
| NZhit10_fwd | GTACCATATGAGCAGGCATAAC |
| NZhit10_rev | GTACGTCGACCTAAGCAGGC |
| NZhit11_fwd | GTACCATATGACCATCTGGCCG |
| NZhit11_rev | GTACCTCGAGTTATTTCGGCTG |
| NZhit12_fwd | GTACCATATGCGAAGAACCG |
| NZhit12_rev | GTACGTCGACCTAAGGTTCC |
| NZhit13_fwd | GTACCATATGGAAACGGCATC |
| NZhit13_rev | GTACGTCGACTCAGGCCATG |
| NZhit14_fwd | GTACCATATGCTGCACCAAC |
| NZhit14_rev | GTACGTCGACTCAGCTGCG |
| | |
| entD KO primers | |
| EntDKOFw | ATGGTCGATATGAAAACTACGCATACCTCCCTCCCCTGTGTAGGCT GGAGCTGCTTC |
| EntDKORv | TTAATCGTGTTGGCACAGCGTTATGACTATCTTTTCCATATGAATA TCCTCCTTAG |
| EntDExnFw | CGTTTGTCATCAGTCTCGAATATGGTCGATATGAAA |
| EntDExnRv | AATCGTACCAGATGTTGTCAATTAATCGTGTTGGCA |

| | |
|---|---|
| CP primers | |
| EntF_fwd | CTAGCATATGGGGCGTGCGCCGAAAG |
| EntF_rev | GTACGTCGACTCATTCGAATCC |
| | |
| For PvdDCP1 substitutions | |
| slPvdDT_Fwd | AGCTCTGCAGAGCTCGTCGAGCGCCCCTATCGAGCGCCCGGTAGC |
| slPvdDT_Rev | AGCTTCTAGACTCCTGGGCGACCTCGCGTTCCAATCCCTGGGCGAA |
| PvdDT1_NNK4_ultramer | GCGCCCGGTAGCGAGCTGGAGCAGCGCATCGCAGCGATCTGGTCGGAGATCCTGGGAGTGGAACGGGTCGGCCTGGACGACAACTTCTTCGAACTGGGCGGTCATTCGTTGCTGGCTNNKCGGGTGATTTCTCGGGTTCGCCAGGAGCAGCAGTTGGACGCAAGCCTGAAGGCGTTGT |
| PvdDT1_NNK_rev | TCCCTGGGCGAACGCTTCCAGAACCGGCCGCTCGAACAACGCCTTCAGGC |
| PvdDT1_NNK_fwd | GCGCCCGGTAGCGAGCTGGAGCAGCGCATCGCAGCGATCTGGTCGGAGAT |
| PvdD_NNK24_ultra | TCCCTGGGCGAACGCTTCCAGAACCGGCCGCTCMNNCAACGCCTTCAGGCTTGCGTCCAACTGCTGCTCCTGGCGAACCCGAGAAATCACCCGGGTAGCCAGCAACGAATGACCGCCCAGTTCGAAGAAGTTGTCGTCCAGGCCGACCCGTTCCACTCCCAGGATCTCCGACCAGATC |
| PvdDT1_NNK4_NNK24_ultra | TCCCTGGGCGAACGCTTCCAGAACCGGCCGCTCMNNCAACGCCTTCAGGCTTGCGTCCAACTGCTGCTCCTGGCGAACCCGAGAAATCACCCGMNNAGCCAGCAACGAATGACCGCCCAGTTCGAAGAAGTTGTCGTCCAGGCCGACCCGTTCCACTCCCAGGATCTCCGACCAGATC |
| | |
| Miscellaneous primers | |
| T7 promoter | TAATACGACTCACTATAGGG |
| T7 terminator | GCTAGTTATTGCTCAGCGG |
| pET Upstream | ATGCGTCCGGCGTAGA |
| DuetUP2 | TTGTACACGGCCGCATAATC |
| BpsA_T_screen | ACGAGCAGATCGGCCACGA |
| PBAD-F | ATGCCATAGCATTTTTATCCA |
| PBAD-R | GATTTAATCTGTATCAGG |

## 2.4 Synthesis of genes

The genes synthesised in this study are listed in Table 2.4. The PPTases Svp and Gsp were synthesised by GenScript (Piscataway, NJ, USA) and the remainder were synthesised as gBlocks by Integrated DNA Technologies (IDT; Carlsbad, CA, USA). All genes were codon optimised for *E. coli* using OPTIMIZER (Puigbò et al., 2007, 2008), using the guided random method. The sequences for all codon optimised synthesised genes are listed in Appendix 1.

**Table 2.4 Genes synthesised in this study**

| Gene |
|---|
| PPTases |
| Sfp |
| Svp |
| Gsp |
| MtaA |
| Carrier protein domains |
| ACMSIII |
| GrsB |
| TycC |
| Pris3 |
| BlmI |
| SgcC2 |
| MtaF |
| TcmM |

## 2.5 Bacterial growth and maintenance

### 2.5.1 Media

Unless otherwise stated all media was made up to the appropriate volume with $ddH_2O$ and autoclaved for 30 min at 121 °C to sterilise prior to use. Media was stored at room temperature or at 4 °C if containing heat sensitive supplements.

#### 2.5.1.1 Lysogeny Broth (LB)*

| Component | Final Concentration |
|---|---|
| Tryptone | 10 $g.L^{-1}$ |
| Yeast extract | 5 $g.L^{-1}$ |
| NaCl | 10 $g.L^{-1}$ |

*LB was supplied as a premixed powder by the manufacturer.

#### 2.5.1.2 ZYP5052 auto-induction medium

##### 2.5.1.2.1 ZY base

| Component | Final Concentration |
|---|---|
| N-Z Amines AS | 10 $g.L^{-1}$ |
| Yeast extract | 5 $g.L^{-1}$ |

#### 2.5.1.2.2 20x NPS

| Component | Final Concentration |
|---|---|
| $NH_4SO_4$ | 66 $g.L^{-1}$ |
| $KH_2PO_4$ | 136 $g.L^{-1}$ |
| $Na_2HPO_4$ | 142 $g.L^{-1}$ |

#### 2.5.1.2.3 50x 5052

| Component | Final Concentration |
|---|---|
| Glycerol | 250 $g.L^{-1}$ (200 $ml.L^{-1}$) |
| D-glucose | 25 $g.L^{-1}$ |
| L α-lactose | 100 $g.L^{-1}$ |

#### 2.5.1.2.4 Autoinduction media

| Component | Final Concentration |
|---|---|
| ZY base* | 929 $ml.L^{-1}$ |
| 20x NPS* | 50 $ml.L^{-1}$ |
| 50x 5052* | 20 $ml.L^{-1}$ |
| $MgSO_4$** | 1 mM |

*Solutions were prepared and autoclaved separately, then mixed immediately prior to use.
**Filter sterilised and added post-autoclave.

### 2.5.1.3 SOC*

| Component | Final Concentration |
|---|---|
| Tryptone | 20 $g.L^{-1}$ |
| Yeast extract | 5 $g.L^{-1}$ |
| NaCl | 0.5 $g.L^{-1}$ |
| KCl | 2.5 mM |
| $MgCl_2$** | 10 mM |
| Glucose** | 20 mM |

*pH adjusted to 7.0 using NaOH pre-autoclave
**Filter sterilised and added post-autoclave

#### 2.5.1.4 GYT medium*

| Component | Final Concentration |
|---|---|
| Glycerol | 100 $ml.L^{-1}$ |
| Yeast extract | 1.25 $g.L^{-1}$ |
| Tryptone | 2.5 $g.L^{-1}$ |

*Filter sterilised

#### 2.5.1.5 Solid Media

For solid media, agar was added to media at a concentration of 1.5% (w/v) prior to autoclaving. Any necessary supplements were added after the agar had cooled to <50 °C post-autoclave. Agar was then poured in approximately 20 ml aliquots into sterile 90 mm diameter Petri dishes and left to solidify at room temperature. Plates were stored at 4 °C.

##### 2.5.1.5.1 Iron limited pigment production solid medium

| Component | Final concentration |
|---|---|
| M9 salts* | 1X |
| Agar* | 1.7% (w/v) |
| L-glutamine* | 100 mM |
| Glycerol** | 0.4% (v/v) |
| $MgSO_4$ ** | 2 mM |
| 2,2-dipyridyl** | 100 µM |
| L-arabinose** | 0.01% (w/v) |
| $CaCl_2$** | 0.1 mM |

*M9 salts were autoclaved separately to agar and L-glutamine (each in half total volume $ddH_2O$)
**Filter sterilised and added post autoclave

### 2.5.2 Media supplements

Antibiotic stock solutions were made up to 1000x final concentration in the appropriate solvent and filter sterilised using a 0.22 µM filter before being stored at -20 °C. Antibiotics and relevant concentration are given in Table 2.5. IPTG stocks were prepared to a final concentration of 100$mg.ml^{-1}$. L-arabinose stocks were prepared to a final concentration of 200 $mg.ml^{-1}$. Stocks of 2,2-dipyridyl were made to a final concentration of 100 $mg.ml^{-1}$.

**Table 2.5 Antibiotic supplements used in this study**

| Antibiotic | Stock solution | Solvent |
|---|---|---|
| Ampicillin | 100 $mg.ml^{-1}$ | $ddH_2O$ |
| Kanamycin | 50 $mg.ml^{-1}$ | $ddH_2O$ |
| Spectinomycin | 50-100 $mg.ml^{-1}$ | $ddH_2O$ |

### 2.5.3 Bacterial growth and storage

Unless otherwise stated, bacteria were grown at 37 °C, and aerated at 200 rpm when grown in liquid culture. For short term maintenance, bacteria on agar plates were stored at 4 °C. For long term storage 600-700 µl of overnight liquid culture was mixed in a 1:1 ratio with autoclaved 80% (v/v) glycerol and stored at -80 °C indefinitely.

## 2.6 Routine molecular biology

### 2.6.1 Polymerase chain reaction (PCR)

PCR amplification reactions for cloning were performed using Phusion™ high fidelity polymerase. For other applications including colony screening, Biomix Red™ was employed. PCR reactions were set up according to the manufacturer's protocols, as outlined in Table 2.6. Standard PCR parameters are given in Table 2.7. Parameters were optimised as required for individual genes.

**Table 2.6 Composition of PCR reactions**

| **Phusion™** | |
|---|---|
| **Component** | **Volume per 20 µl reaction** |
| 5x Phusion™ HF buffer | 4 µl |
| 10 mM dNTPs | 0.4 µl |
| 10 µM fwd primer | 1 µl |
| 10 µM rev primer | 1 µl |
| 50 ng/µl Template DNA | 0.5 µl |
| DMSO | 1 µl |
| $ddH_2O$ | to 20 µl |

| **Biomix Red™** | |
|---|---|
| **Component** | **Volume per 20 µl reaction** |
| 2x Biomix™ master mix | 10 µl |
| 10 µM fwd primer | 1 µl |
| 10 µM rev primer | 1 µl |
| 50 ng/µl Template DNA* | 0.5 µl |
| $ddH_2O$ | to 20 µl |

*For colony screen PCR reactions, a small amount of *E. coli* colony picked from an agar plate with a sterile pipette tip was used as the template

**Table 2.7 Standard PCR cycling parameters**

| Temperature (°C) | Time | |
|---|---|---|
| **Phusion™** | | |
| 98 | 1 min | |
| 98 | 10 s | 20-25 cycles |
| 56 | 20 s | |
| 72 | 30 s/kb | |
| 72 | 5 min | |
| **Biomix Red™** | | |
| 95 | 5 min | |
| 95 | 20 s | 10-15 cycles |
| 56 | 30 s | |
| 72 | 30 s/kb | |
| 95 | 20 s | 15-20 cycles |
| 52 | 30 s | |
| 72 | 30 s/kb | |
| 72 | 5 min | |

### 2.6.2 Agarose gel electrophoresis

To assess the size and quality of DNA products including those of PCR reactions, DNA was run on a 1% (w/v) agarose gel, containing 1 $\mu g.ml^{-1}$ ethidium bromide. Gels were run submerged in 1x TAE buffer (40 mM Tris-acetate, pH 8.0, 1 mM EDTA) at 120-130 V for 30-40 min. Hyperladder I (Bioline, London, UK) was run alongside the samples for size comparison and bands were visualised under ultra-violet (UV) light.

### 2.6.3 PCR product purification

PCR products were purified using a Zymo-Spin™ column (Zymo Research; Irvine, CA, USA) according to the manufacturer's instructions. Products were eluted in autoclaved $ddH_2O$ and stored at -20 °C.

### 2.6.4 Restriction enzyme digests

Restriction digests were performed according to the manufacturer's instructions and heat inactivated at 65 °C for 20 min prior to use in further ligation reactions. Where enzymes were unable to be heat inactivated, digests were immediately clean and concentrated using a Zymo-Spin™ column. Digests that were not used immediately were stored at -20 °C.

### 2.6.5 Ligation

Ligations were performed according to the manufacturer's instructions, using a 1:3 molar ratio of vector to insert, with each reaction containing 100-150 ng of total DNA. Reactions were incubated at 25 °C for 90 min. Ligations were then used to transform competent cells or else stored at -20 °C.

### 2.6.6 Bacterial cell transformation and electroporation

#### 2.6.6.1 Preparation of chemically competent cells

The *E. coli* strain to be made competent was inoculated from a glycerol stock and grown in 3 ml LB amended with 10 mM $MgCl_2$ and any relevant antibiotics at 37 °C, 200 rpm overnight. The following day 50 ml of LB supplemented with 10 mM $MgCl_2$ and any relevant antibiotics was inoculated from this overnight culture to give a final $OD_{600}$ of 0.05. Cells were incubated at 37 °C, 200 rpm until an $OD_{600}$ of 0.4-0.6 was reached. Cells were then cooled on ice for 15 min then centrifuged at 2700 g for 15 min at 4 °C. The cell pellet was resuspended in 1 volume of ice cold TFBI buffer (buffers used in this protocol are given in Table 2.8. Cells were then incubated on ice for 2 h before being centrifuged again at 2700 g for 15 min at 4 °C. The resultant cell pellet was resuspended in 0.1 volume ice cold TFBII buffer. Cells were transferred into sterile, pre-chilled 1.5 ml microcentrifuge tubes in 100-200 µl aliquots, snap frozen and stored indefinitely at -80 °C.

**Table 2.8 Competent cell protocol buffer recipes**

| 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) |

#### 2.6.6.2 Transformation of chemically competent cells

Frozen cell aliquots were removed from -80 °C storage and thawed on ice prior to addition of plasmid DNA. For each transformation, DNA in a volume of less than 10 µl was added to 100 µl of cells. This mixture was left on ice for at least 20 min after which transformation was induced by

by heat shock at 42 °C for 90 s. Cells were then returned to ice for 1-2 min and then recovered in 900 µl LB at 37 °C, 200 rpm for 60-90 min. Finally cells were spread on LB agar plate supplemented with appropriate antibiotics and grown overnight at 37 °C.

#### 2.6.6.3 Preparation of electrocompetent cells

The *E. coli* strain to be made competent was inoculated from a glycerol stock and grown in 10 ml LB supplemented with any relevant antibiotics at 37 °C, 200 rpm overnight. This overnight culture was then used to inoculate 400 ml LB containing any relevant antibiotics. This culture was grown until an $OD_{600}$ of 0.30-0.35 was reached, at which point it was transferred to 8 sterile 50 ml tubes and cooled on ice for 30 min. Cells were then centrifuged for 30 min at 1000 g, 4 °C. For the first wash step, pellets were resuspended in a total volume of 400 ml ice cold sterile $ddH_2O$ by gentle pipetting. Cells were then centrifuged again for 30 min at 1000 g, 4 °C and the resulting pellet was washed in 200 ml ice cold sterile 10% (v/v) glycerol. Cells were pelleted again through centrifugation, and this time the pellet was resuspended in 100 ml ice cold sterile 10% (v/v) glycerol. Cells were then finally collected through centrifugation and the pellet was resuspended in 300-500 µl of ice cold sterile GYT. The $OD_{600}$ of a 1/100 dilution was then measured ($OD_{600}$ of 1.0 = $2.5x10^8$ cells/ml) and the final concentration of cells was adjusted to 2-3 x $10^{10}$ cells/ml with the addition of further GYT if required. 40 µl aliquots were transferred to pre-chilled 1.5 ml microfuge tubes on ice. Cell aliquots were then snap frozen in a metal tube block which had been pre-chilled to -80 °C. Aliquots of competent cells were stored at -80 °C until needed.

#### 2.6.6.4 Transformation of electrocompetent cells

Frozen cell aliquots were removed from -80 °C storage and thawed on ice. Plasmid DNA was then added to the cells, typically 50 ng in a volume of sterile $ddH_2O$ not exceeding $1/10^{th}$ of the cell aliquot. Contents of the tube were then gently mixed and transferred to an ice cold sterile 2 mm gap electroporation cuvette. Cells were left on ice for 15 min. Cells were then electroporated at 2.5 kV, 25 µF, 100 Ω, following which 10 volumes of SOC broth were immediately added to the cuvette, which was subsequently transferred to a sterile 15 ml tube. Cells were incubated at 37 °C, 200 rpm for 1 h for recovery. Finally, cells were plated on LB agar plates containing the appropriate antibiotics. Undiluted cells (100 µl), as well as 1/10 and 1/100 dilutions, were plated initially to determine the optimal dilution for single colonies to be obtained on agar plates. The remainder of the transformation was mixed in a 1:1 ratio with sterile 80% (v/v) glycerol and stored at -80 °C until needed.

### 2.6.7 Isolation of plasmid DNA

#### 2.6.7.1 Miniprep

Isolation of plasmid DNA was achieved using either the Zyppy™ Plasmid Miniprep Kit (Zymo Research; Irvine, CA, USA) or the High-Speed Plasmid Mini Kit (Geneaid Biotech Ltd; New Taipei City, Taiwan) according to the manufacturer's instructions.

#### 2.6.7.2 Midiprep

Where high concentrations of plasmid were required for library generation, midipreps were performed using the Geneaid™ Midi Plasmid Kit (Geneaid) according to the manufacturer's instructions.

### 2.6.8 DNA quantification

Measurement of the concentration and purity of DNA was achieved using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific; Waltham, MA, USA), according to the manufacturer's instructions.

### 2.6.9 DNA sequencing

All DNA sequencing performed in this study was carried out by Macrogen Inc. (Seoul, South Korea).

## 2.7 Gene knock out in *E. coli*

Gene knock out of *entD* in *E. coli* was based on the protocol developed by Datsenko and Wanner (2000). Primers were designed to contain a 30 bp sequence homologous to the start and end of the gene at the 5' end and 20 bp homologous to the pKD4 plasmid at the 3' end. These primers were used to amplify a ~1500 bp section of the pKD4 plasmid, resulting in the amplification of a kanamycin resistance gene flanked by FRT (FLP recognition target) sites, with regions of the *entD* gene flanking this. This PCR product was concentrated to ensure a final concentration of ≥ 100 $ng.\mu l^{-1}$. Meanwhile the *E. coli* strain to be altered, BL21, was transformed by the heat sensitive plasmid pKD46, and transformation plates were grown at 30 °C. BL21 containing pKD46 was then induced using L-arabinose before being made electrocompetent and transformed by the pKD4 derived PCR product where, through homologous recombination between the regions of gene homology on the outside of the kanamycin cassette, the gene *entD* was knocked out and replaced by the kanamycin cassette. The pKD46 plasmid was then cured through growth at 42 °C. To

remove the kanamycin resistance cassette, the heat sensitive plasmid pCP20 was transformed into this strain. This plasmid contains the FLP recombinase gene which recognises the FRT sites to remove the kanamycin resistant cassette, leaving a 200 bp "scar" in the place of the *entD* gene. Finally the pCP20 plasmid was cured via growth at 42 °C.

## 2.8 Protein expression and purification

All proteins to be expressed were cloned into His-tag vectors for purification via nickel affinity chromatography.

### 2.8.1 Expression in LB at low temperatures

A single colony or glycerol stock of the expression strain was used to inoculate 10 ml LB supplemented with appropriate antibiotics and grown overnight at 37 °C, 200 rpm. The following day this overnight culture was used to inoculate 400 ml of LB containing any appropriate antibiotics, giving a final $OD_{600}$ of 0.05-0.1. This was incubated at 37 °C, 200 rpm until an $OD_{600}$ of 0.4-0.6 was reached. Cultures were then transferred to 18 °C for 30-60 min before addition of 0.5 mM IPTG. Cultures were then grown for an additional 24 h before cells were harvested through centrifugation at 2700 g, 20 min, 4 °C. The supernatant was discarded and cell pellets were stored at -20 °C until required.

### 2.8.2 Expression in LB containing additional supplements at low temperatures

A single colony or glycerol stock of expression strain was used to inoculate 10 ml of LB supplemented with appropriate antibiotics as well as 2.5 mM betaine and 1 M D-sorbitol and grown overnight at 37 °C, 200 rpm. The following day 400 ml of the same medium was inoculated by the overnight culture to give a final $OD_{600}$ of 0.05-0.1. This was incubated at 37 °C, 200 rpm until an $OD_{600}$ of 0.4-0.6 was reached. Cultures were then transferred to 18 °C for 30-60 min before addition of 0.5 mM IPTG. Cultures were then grown for an additional 24-28 h before cells were harvested through centrifugation at 2700 g, 20 min, 4 °C. The supernatant was discarded and cell pellets were stored at -20 °C until needed.

### 2.8.3 Expression using auto-induction media

A single colony or glycerol stock was used to inoculate 10 ml of LB supplemented with the appropriate antibiotics and grown overnight at 37 °C, 200 rpm. The following day 400 ml ZYP5052

medium containing appropriate antibiotics was inoculated by the overnight culture to give a final $OD_{600}$ of 0.05-0.10. This was incubated at 18 °C, 200 rpm for 24-48 h before being harvested by centrifugation as in 2.8.1.

### 2.8.4 Cell lysis and soluble fraction separation

For protein purification, cell pellets were resuspended in 1/20 vol of the expression culture of 1x His•bind™ binding buffer, minus imidazole. For BpsA purification this binding buffer contained 12.5% (v/v) glycerol, and for PPTase purification this binding buffer contained 25% (v/v) glycerol. Cells were lysed by 2-3 passages through a French press chamber, at 40,000 psi. The French press chamber was cooled to 4 °C prior to use and buffer was kept ice cold. Following lysis the soluble and insoluble fractions were separated by centrifugation at 26890 g, 30 min, 4 °C.

### 2.8.5 Protein purification via Ni-NTA affinity chromatography

#### 2.8.5.1 Purification of BpsA

Purification of $His_6$-tagged proteins was achieved using Novagens His•bind™ Ni-NTA chromatography kit. The following changes to the manufacturer's instructions were made for the purification of BpsA (due to the weak binding of BpsA to the resin). All elution buffers contained 12.5% (v/v) glycerol and all reagents and equipment were kept at ice cold temperatures for the duration of the purification. A large volume, approximately 6-7 ml, of resin was used. The first pass through of the soluble fraction through the resin was collected, and passed through the column a second time. Columns were washed using 15 ml of binding buffer containing no imidazole and then 50 ml of standard binding buffer. The wash buffer step that is usually employed at this stage of purification was omitted for BpsA. For the elution the resin was resuspended in 8 ml elution buffer, and protein was eluted in an 8-10 ml total volume. Buffer exchange and concentration of the eluted protein was performed in 100 kDa molecular mass cut-off Millipore Amicon® ultra-15 centrifugal filter units (Merck Millipore, Billerica, MA, USA), into 50 mM sodium phosphate buffer, pH 7.8, 12.5% (v/v) glycerol. Final buffer composition was then adjusted to 40% (v/v) glycerol and aliquots were stored at -20 °C.

#### 2.8.5.2 Purification of PPTases

Amendments were also made to the standard protocol for PPTase purification, due to the instability and low yield typically observed with these proteins. All elution buffers contained 25% (v/v) glycerol and all reagents and equipment were kept at 4 °C for the duration of the experiment.

As with BpsA, the soluble fraction for PPTase purification was also passed through the resin column twice. Columns were washed using 10-15 ml binding buffer (no imidazole), 10-15 ml standard binding buffer, 10-15 ml wash buffer (half standard imidazole concentration), and 5 ml standard wash buffer. Following elution, protein was immediately desalted using GE Healthcare HiTrap™ desalting columns, according to the manufacturer's protocol. Desalting buffer for PPTases was 50 mM Tris-Cl pH 7.5, 12.5% (v/v) glycerol. Glycerol was added to desalted protein to bring final concentration to 50% (v/v) glycerol and aliquots were stored at -80 °C.

#### 2.8.5.3 Purification of carrier proteins

CPs were purified following the same protocol as for PPTases.

### 2.8.6 Determination of protein concentration

The final concentration of the purified PPTase proteins was determined using a DC™ protein assay kit (Bio-Rad, Hercules, CA, USA), following the manufacturer's instructions. Protein standards were made with bovine serum albumin (BSA) in the appropriate buffer and glycerol concentrations, ranging from 0.25-1.5 $mg.ml^{-1}$. The concentration of BpsA was first determined using absorbance at 280 nm using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific), according to the manufacturer's instructions.

## 2.9 SDS PAGE

For qualitative assessment of proteins, 15-17% SDS PAGE gels were run to visualise proteins. These were cast and run on a 1 mM thickness Bio-Rad Protean II™ apparatus. For each gel 5 ml of separating gel was loaded into the apparatus before being overlaid with sterile $ddH_2O$ and left to set for 45 min. The $ddH_2O$ was then removed and 1.5 ml of stacking gel was poured on top. A multi-well comb was inserted and this was left to set for at least 45 min. If not used immediately gels were stored for up to three days at 4 °C, wrapped in wet paper towel and tin foil. Samples were added to 3x loading buffer in a 2:1 ratio, and incubated at 95 °C for 10 min. Gels were run in 1x SDS run buffer, at a constant voltage (180 V) for 60 - 75 min, until the dye front reached the end of the gel. Gels were then stained in Coomassie blue stain and gentle shaking was applied for 30-60 min. Gels were then rinsed using tap water and left in destain solution with gentle rocking until the desired level of destain was reached. All recipes for SDS protocol can be found in Table 2.9.

**Table 2.9 Recipes for SDS PAGE materials**

| Component | Final concentration |
|---|---|
| **15% separating gel** | **Per 10 ml** |
| 40% acrylamide solution (w/v) | 3.65 ml |
| 2% Bis-acrylamide solution (w/v) | 2.0 ml |
| 1.5 M Tris-Cl pH 8.8 | 2.5 ml |
| 10% SDS (w/v) | 100 µl |
| 10% APS* (w/v) | 100 µl |
| TEMED* | 6 µl |
| $ddH_2O$ | 1.75 ml |
| | |
| **17% Separating gel** | **Per 10 ml** |
| 40% acrylamide solution (w/v) | 4.14 ml |
| 2% Bis-acrylamide solution (w/v) | 2.28 ml |
| 1.5 M Tris-Cl pH 8.8 | 2.28 ml |
| 10% SDS (w/v) | 100 µl |
| 10% APS* (w/v) | 76 µl |
| TEMED* | 7.5 µl |
| $ddH_2O$ | 1.12 ml |
| | |
| **5% stacking gel** | **Per 6 ml** |
| 40% acrylamide solution (w/v) | 310 µl |
| 2% Bis-acrylamide solution (w/v) | 580 µl |
| 0.5 M Tris-Cl pH 6.8 | 1.5 ml |
| 10% SDS (w/v) | 60 µl |
| 10% APS* (w/v) | 46 µl |
| TEMED* | 7.5 µl |
| $ddH_2O$ | 3.5 ml |
| | |
| **3x SDS loading buffer** | **Final concentrations** |
| Tris-Cl pH 6.8 | 150 mM |
| SDS | 6% (w/v) |
| Bromophenol blue | 0.3% (w/v) |
| Glycerol | 30% (w/v) |
| β-mercaptoethanol | 300 mM |
| | |
| **1x run buffer** | **Per 1 L** |
| Glycine | 14.4 g |
| Tris | 3.03 g |
| SDS | 1 g |
| | |
| **Coomassie blue stain** | **Per 1 L** |
| Coomassie brilliant blue | 2.5 g |
| 100% ethanol | 450 ml |
| 100% acetic acid | 100 ml |
| $ddH_2O$ | 450 ml |

| **Destain solution** | **Per 1 L** |
|---|---|
| 100% methanol | 400 ml |
| Acetic acid | 100 ml |
| dd$H_2O$ | 500 ml |

*Added immediately prior to pouring gel
APS is Ammonium persulfate

# 2.10 *In vivo* assessment of indigoidine production

## 2.10.1 Indigoidine production as assessed on solid media

Strains to be tested were plated out on either LB or auto-induction agar plates containing appropriate antibiotics and 100 mM L-glutamine, at a density of approximately 100 colonies per plate, and incubated overnight at 37 °C. For LB plates, following overnight growth, LB agar slabs were lifted out from the Petri dish and placed on a sterile lid. For each plate, 100 µl of 2.5% IPTG (w/v) was spread evenly across the base of the empty Petri dish and the agar slab was replaced on top. For auto-induction agar plates no further action was required. Plates were then incubated at room temperature for up to four days to observe development of pigmentation.

## 2.10.2 Indigoidine production as assessed in liquid media

Glycerol stocks of appropriate strains were used to inoculate 350 µl of LB containing appropriate antibiotics and then incubated overnight at 37 °C, 200 rpm. 20 µl of each culture were used to inoculate quadruplicate wells within a 96-well microplate, each containing 130 µl LB amended with 100 mM L-glutamine, 0.6 mM IPTG, and appropriate antibiotics. Plates were wrapped in foil and incubated at 18 °C, 200 rpm for 24-48 h. At appropriate time points, $OD_{590}$ and $OD_{800}$ values were recorded using an EnSpire 2300 Multilabel Reader (Perkin Elmer; Waltham, MA, USA). After a reading was taken the plate was again incubated at 18 °C, 200 rpm.

### 2.10.2.1 Normalisation of indigoidine absorbance

A normalised value for indigoidine absorbance was calculated using the method described by (Beer et al., 2014), as adapted from (Myers et al., 2013). The OD values recorded represent the indigoidine sensitive wavelength ($OD^S$) of 590 nm; as well as the robust wavelength ($OD^R$) of 800 nm, which accounts for the cellular components present. $OD^S_{B,+I}$ is the OD at 590 nm which contains absorption from both bacterial cells and the indigoidine pigment, while $OD^S_{B,-I}$ accounts for the scattering of just the cellular components at $OD^S$ (i.e. at 590 nm). $OD^S_{B,-I}$ is calculated as:

$$OD^{S}_{B,-I} = \delta.OD^{R}$$

where δ is the correction factor and is calculated by measuring the ratio of pure cell culture at both $OD^{S}$ (590 nm) and $OD^{R}$ (800 nm). The normalised calculation for indigoidine is then calculated as:

$$\text{relative indigoidine} = OD^{S}_{B,+I} - \delta.OD^{R}$$

Background correction was achieved by first subtracting the mean of "culture medium alone" replicates from every value.

## 2.11 *In vitro* purified protein assays

### 2.11.1 Activation of BpsA by 4'-phosphopantetheine attachment

For pre-activation of BpsA, reactions were set up containing the following: 3.4 µM BpsA, 0.25 µM Sfp, 10 mM $MgCl_2$, 100 µM CoA and 50 mM sodium phosphate buffer pH 7.8. Reactions were incubated for 20-30 min at 30 °C.

### 2.11.2 Determination of kinetic parameters for BpsA

For determination of the kinetic parameters of BpsA with L-glutamine as the substrate, a triplicate 2-fold serial dilution of L-glutamine from 8 to 0.004 mM, in a final volume of 50 µl, was established in a 96-well microplate. 100 µl of reaction mix (5 mM ATP, 15 mM $MgCl_2$, and 75 mM Tris-Cl pH 7.8) was then added to each well. Reactions were initiated via the addition of 50 µl pre-activated BpsA mix to each well. The microplate was mixed at 1000 rpm for 10 s then $OD_{590}$ values were recorded every 8-10 s for 15-30 min using an EnSpire 2300 Multilabel Reader. Data was analysed in Microsoft Excel, with the slope function employed to determine the reaction velocity. The maximum velocity for each reaction, as determined by the steepest gradient, was analysed using GraphPad Prism® to derive the kinetic parameters as described by Owen et al. (Owen et al., 2011).

### 2.11.3 Preliminary assessment of PPTase activity

For initial assessment of PPTase activity levels, reactions were set up in triplicate two-fold serial dilutions of the PPTase to be tested, across 6 columns of a 96-well microplate. The initial wells in the series contained 50 µl of undiluted enzyme. Reaction mix (100 µl total, 100 mM Tris-Cl, pH 8.0, 8 mM L-glutamine, 5 mM ATP, 20 mM $MgCl_2$, 50 µM CoA) was then added to each well. Reactions were initiated via the addition of 50 µl of 3.2 µM BpsA to each well, after which plates were mixed

at 1000 rpm for 10 s. Reaction progress was measured via the recording of $OD_{590}$ values every 15-30 s for 30-90 min (depending on the activity of the PPTase).

### 2.11.4 Determination of kinetic parameters for PPTases via BpsA

Reactions were set up in a triplicate two-fold serial dilution from 100-0.049 µM CoA in 50 µl volumes in a 96-well microplate. 100 µl of reaction mix (100 mM Tris-Cl pH 8.0, 8 mM L-glutamine, 5 mM ATP, 20 mM $MgCl_2$, and 0.1-1 µM PPTase (depending on activity level of PPTase)) was added to each well. To start the reaction, 50 µl of 3.2 µM BpsA was added to each well. The plate was mixed at 1000 rpm for 10 s and $OD_{590}$ values were recorded every 15-30 s for 30-90 min using an EnSpire 2300 Multilabel Reader. Data was analysed in Microsoft Excel using the slope function to derive velocity values. The slope function was then applied to these velocity values, and the maximum value for each reaction taken as the maximum PPTase velocity. For lower activity enzymes, the slope function measured a larger number of data points. The maximum velocity for each reaction was analysed using GraphPad Prism® to derive kinetic parameters as described by Owen et al. (2011).

### 2.11.5 CP competition assay

A triplicate 2-fold serial dilution of carrier protein in a final volume of 50 µl was set up in a 96-well microplate. 50 µl of reaction mix (3 µM apo-BpsA, 0.75 µM CoA, 40 mM $MgCl_2$ and 400 mM Tris-Cl pH 8.0) was added to each well. Reactions were initiated via the addition of 50 µl of a PPTase solution (at a final concentration of 0.1-0.25 µM PPTase) to each well. Plates were mixed at 1000 rpm for 10 s and incubated for 15-30 min at 30 °C, allowing the PPTase reactions to proceed until completion. Following this, indigoidine synthesis was initiated via addition of 50 µl of a reaction mix containing 4 mM L-glutamine and 4 mM ATP. Plates were mixed at 1000 rpm for 10 s and $OD_{590}$ values were recorded every 15 s for 50 min using an EnSpire 2300 Multilabel Reader. Velocity values were derived using the Microsoft Excel slope function and these were converted into percentage of maximum velocity for a triplicate reaction. For generation of $IC_{50}$ curves, data from each experiment was analysed in GraphPad Prism® using four-parameter dose-response curves.

## 2.12 Screening of metagenome libraries

Two small insert environmental DNA (eDNA) libraries, each described in section 4.1.4., were screened during the course of this study. Each was transformed into electrocompetent EcoBlue1, EcoBlueP and EcoBlueE cells, with 50 ng of DNA used for each transformation. Following recovery

cells were mixed 1:1 with sterile 80% (v/v) glycerol, and 0.5 ml aliquots were dispensed and stored at -80 °C. When required for use a single aliquot was thawed on ice and serial dilutions plated on auto-induction agar plates containing 100 mM L-glutamine and appropriate antibiotics, in order to determine the optimal dilutions that would result in approximately 5,000 to 10,000 insert containing clones per plate. Libraries were stored as frozen transformation aliquots and found to remain viable in this state for at least 3 months.

For screening, cells were thawed on ice and plated using an optimal (empirically determined, as above) dilution on auto-induction agar plates containing 100 mM L-glutamine and appropriate antibiotics. After 16 h of incubation at 37 °C plates were then maintained at room temperature and monitored for pigment development. Hits were taken as colonies that developed a visible blue pigmentation. Where single colonies were unable to be recovered directly clones were picked as accurately as possible, resuspended in 100 µl GYT, and then streaked on to a new auto-induction agar plate. Isolated hits were picked from plates and grown in 3 ml LB cultures containing 0.2% glucose (w/v) and appropriate antibiotics. Glycerol stocks were made from 500 µl of overnight culture and plasmid DNA was isolated from the remainder of the culture via plasmid miniprep. The resulting DNA was transformed into chemically competent DH5α cells in order to separate the insert-containing pRSETB or pETDuet1 plasmids from the pCDFDuet1 plasmid bearing the co-expressed *bpsA* reporter, with cells being plated on to LB agar plates containing ampicillin only.

### 2.12.1 Analysis of eDNA hits

Sequencing was performed by Macrogen Inc. DNA samples and primers for sequencing were prepared according to the company's specifications of concentration and purity. Inserts were sequenced using the T7 promoter and T7 terminator primers for the pRSETB inserts or the DuetUP2 and T7 terminator primers for the pETDuet1 inserts, which anneal to plasmid sequences immediately upstream and downstream of the insert respectively. Internal primers for sequence extension by primer walking were then designed as required. Individual sequences were assembled into contiguous sequences using the CAP3 program (Huang and Madan, 1999). Sequences were then entered into the GENtle software (http://gentle.magnusmanske.de/) and ORFs within each eDNA insert were identified. These were then annotated by BLASTX searching and alignment (http://blast.ncbi.nlm.nih.gov/). The sequences of the eDNA inserts recovered in this study were submitted to GenBank under the continuous accession numbers JN886564 through to JN886585.

## 2.13 Directed evolution of EntD

### 2.13.1 Vector preparation

The pBAD vector was isolated as in section 2.6.7.2. For digestion, 16 µg plasmid DNA was heated to 70 °C for 20 min, to relieve supercoiling, and was digested using 50 U of each restriction enzyme (XhoI and HindIII) in a final volume of 400 µl, following the manufacturer's instructions for buffer composition and temperature. After 6 h incubation a further 20 U of each enzyme was added and reactions were incubated for an additional 16 h. Digests were then heat inactivated at 80 °C for 20 min and purified using a 20 µg capacity Zymo-Spin™ clean up column. Digested vector was eluted in 30 µl sterile dd$H_2O$ and the concentration of the digested vector was determined. A 200 ng sample was run on an agarose gel to ensure no degradation had occurred. The rest of the digestion was then aliquoted in 5 µl portions and stored at -20 °C until required.

### 2.13.2 Assessment of vector quality

It was important to assess vector quality before use in directed evolution library generation experiments. The *entD* gene was amplified using Phusion high fidelity polymerase and prepared for ligation as described in section 2.13.4. A ligation using a total of 500 ng DNA was set up as in section 2.13.5, using the Phusion amplified *entD* gene in place of the error prone PCR (epPCR) amplified *entD*. This ligation was then used to transform chemically competent BL21 ΔentD cells harbouring a pCDFDuet:*bpsA* plasmid. The transformation was plated onto LB agar plates containing 100 mM L-glutamine, 0.01% (w/v) L-arabinose, ampicillin and spectinomycin. Plates were incubated at 37 °C overnight and *bpsA* was induced as described in section 2.10.1. Plates were incubated at room temperature. Transformation and ligation efficiency were determined by comparing numbers of blue colonies *vs.* white colonies after 12-16 h. Blue colonies indicated a successful ligation, while white indicated a self-ligation or uncut vector. A successful ligation rate of 70% or higher was deemed sufficient for use in further experiments.

### 2.13.3 Error prone PCR

epPCR was carried out on *entD* using the GeneMorph® Mutazyme® II kit, according to the manufacturer's instructions. Template was prepared by amplification of *entD* using Phusion polymerase. A high error rate was chosen, which was achieved through using 100 ng purified PCR product as template in a 50 µl reaction using 30 amplification cycles. Prior to thermocyling, reactions were divided into equal amounts (12.5 µl) across 4 tubes to reduce likelihood of clonal

mutations in the final library. The size and quality of the amplicons was determined by running a 3 µl aliquot of total pooled reactions on an agarose gel.

### 2.13.4 Insert preparation

The 50 µl epPCR reactions were purified using 5 µg capacity Zymo-Spin™ columns, and eluted in 20 µl sterile $ddH_2O$. The concentration of each was determined by Nanodrop analysis of a 1 µl aliquot. The rest of the eluent was then digested using 30 U of each restriction enzyme, for 5 h, in a final volume of 50 µl. Buffer composition and temperature were as recommended by the manufacturer. The digest was then heat inactivated by incubation at 80 °C for 30 min.

### 2.13.5 Ligation and transformation

The total amount of DNA per ligation was 1- 1.5 µg, with a molar ratio of vector to insert of 1:6. Ligations were incubated at 22 °C for 1 h and then subsequently at 16 °C overnight. Following this, an additional 2 U of ligase was added and the reaction was incubated for a further 7 h. Ligations were purified using a 5 µg capacity Zymo-Spin™ column and eluted in 20 µl of $ddH_2O$. Ligations were used to transform electrocompetent EcoBlue1 cells. Electrocompetent cells were prepared and transformed as in sections 2.6.6.3 and 2.6.6.4.

### 2.13.6 First tier screening

For screening, cells were thawed on ice and plated on iron limited pigment production plates as described in section 2.5.1.5.1, and incubated at 37 °C for 72 h. After this time, expression of *bpsA* was induced with the addition of IPTG as described for LB agar plates in section 2.10.1. Plates were incubated at 25 °C and monitored for colour development. After 24 h hits were selected, on the basis that the colonies appeared to be of comparably larger size and normal morphological shape as well as having less pigmentation than surrounding colonies. These were streaked on LB agar plates containing 100 mM L-glutamine, 0.01% (w/v) L-arabinose and appropriate antibiotics, incubated at 37 °C overnight and induced by addition of IPTG as described in section 2.10.1. Plates were incubated at room temperature, and potential hits were taken as those that appeared to have a loss of pigment production.

#### 2.13.6.1 Conformation of ability to grow on iron restricted plates

Due to confounding satellite colonies, clones picked as potential hits in section 2.13.6 were confirmed to have retained the ability to activate EntF (the native CP from *E. coli* that *entD* was evolved to retain the ability to activate) by being re-plated individually on iron limited plates at a density of approximately 100 colonies per plate. Plates were incubated at 37 °C for 72 owenh. Clones that were able to grow were selected as positive hits from the first stage of screening.

#### 2.13.6.2 Screening potential hits in EcoBlueE

As an alternative method of confirming the retention of activity of potential hits with EntF, hits were tested for their ability to activate the EntF-CP substituted BpsA. The plasmids containing potential hits as identified in section 2.13.6 isolated via miniprep and then used to transform chemically competent EcoBlueE cells. These were plated on LB agar plates containing 100 mM L-glutamine, 0.01% (w/v) L-arabinose and appropriate antibiotics, and incubated overnight at 37 °C. Plates were induced with IPTG as in section 2.10.1, and incubated at room temperature to monitor pigment production. Those that produced pigment within the first 24 h were selected as hits.

### 2.13.7 Analysis of hits

Each hit found in sections 2.13.6.1 and 2.13.6.2 was picked and grown in 3 ml LB cultures containing 0.2% (w/v) glucose, ampicillin and spectinomycin. Glycerol stocks were made using 500 µl of the overnight culture and plasmid DNA was isolated from remaining cultures. The resulting DNA was transformed into chemically competent DH5α cells and in order to separate the variant *entD*-containing pBAD from the pCDFDuet1 plasmid containing the co-expressed *bpsA* reporter, with transformations plated onto LB agar plates containing ampicillin only. The resultant single plasmid containing cells were then grown overnight in 3 ml LB containing ampicillin, and plasmids were isolated. These were then used to transform both EcoBlue1 and EcoBlueE chemically competent cells. The ability of each of these to produce indigoidine was tested as in section 2.10.1, using LB agar plates containing 100 mM L-glutamine, 0.01% (w/v) L-arabinose and appropriate antibiotics. Evolved variants of *entD* that exhibited indigoidine production in EcoBlueE and less or none in EcoBlue1 were sent for sequence analysis.

#### 2.13.7.1 Quantification of indigoidine production for evolved variants

Variants of *entD* in each EcoBlue1 and EcoBlueE were assessed for ability to produce indigoidine as in section 2.10.2, with the amendment that the medium also contained 0.2% (w/v) L-arabinose.

# Chapter 3: Using BpsA to probe Carrier Protein domain interactions

## 3.1 Introduction

Understanding the interactions between domains and modules within NRPS enzymes is a key requirement to understanding how these enzymes function. It is also an important first step in being able to swap modules or domains between different NRPS enzymes, with the ultimate goal of creating entirely new bioactive products. The work in this chapter builds on previous work that was performed in our lab by (then PhD student) Dr Jeremy Owen, and focuses on the interactions between CP domains and their downstream partner domains. CP domains are typically located upstream from either condensation (C) domains (in elongation modules) or thioesterase (TE) domains (in terminal modules). In this previous work, seven different CP domains were substituted into BpsA, then subsequently random mutagenesis was employed to improve the activity of certain swapped CP domains (Owen, 2010). One observation was that CP domains which are located upstream of a TE domain in their native context were more likely to result in a functional enzyme when swapped into BpsA. The native BpsA CP domain is itself located upstream of a TE domain, suggesting these CP domains may have evolved certain features that enable them to interact with TE domains preferentially as opposed to C domains or other downstream domains. Directed evolution to improve BpsA CP substitution constructs provided insight into which residues were particularly important, establishing a model of CP domain interaction. The work described in this chapter focuses on two of the key positions that were found to be important. These positions are the +4 and +24 residues, relative to the invariant serine residue found in every CP domain. This work, as well as aiding in our general understanding of NRPS interactions, led to the creation of evolved BpsA variants which themselves have various applications, including expanding the utility of a metagenomic screening system, the focus of chapter 4.

### 3.1.1 CP interactions within NRPS enzymes

Directly swapping one NRPS domain or module for another generally leads to enzymes with drastically reduced activity, if any activity at all (Calcott and Ackerley, 2014). The desire to reliably manipulate multi-modular biosynthetic enzymes such as NRPSs means much work has been done on understanding the interactions between different domains and modules, in order to improve

efficiency and activity of potential module or domain substitutions. The ultimate goal of these efforts is to selectively modify the existing peptide products made by these enzymes, or to create entirely new products. CP domains are important to the overall activity of the enzyme as they have to interact with a large number of other enzymes and domains. This includes upstream A domains and downstream C, E or TE domains as well as the PPTase enzymes that activate them. CP domains are required to be highly flexible, with the attached phosphopantetheinyl arm transporting the growing peptide product from one active site to the next (Mercer and Burkart, 2007). The inherent flexibility of CPs has made structural studies difficult, with limited crystal and NMR structures solved (Lohman et al., 2014). One structural study of the type I PCP TycC3 (Koglin et al., 2006) suggested that the CP domain had three conformational states. Two states were exclusive to either the apo form (A state) or holo form (H state) of the CP domain. The third state (A/H) appeared to be shared by both the apo and holo forms. Further work however has shown that the A and H states are likely to have been artefacts (perhaps as a result of excising the domain from its native context) and the A/H state may be the only conformational state for a CP domain regardless of whether it is in the apo or holo form (Haslinger et al., 2015; Lohman et al., 2014; Tufar et al., 2014). It has recently been suggested that the flexibility of the CP domains comes from subtle rearrangements rather than larger conformational changes (Haslinger et al., 2015).

Mutagenesis studies are an important alternative tool to elucidate key residues involved in domain interactions. Several studies have used alanine scanning mutagenesis to look at the interaction between the CP and neighbouring domains. This technique examines which residues can tolerate a sequence change to an alanine, while still allowing the original protein to retain activity. When a residue is resistant to change this implies that that particular position is important to the protein's function. One study (Lai et al., 2006) used scanning alanine mutagenesis to examine the EntB CP domain within the enterobactin complex. The downstream interaction partner is the C domain of the EntF module (an *in trans* protein interaction). Helix II and helix III of the CP were mutated, as they had previously been shown to be generally important in these interdomain interactions (Lai et al., 2006). A selection was applied and surviving clones were sequenced. For each position, a ratio comparing how many times the WT residue was found as opposed to an Ala residue was determined. A higher ratio indicated Ala was not tolerated well at this position and the WT residue was said to be conserved. The most conserved WT residues were those at the +19 and +23 positions (in relation to the invariant serine residue), with the +4 and +24 positions being the next most conserved. These residues were considered to be important in protein function and interaction with the downstream EntF C domain. This same group next used scanning alanine mutagenesis to examine the EntF CP domain (Zhou et al., 2006), and found

substitutions to Ala at residues +21 and +24 to be the most detrimental to protein function, concluding that these residues were important in an interaction surface between the CP domain and the downstream TE domain. These two studies looked at downstream interaction of the CP domain with different domains (a C domain and a TE domain) and found residues in similar locations (all on helix III of the CP domain, with the +24 residue found in both cases), which may indicate that this region is particularly important for downstream interactions.

### 3.1.2 BpsA as a probe for investigating CP interactions

Previous work in our lab used BpsA as a tool for probing the interaction between the CP and other domains within the NRPS enzyme, in particular the TE domain (Owen, 2010). BpsA is an ideal NRPS for these experiments, given its relatively small size (for an NRPS) and its readily detectable product. The small size of BpsA means it can be relatively easily genetically manipulated, and addition of only one substrate (L-glutamine) is required to form the blue pigment indigoidine. Also, BpsA contains no C domain, which simplifies its use as a model NRPS as there are no confounding interactions with an upstream C domain. CP domains from seven foreign NRPS enzymes were substituted into BpsA, most of which initially had very low activity (if any), not unusual for NRPS domain substitution experiments (Calcott and Ackerley, 2014). A selection of these CP-substituted variants were improved via random mutagenesis, with superior variants recognised on the basis of improved indigoidine production. Although only a small sample size of seven CP domains was examined, the three CP domains that were directly upstream from TE domains in their native setting all retained some initial activity when substituted, while those four CP domains that were upstream from a C domain in their native setting did not. This led to the conclusion that CP domains that were upstream of TE domains in their native enzyme may have naturally evolved to preferentially interact with the TE domain over other possible downstream domains, and hence were better able to function when substituted immediately upstream of the TE domain in BpsA.

Consistent with this, the CP domain from the first module of the PvdD NRPS from *P. aeruginosa* (PvdDCP1), which is located immediately upstream from a C domain in its native context, gave no activity when substituted into BpsA. The CP domain from the second module (PvdDCP2) on the other hand (upstream from a TE domain in its native context) did show activity (Figure 3.1). These two PvdD CP domains are of interest as the two PvdD modules have clearly risen from a gene duplication event, sharing a high sequence identity and both encoding for the same amino acid residue in the final NRPS product (Ackerley et al., 2003). While the A domains of PvdD share 99% amino acid identity and the C domains 80%, the CP domains of these two modules are more diverse at 72% identity. The divergence in sequence identity in the CP domains is almost entirely

downstream of the invariant serine, with 97% identity conserved upstream of this site and only 20% identity downstream of this site. This likely reflects that these domains each have to interact with different downstream partners (the C domain for PvdDCP1 and the TE domain for PvdDCP2), and so any differences in sequence may reflect key residues involved in these interactions. In order to pinpoint residues which may be involved in the specialised downstream interaction, the BpsA PvdDCP1 substitution was the subject of a directed evolution effort to improve function via random mutagenesis. The sequence of the improved mutants highlighted residues that seemed to be important in restoring the interaction between the CP and TE domain. The variant with the highest activity following the first round of evolution was "3kf0", which had residue substitutions to Ile at both the +4 and +24 positions (in WT PvdDCP1 the wild type residues are Thr and Phe respectively). Based on modelling, it was hypothesised that these two positions were involved in the interaction with the hydrophobic TE interface (Owen, 2010).

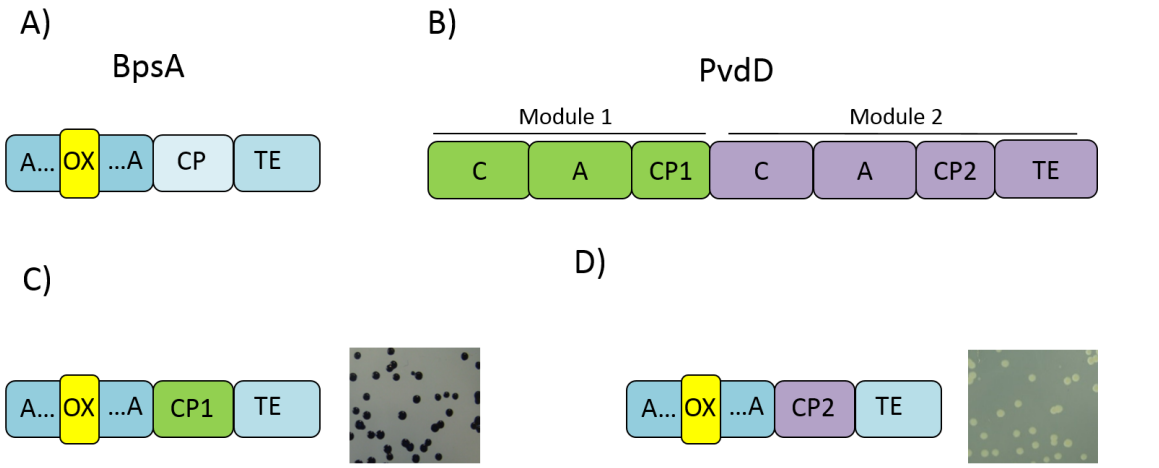



**Figure 3.1 PvdD CP substitutions into BpsA.** A) Module structure of BpsA. B) Module structure of PvdD. C) PvdDCP1 substitution results in a functional BpsA enzyme and indigoidine production. D) PvdDCP2 substitution results in a non-functional BpsA enzyme with no indigoidine production.

One conclusion from this previous work was that a hydrophobic amino acid at the +4 position was important for the interaction between the CP and TE domains, although due to lack of a solved structure for BpsA it was unclear exactly why. The +4 position was commonly mutated to a codon encoding a hydrophobic residue in evolved variants of the PvdDCP1 substitution, with the equivalent change frequently observed in a different CP domain substitution construct that was also evolved, DhbF. The third CP (EntF CP domain) substitution that was evolved already had a hydrophobic residue (Met) at the +4 position, and this residue was not altered in any of the evolved variants that were recovered (Owen, 2010).

The importance of a hydrophobic residue at the +4 position was also supported by CP domain alignments (Owen, 2010). Fifty four CP domains (with the corresponding downstream domains of each split between C, E or TE domains) from the NRPS of various *Pseudomonas* species were aligned. In the case of the 20 CP domains located upstream from TE domains, a hydrophobic residue was always present at the +4 position. For the 17 CP domains upstream from C domains, residues at this position were variable. For example PvdDCP1, upstream of a C domain in its native setting, contains a polar Thr at this position; whereas PvdDCP2, upstream of a TE domain, contains a hydrophobic Leu. It was hypothesised that one reason the initial PvdDCP1 substitution had failed to show any activity was because of the presence of the non-hydrophobic Thr (Owen, 2010). However, subsequent alignments (a total 103 CP domains sourced from 10 non-*Pseudomonas* bacterial species, including 15 CP domains upstream from TE domains) performed by Dr Mark Calcott identified four (out of 15) CP domains that were upstream from TE domains and yet had a Thr at the +4 position (Calcott, 2014). The remaining 11 CP domains did all have hydrophobic residues at this position. This chapter addresses the question of whether these four CP domains might still be able to function within BpsA, to see if the +4 Thr would be prohibitive to effective function post-substitution, or whether they are still optimised for interaction with a downstream TE domain via some alternative evolutionary solution.

From Dr Owen's work it seemed clear that both the +4 and +24 positions were important, with hydrophobic residues being particularly key, however random mutagenesis via error prone PCR is limited in terms of the range of mutations it can reasonably access. Random changes can be introduced at any point in a gene but in practice it is unlikely that two or more changes will occur within the same codon, resulting in a limited number of amino acids are possible at each position (Hermes et al., 1990). In the case of PvdDCP1, this meant that the +4 position was effectively limited to just five possible amino acid changes (Ile, Asn, Pro, Ser and Ala) and the +24 position was limited to six possible changes (Leu, Ser, Tyr, Cys, Ile and Val). With these residues identified as being important, we wanted to take a closer look at which amino acid changes could be tolerated at this position and their relative effects on activity. The second aim of this chapter was to perform random codon mutagenesis of the +4 and +24 positions in PvdDCP1, both alone and in combination, to identify substitutions at these positions that substantially improved activity of the PvdDCP1 BpsA swap.

### 3.1.3 Aims

- Substitute into BpsA the four CP domains, found upstream of TE domains in their native context, which have a Thr at the +4 position and test for activity.

- Identify all amino acid residues at the +4 and +24 positions of PvdDCP1 that are able to promote effective function when substituted into BpsA.
- Quantify the contributions of each of the tolerated +4 and/or +24 mutations, identifying any additive or synergistic effects between the two.
- Kinetically characterise the top BpsA-PvdDCP1 substitution variants.

# 3.2 Results

## 3.2.1 Activity of CP domain BpsA substitutions having Thr at the +4 position

The four CP domains identified in alignments by Dr Mark Calcott (Table 3.1), which were all immediately upstream of a TE domain in their native enzymes and yet contained a non-hydrophobic Thr at their +4 positions, were investigated here. These four CP domains were selected to be substituted into BpsA to see which conclusion (stated in section 3.2.1) would hold true - whether these CPs located upstream of TE domains in their native context would function when substituted into BpsA, or whether CP domains with a Thr at the +4 position are not able to function when substituted into BpsA. Each of the four was cloned into *bpsA* using silent restriction sites. The ability of each to function was first assessed qualitatively using an agar plate based screen, followed up by a quantitative liquid assay for those that were shown to have activity.

**Table 3.1 Origin of carrier proteins immediately upstream of TE domains that also have a Thr at +4 position**

| Gene name | Species | Accession number |
|---|---|---|
| ACMSIII | *Streptomyces anulatus* | ADG27359.1 |
| GrsB | *Brevibacillus brevis* | CAA43838.1 |
| TycC | *Brevibacillus brevis* | AAC45930.1 |
| Pris3 | *Streptomyces pristinaespiralis* | O07944 |

### 3.2.1.1 Solid media activity test

Plasmids bearing *bpsA* variants containing each of the four CP domains were used to transform a *sfp* containing BL21 strain and the resulting transformants were plated out on L-glutamine containing agar plates. Due to the toxicity of indigoidine, expression of *bpsA* was delayed until after initial colony formation. After an overnight incubation, the agar slab was scooped out of the plate, set on a sterile lid and IPTG solution was spread evenly on the bottom of the plate. Agar slabs were then replaced on top and diffusion of the IPTG through the agar resulted in development of blue pigmentation in transformants that possessed an active BpsA variant. A

negative control of BpsA containing a deleted CP domain was included, as well as a control with the WT *PvdDCP1* substituted in, which contained Thr at the +4 position and had already been shown not to function (Owen, 2010). Neither of the negative controls developed any blue colour. In contrast, each of the four substitutions showed at least minimal activity, with the number of days taken for colonies to turn blue listed in Table 3.2. Figure 3.2 shows a picture of each of the four plates after 7 days, with different degrees of blue pigmentation visible. Two of the CP domain substitute variants (ACMSIII and TycC) gave a stronger activity than the other two, with indigoidine production visible after two and three days respectively. The other two variants only developed a faint colour after six days, indicating extremely low levels of activity. All four of the CP-substitution variants tested were capable of producing indigoidine to some degree.

**Table 3.2 Time to develop blue pigmentation for different CP domains**

| CP domain origin gene | Time for colour to develop on solid media (days) |
|---|---|
| ACMSIII | 2 |
| GrsB | 6 |
| TycC | 3 |
| Pris3 | 6 |

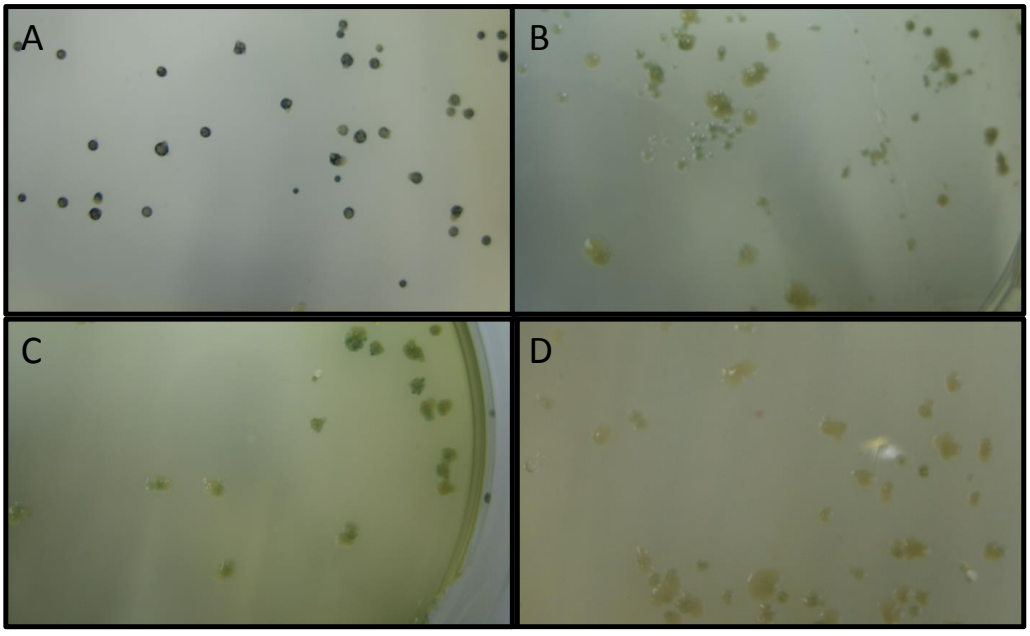



**Figure 3.2. Initial assessment of BpsA CP-substitution activity on agar plates.** The four CP-substitution strains were grown on L-glutamine containing plates and induced with IPTG. Photos taken after seven days of incubation at room temperature A) ACMS III B) GrsB C) TycC D) Pris3.

#### 3.2.1.2 Liquid media activity tests

The ability of strains expressing each of the four BpsA CP-substitution variants to produce indigoidine was tested quantitatively in liquid culture assays. The relative concentrations of indigoidine were determined using the method described by (Beer et al., 2014) and described in section 2.10.2. Cultures were measured at set time points at absorbance's of both 590 and 800 nm. In these assays, indigoidine is measured at an absorbance of 590 nm (Takahashi et al., 2007) while absorbance at 800 nm is used to normalise for cell density. This method was used in preference to a previous method where 96 well plates were spun down to remove cells and absorbance of indigoidine alone was measured at 590 nm (Owen, 2010). The dual absorbance reading method allowed for the same culture to be measured over multiple time points, as well as generally giving more consistent results in pilot studies (not shown).

The results in Figure 3.3 show indigoidine production as measured after 30 h incubation at 18 °C. The ASMSIII substitution was the only CP-substituted BpsA displaying any activity, consistent with this substitution being the most active in the solid media activity test (Figure 3.2). The other three substitutions had no detectable activity above the level of a negative control in the liquid culture assay. This was not surprising as it was generally found that if a BpsA variant had poor enough activity that it did not turn its host colony blue on an agar plate assay after 48 h, it was unlikely to show any activity in the liquid culture assay over a measurable time period.

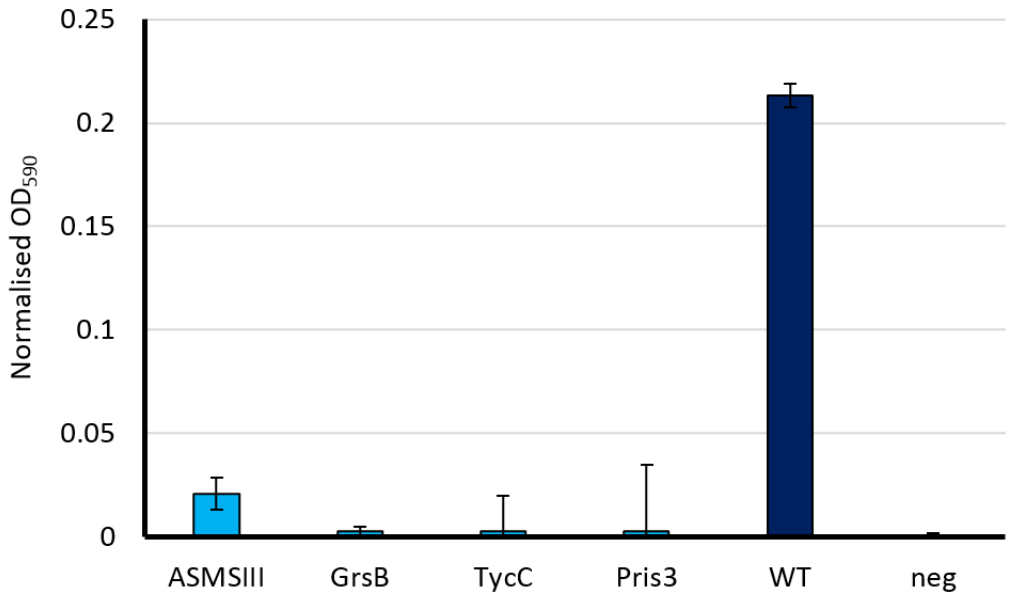



**Figure 3.3. Quantitative assessment of the relative indigoidine synthesis activities of +4 Thr CP-substitution strains.** +4 Thr CP-substitution strains were set up in quadruplicate in a 96-well microplate and incubated at 18 °C. After 30 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. WT BpsA is included as a positive control (dark blue bar) and BpsA with a deleted CP domain is included as a negative control (red bar ("neg")). Error bars are +/- one standard deviation.

Even though the activity of three of the four CP-substitutions tested was undetectable in the liquid based assay, the fact that each of the four CP-substitutions were all at least minimally functional in the agar plate assay (Table 3.2) supports the hypothesis that CP domains found upstream from TE domains (as opposed to upstream from C domains) in their native context are able to function more easily when substituted upstream from other TE domains. The fact non-hydrophobic Thr at the +4 position is not strictly necessary for the function of a CP-substitution with BpsA. However, the extremely low levels of activity of each of these substitutions implies that this +4 position may still be important, and this non-hydrophobic residue may be severely impairing the substitutions.

### 3.2.2 Identification of amino acid residues at the +4 position of PvdDCP1 that result in partial restoration of indigoidine biosynthesis

In previous work the +4 position in the CP domain was identified as being important in improving activity of CP substitutions within BpsA, with many evolved variants gaining a hydrophobic amino acid at this position (Owen, 2010). The work described in Section 3.2.1 showed that a hydrophobic residue at the +4 position was not strictly necessary for function, with four CP domains with +4 Thr able to function when substituted into BpsA, albeit minimally in most cases. In contrast, when

PvdDCP1 had been substituted into BpsA it resulted in a completely non-functional enzyme. Directed evolution was previously employed to improve the activity of this enzyme, and various activity-enhancing mutations were found (Owen, 2010). All but one of the mutations at the +4 position was Ile, with one Ala found in conjunction with another mutation. Given the practical limitations of error prone PCR (as discussed in section 3.1.2), these mutations found were two out of the five possible mutations that were practically available at this position, insufficient to confirm the pattern that hydrophobic residues were required.

In order to determine which of the 20 amino acids at the +4 position of PvdDCP1 would yield a functional CP-domain substituted BpsA enzyme, a mutant library for the PvdDCP1 domain was generated, with an NNK codon at the +4 position. For the NNK codon, N stands for any base and K can be either a thymine or guanine. This combination results in 32 of the possible 64 codons being represented, collectively specifying all 20 possible amino acids. Primerless PCR was performed with Phusion® High-Fidelity polymerase using an ultramer (178 bp oligonucleotide) containing the degenerate codon NNK at the +4 position together with a reverse primer, thereby creating a double stranded product of the *PvdDCP1* gene with the NNK codon at the +4 position. This was then amplified using BioMix™ Red and ligated into the BpsA substitution construct that had a deleted CP domain. This PvdDCP1 +4 NNK library was then transformed into BL21 *ΔentD* cells that contained an activating PPTase gene, s*fp*. To screen for active mutants the library was then plated on L-glutamine containing agar plates and *bpsA* expression was induced after initial overnight growth of colonies as in section 3.2.1.1. Plates were monitored over a period of four days and any blue hits were selected. The timing of discovery of each of the hits was recorded, ranging from a few hours to four days following induction. This gave the first indication of differences in activity levels between different mutants.

Prior to selection, eight randomly selected clones from the library were sequenced to confirm that a reasonable diversity was present. With six out of the eight hits being unique clones, diversity was confirmed. In total 1-2 x $10^3$ clones were screened, with approximately 47% of colonies giving a positive blue result. As only one site was varied, there was likely a lot of redundancy and only 165 hits were ultimately isolated for further analysis. From these, a total of 51 clones were sequenced (with varying activity levels based on time to turn blue in initial screen). Unexpectedly, sequencing revealed that 20 of these clones contained additional mutations, at locations other than the targeted +4 position. These additional mutations perhaps contributed to the unusually high rate (47%) of hits that were seen. Subsequent to this, the experiment was repeated only amplifying the PvdDCP1 +4 NNK gene with a high fidelity polymerase, Phusion™, compared to the

previous PCR system Biomix Red™. A further 1000 colonies were screened, this time yielding just a 15% blue hit rate. Seventeen of these were sequenced, and while there were still three found that had additional mutations present, the majority were mutated at the +4 position alone.

Overall six unique clones were found which had mutations solely at the +4 position. These amino acids were Ile, Phe, Leu, Tyr, Met and Ala. The WT residue, Thr, was also found amongst the hits sequenced, but later found to have no activity when re-streaked on L-glutamine containing agar plates. It was noted during this process that false positives were more likely to be found after several days of incubation on an agar plate. After this amount of time some of the darker colonies leach indigoidine into the surrounding agar, making colonies producing their own indigoidine harder to identify. The time for each clone to produce indigoidine pigment was tested on agar plates, as in section 3.2.1.1. This gave an indication of activity level of each clone, as well as allowing identification of any false positives. The time taken for each of the active +4 variant substitution strains to turn blue is summarised in Table 3.3.

**Table 3.3 Amino acid residues at the +4 position that resulted in improved BpsA activity**

| Amino acid at +4 | Time for colour to develop on solid media (days) |
|---|---|
| Alanine | 2 |
| Leucine | 2 |
| Isoleucine | 2 |
| Phenylalanine | 3 |
| Tyrosine | 3 |
| Methionine | 4 |

### 3.2.2.1 Quantitative assessment of hit activity

In the previous section, six amino acid changes at the +4 position in PvdDCP1 were identified qualitatively as having the ability to generate functional BpsA variants, i.e. yielding blue coloured colonies on agar plates. To quantitatively assess the effect of the different mutations at this position, each unique clone capable of giving rise to the blue colouration on plates was then subjected to an assay in liquid media, where the amount of indigoidine was calculated through absorbance readings. The results for single mutations at the +4 position can be seen in Figure 3.4. The variant strains that displayed the greatest increase in indigoidine production were those with Ile and Phe at the +4 position, followed by Leu. All of the variants were much lower than wild type activity levels. The other three variants, with Tyr, Met and Ala at the +4 position, did not show any more activity than the negative control. The relative levels of activity do not correlate with those seen in the agar plate assays (Table 3.3), where the Ala substituted variant appeared to have equal

activity to Leu and Ile substituted variants, and the Phe substituted variant appeared to have the same level of activity as the Tyr substituted variant.

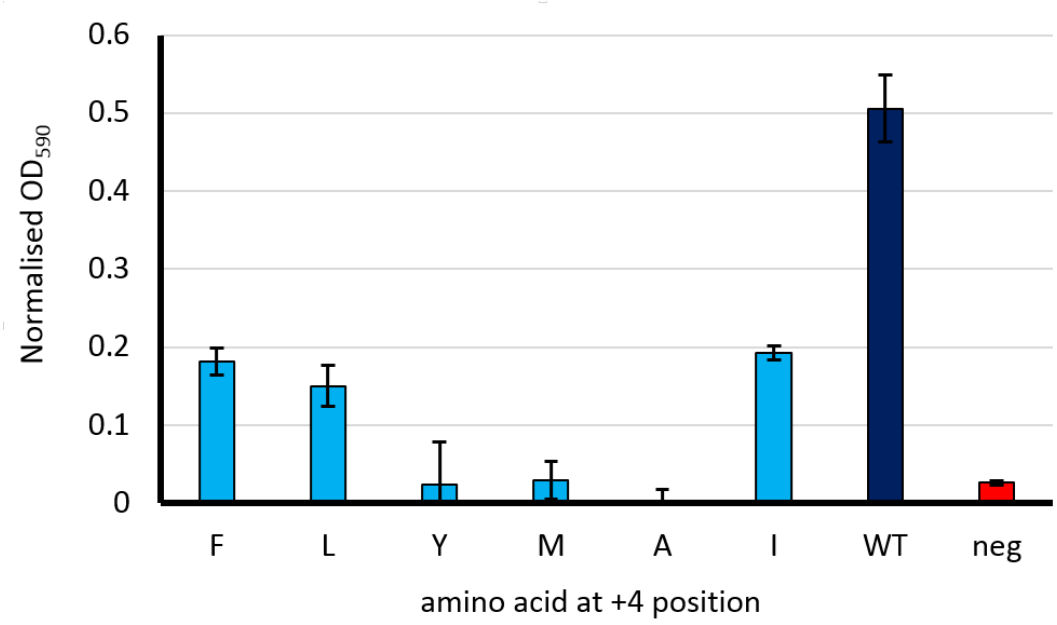



**Figure 3.4. Quantitative assessment of the relative indigoidine synthesis activities of active PvdDCP1 +4 mutation substitution strains.** PvdDCP1 +4 variant strains were set up in quadruplicate technical replicates in a 96wp and incubated at 18 °C. After 48 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. WT BpsA is included as a positive control (dark blue bar) and BpsA with a deleted CP domain is included as a negative control (red bar). Error bars are +/- one standard deviation.

### 3.2.3 Identification of amino acid residues at the +24 position that result in partial restoration of indigoidine biosynthesis

As described previously for the +4 position in section 3.2.2, a +24 NNK library for the *pvdDCP1* gene was generated and cloned into the CP domain substitution BpsA vector. This library was transformed into a strain containing the activating PPTase gene *sfp*, plated and screened for activity on L-glutamine containing agar plates. After induction with IPTG, plates were incubated at room temperature and blue colony "hits" were selected. Approximately 500 colonies were screened in total, with 17% of these being positive blue hits. Of these, 32 clones were sequenced, each of which only contained a change at the +24 site, with no additional mutations being found. Twelve unique clones were recovered, and each was plated on L-glutamine containing plates and induced with IPTG. The time taken for colonies to turn blue was recorded for each. The amino acids identified at this position together with the timing for blue colour to develop are summarised in Table 3.4. The time for colour to develop mostly corresponded with the day that hits were observed to appear in the original screen.

**Table 3.4 Residues at +24 position resulting in improved BpsA activity**

| +24 | Time for colour to develop on solid media (days) |
|---|---|
| L | 1 |
| V | 1 |
| M | 1 |
| I | 2 |
| S | 2 |
| C | 2 |
| Q | 2 |
| T | 2 |
| A | 3 |
| G | 3 |
| Y | 4 |
| P | 4 |

#### 3.2.3.1 Quantitative assessment of hit activity

To quantitatively assess the effect of the different mutations at the +24 position, each unique hit was then subjected to an assay in liquid media, where the amount of indigoidine was calculated via absorbance readings. Only the top performing clones were tested, i.e. those that turned blue within 48 h on an agar plate (reasoning that anything beyond this was unlikely to show any activity in the liquid assay). Figure 3.5 illustrates the relative activity levels of these mutants. The apparent best clone had Ile at the +24 position, followed by Leu, Val and Met. All of these top residues are hydrophobic. Overall these results are in agreement with the conclusion drawn in the previous section, i.e. that changes at just one position, if it is the right position, can cause dramatic changes in overall activity. Based on these results it was next sought to determine whether changing two residues at the same time could achieve additive or even synergistic effects.

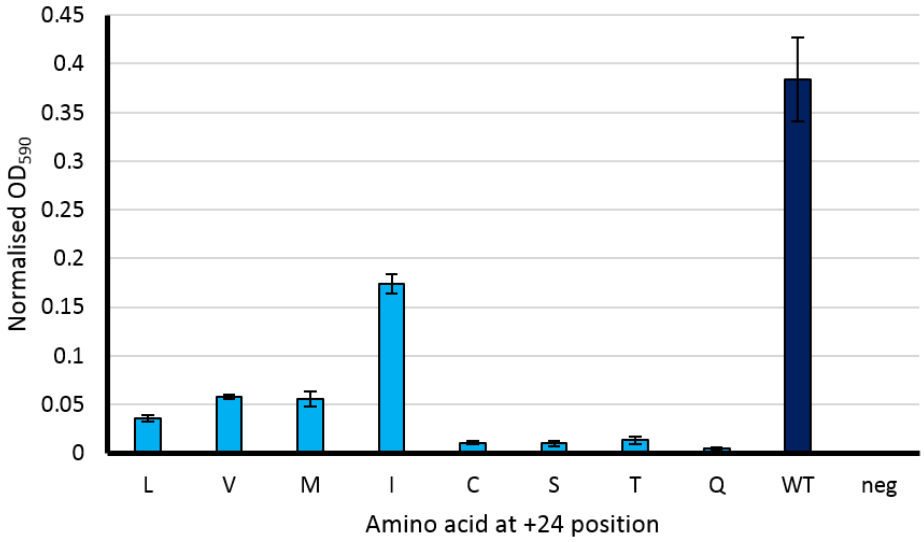



**Figure 3.5. Quantitative assessment of the relative indigoidine synthesis activities of active PvdDCP1 +24 mutation substitution strains.** PvdDCP1 +24 variant strains were set up in quadruplicate in a 96wp and incubated at 18 °C. After 48 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. WT BpsA is included as a positive control (dark blue bar) and BpsA with no CP present is included as a negative control (red bar ("neg")). Error bars are +/- one standard deviation.

### 3.2.4 Identification of amino acid combinations at the +4 and +24 positions that restore indigoidine biosynthesis

Sections 3.2.2 and 3.2.3 show that changing either the +4 or +24 position can in isolation have a dramatic effect on the activity of a protein, in some cases transforming a completely inactive enzyme into a relatively active one. Next the effect of combining mutations at these positions was tested. A degenerate *pvdDCP1* gene construct that contained NNK codons at both the +4 and +24 sites was cloned into the CP-substitution BpsA vector. This library was then screened in the same manner as previously. Over 700 clones were screened, with 152 blue clones recovered, an approximately 21% hit rate (higher than observed for mutation of either the +4 or +24 residue in isolation). A total of 39 hits were sequenced. Interestingly several hits were found that contained the PvdDCP1 WT residue (Thr) at the +4 position, making them equivalent to variants found in section 3.2.3 that had a mutation solely at the +24 position. In contrast, no hits were found with the PvdDCP1 WT residue, Phe, at the +24 position. The sequences of hits and the time taken for each to yield blue colonies when subsequently tested in *E. coli* on solid media is summarised in Table 3.5.

**Table 3.5 Sequence and activity level of clones with mutations at +4 and +24 residues**

| +4 | +24 | Time for colour to develop on solid media (days) |
|---|---|---|
| M | P | 1 |
| V | I | 1 |
| M | Y | 1 |
| L | T | 1 |
| M | L | 1 |
| V | A | 1 |
| T | L | 1 |
| T | V | 1 |
| C | M | 1 |
| L | P | 1 |
| L | M | 1 |
| V | L | 1 |
| C | V | 1 |
| T | M | 1 |
| T | A | 2 |
| P | M | 2 |
| E | L | 2 |
| G | I | 2 |
| D | L | 2 |
| S | A | 2 |
| T | C | 2 |
| F | P | 2 |
| L | N | 2 |
| G | L | 2 |
| H | A | 2 |
| P | C | 2 |
| H | S | 2 |
| A | S | 2 |
| T | N | 3 |
| G | T | 4 |

### 3.2.4.1 Quantitative assessment of hit activity

To quantitatively assess the effect of the combined mutations at the +4 and +24 positions, each unique clone capable of giving rise to blue colony colouration on plates was then subjected to an assay in liquid media, with the amount of indigoidine calculated through absorbance readings. Again only the top activity mutants were chosen, in this case the clones that had turned blue within 24 h in the agar plate assay (restricting the time period to just one day due to the overall larger number of clones found). The results for this are illustrated in Figure 3.6. Due to the higher average activity of the clones, activity levels were able to be clearly distinguished after 24 h. The clones that displayed the highest activity levels were those with +4/+24 combinations of L/M, M/L and V/I. Each of these residues is strongly hydrophobic. Most of the top mutations, for instance L/M and M/L, had each of the combined mutations appearing individually in the +4 or +24 screens.

There is evidence to suggest some mutations may be synergistic, for example, the +4 L and +24 M variants individually were able to restore only a moderate degree of indigoidine synthesis after 48 h, but when these two mutations were combined in the top performing variant in this assay (L/M, Figure 3.6), indigoidine synthesis approached WT levels after just 24 h. The degree to which the two positions interact is explored further in the next section.

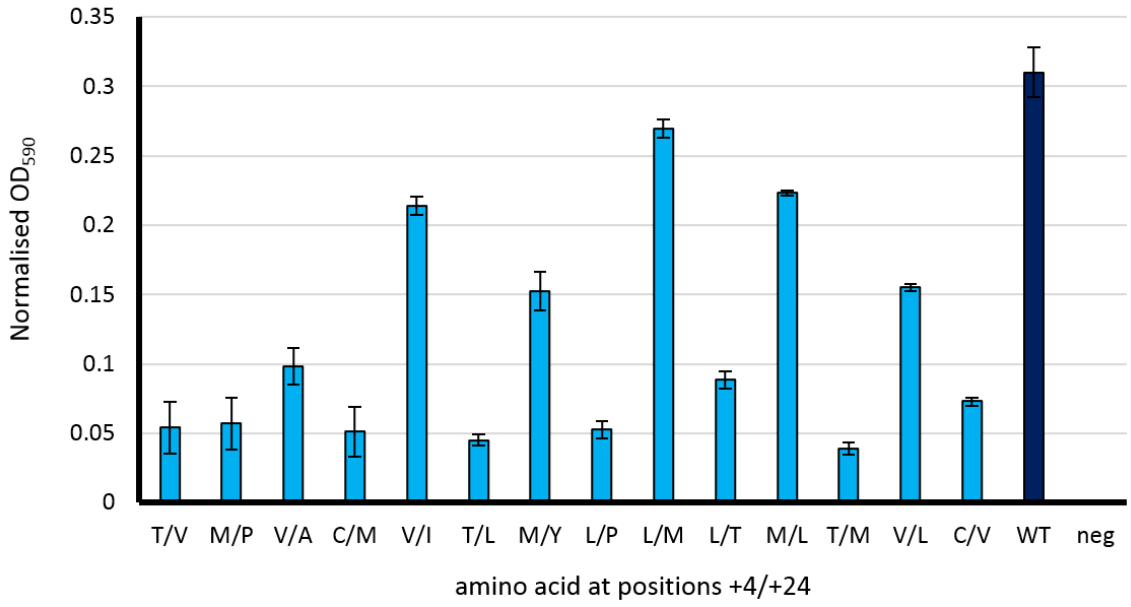



**Figure 3.6 Quantitative assessment of the relative indigoidine synthesis activities of active PvdDCP1 +4/+24 substitution strains.** Top PvdDCP1 +4/+24 substitution variant strains were set up in quadruplicate in a 96wp and incubated at 18 °C. After 24 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. WT BpsA is included as a positive control (dark blue bar) and BpsA with a deleted CP domain is included as a negative control ("neg"). Error bars are +/- one standard deviation.

### 3.2.5 Combined effects of mutations

Many of the amino acid changes that were found in the combined +4/+24 clones had also been found individually, allowing a comparison of the effects when a particular mutation was found either alone or in combination with another, and the identification of additive, synergistic or antagonistic mutations. In five separate cases the data were available to compare the +4 and +24 mutations both together and for each in isolation (Figure 3.7). Clones analysed included +4/+24 I/I, the top clone from Dr Owen's first round of directed evolution, as each of the mutations (+4 Ile and +24 Ile) was recovered individually during the course of this study. The relative activity level of each can be seen in Figure 3.7, with the effects from combining these particular mutations all appearing to be either additive or synergistic. In the case of the +4 Ala and +24 Ser substituted

variants the effect may have been additive, but overall the activity from these three clones was too small to draw a strong conclusion. The rest of the clones for which all +4/+24 combinations were available- +4/+24 L/T, M/L, L/M and I/I – each looked to be synergistic, with the activity of the +4/+24 clones being substantially greater than the activity of the individual clones combined. These included the two top mutants found in this screen, M/L and L/M as well as I/I, the top mutant found in previous first tier screening.

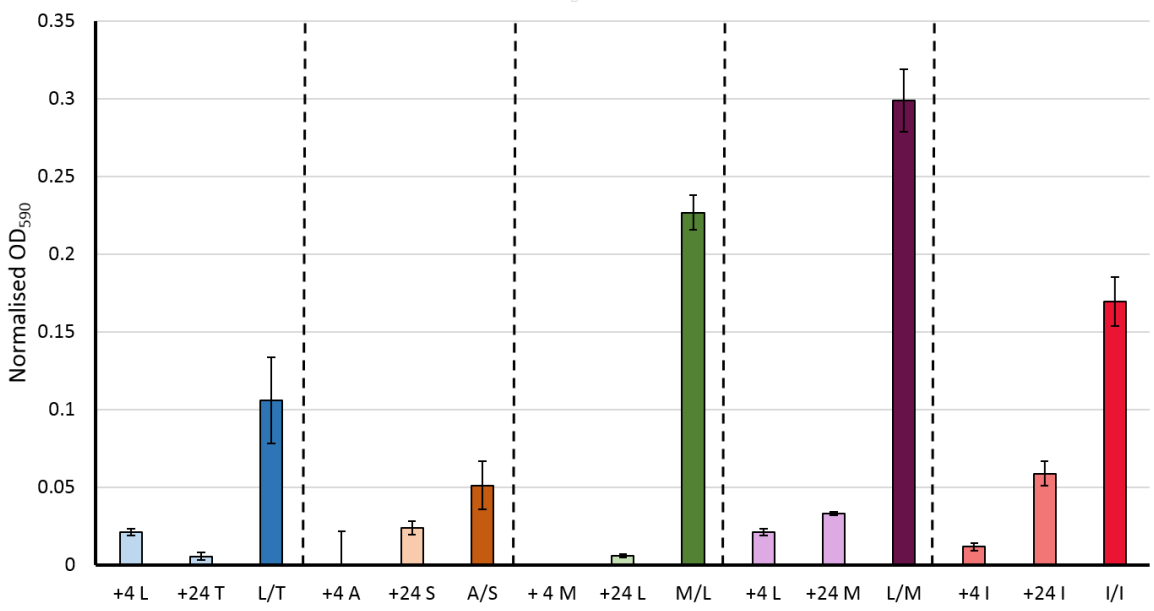



**Figure 3.7. Quantitative assessment of the relative indigoidine synthesis activities of PvdDCP1 substitution strains for which the same +4 and +24 mutations had been found both alone and in combination.** PvdDCP1 variant substitution strains were set up in quadruplicate in a 96wp and incubated at 18 °C. After 48 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. Error bars are +/- one standard deviation.

Another example of synergy seen was when a mutation did not appear to have enough activity on its own to even be identified in the +4 screen, yet acted synergistically with a second mutation at +24. For example Val at the +4 position on its own appeared to have no effect (based on the fact it was never recovered from the NNK +4 library), but synergistic effects were seen in combination with Ala, Leu or Ile at the +24 position (Figure 3.8). In fact the +4/+24 V/I clone was the third top mutant overall.

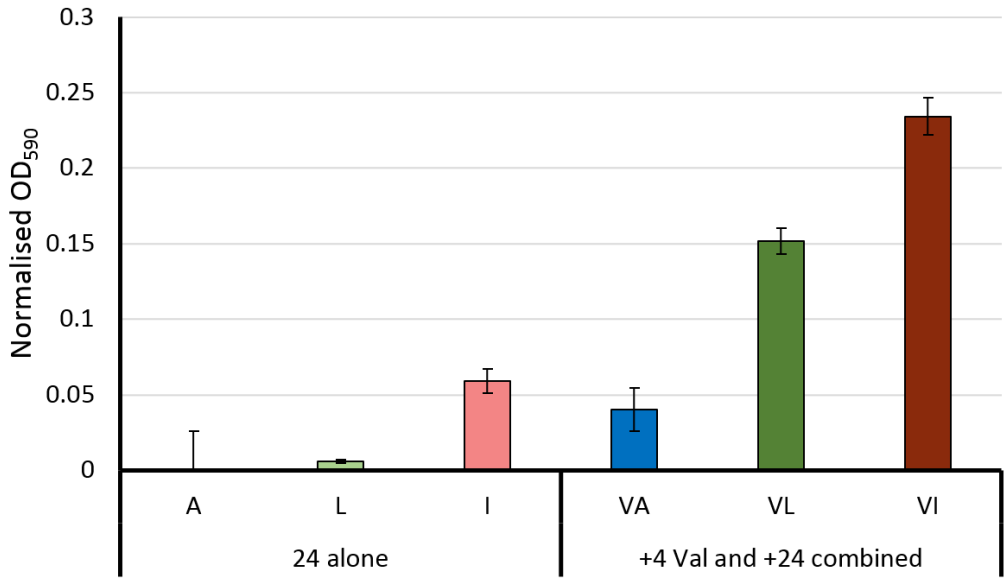



**Figure 3.8. Quantitative assessment of activity of PvdDCP1 substitutions that contained +24 mutations found alone or in combination with Valine at +4 position.** PvdDCP1 variant substitution strains were set up in quadruplicate in a 96wp and incubated at 18 °C. After 48 h, absorbance at 590 nm and 800 nm was measured using a microplate reader. Relative indigoidine concentration was calculated using these values, as described in section 2.10.2.1. Error bars are +/- one standard deviation.

We can also compare cases where the +24 mutant was found on its own, and then recovered along with a variety of +4 mutations within the combined +4/+24 screen. In many cases it is clear that the +4 mutations found in the combined screening were passenger mutations, i.e. only there because another dominant substitution such as Leu was present at the +24 position. In some cases these may in fact be antagonistic mutations, where the +4/+24 variant had lower activity than the +24 alone. For example, Leu at +24 on its own had a greater activity than if Asp, Glu or Gly were present at the +4 position as well (+24 Leu yielded indigoidine after one day on an agar plate, while the +4/+24 variants D/L, E/L and G/L all yielded indigoidine after two days) (Table 3.5). This was not generally the case the other way around, i.e. no dominant mutations were observed at the +4 position that were able to support non-contributing mutations at the +24 position, indicating that the +24 position in this context was the more important mutation of the two for improving activity.

### 3.2.6 Kinetic analysis of top mutants

To verify trends that were seen at an *in vivo* level, the top performing PvdDCP1 substitution variants of BpsA from each library were selected for purification in order to enable kinetic analyses. The top variants chosen were the +4 Phe and +24 Ile substituted variants, as well as the top +4/+24 combination variant, L/M, and the most active variant from Dr Owen's initial round of

evolution, +4/+24 I/I. Each of these mutants was already cloned into the pCDFDuet plasmid, which contains a His6 tag to allow for nickel affinity chromatography purification. Top mutants were each transformed into the *E. coli* BL21 *ΔentD* strain, i.e. an expression host lacking an activating PPTase to ensure the purification of un-activated *apo* BpsA. Enzymes were purified following the protocol for WT BpsA (Owen et al., 2011), with mutations having no noticeable effect on solubility or expression levels. Velocity values were derived at different L-glutamine concentrations and all variants were found to fit a Michaelis-Menten model (curves for each are shown in Figure 3.9). Kinetic parameters were derived for each of the top mutant enzymes as well as WT BpsA (Owen et al., 2011), and are given in Table 3.6.

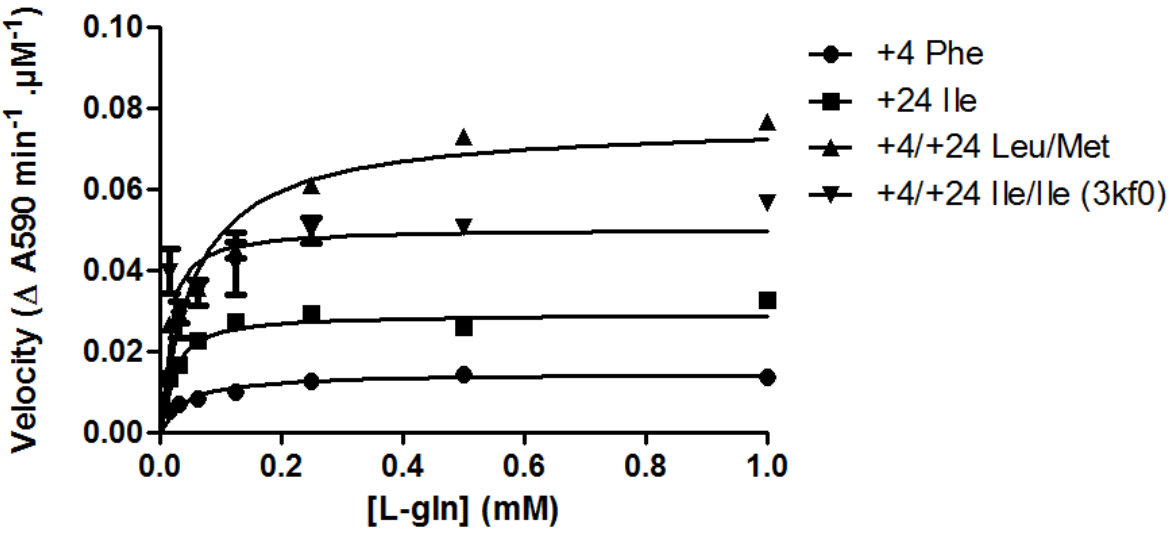



**Figure 3.9. Michaelis-Menten curves for each substitution variant tested.** Plots shown indicate the maximum velocity of indigoidine formation for each variant at different concentrations of L-glutamine. Curves were fitted using Graphpad Prism® and used to derive kinetic parameters for each BpsA variant with L-glutamine as substrate. Data points were generated from three repeats and error bars are +/- one SEM.

**Table 3.6 Kinetic parameters for BpsA and PvdDCP1 substituted improved variants**

| BpsA variant | $V_{max}$ (Δ $A_{590}$ $min^{-1}$ .$\mu M^{-1}$) | $K_m$ (mM) | $V_{max}/K_m$ (Δ $A_{590}$ $min^{-1}$ .$\mu M^{-1}$.$mM^{-1}$) |
|---|---|---|---|
| WT | 1.16 ± 0.023 | 0.67 ± .046 | 1.73 ± 0.12 |
| +4/+24 Ile/Ile | 0.05 ± 0.002 | 0.012 ± 0.004 | 4.17 ± 1.40 |
| +4 Phe | 0.015 ± 0.0003 | 0.038 ± 0.004 | 0.39 ± 0.42 |
| +24 Ile | 0.028 ± 0.0007 | 0.017 ± 0.003 | 1.65 ± 0.29 |
| +4/+24 Leu/Met | 0.076 ±0.002 | 0.056 ± 0.008 | 1.36 ± 0.20 |

All variants were substantially impaired in maximal velocity relative to WT BpsA (being 1-2 orders of magnitude slower). However, in all cases the $K_m$ had also diminished. The $V_{max}/K_m$ provides a measure of the relative catalytic efficiencies for each enzyme. Taking this into account, the +4/+24 I/I variant (3kf0) has a higher efficiency than the WT BpsA enzyme. The +24 Ile substituted variant

and +4/+24 L/M combination variant both have similar efficiencies to the WT BpsA enzyme. The +4 Phe substituted variant has a lower activity than all the other tested variants, which is consistent with results seen in the liquid culture assays.

## 3.3 Discussion

The aim of this chapter was to investigate features of CP domains that were important for the creation of functional BpsA CP-substitution variants. This involved optimising the interactions between the foreign CP domains and the downstream TE domain of BpsA. The particular features studied were the location of the foreign CP to be substituted in its original enzyme, as well as two key residues, the +4 and +24 positions (relative to the invariant Serine).

Four BpsA substitution variants having CP domains that were located upstream from TE domains in their native context were all found to be functional, with the most active variant producing indigoidine after two days in a solid media assay and all variants producing pigment within six days. The fact that they were all at least minimally functional is consistent with the previous observations that CP domains found upstream of TE domains in their native context are generally able to function when substituted into BpsA (Owen, 2010). However, the four CPs were specifically chosen because their sequences were inconsistent with a pattern previously noted; that is, they each contained a Thr residue at the +4 position instead of the hydrophobic residue that Dr Owen had found to be present in all of the CPs upstream from TE domains in his study. As each of these four CP-substitutions was functional, this implies that a hydrophobic residue at +4 is not necessary for a functional BpsA CP domain swap. These four CP domains must therefore contain other residues involved in optimising the interaction with the TE domain, which compensate in some manner for the +4 Thr. Nonetheless, activity for three out of the four substitutions was very low (not detectable in liquid culture assay), indicating further opportunity to evolve and improve these substitutions. It may be that targeted mutagenesis of the +4 site to a hydrophobic residue would be a good starting point.

Next the substitution of PvdDCP1 into BpsA was examined. Located upstream from a C domain in its native context, and so unable to result in a functional variant when directly substituted into BpsA, this CP domain substitution had previously been improved using error prone PCR (Owen, 2010). Mutations at the +4 and +24 positions had been found to restore activity to the substitution, in particular when these positions were changed to hydrophobic residues. This research confirmed that a mutation to a hydrophobic residue at the +4 position could restore activity to the substitution, revealing four other residues, Leu, Phe, Met and Tyr, in addition to Ile

and Ala (which were the only two residues found via error prone PCR) that resulted in a functional BpsA variant. The variants with the highest activity as measured by the liquid culture assay (Ile, Leu and Phe) are the more hydrophobic residues (Kyte and Doolittle, 1982). While the earlier substitutions of the four CP with the +4 Thr had shown that a hydrophobic residue at the +4 position was not necessary for function, in the case of PvdDCP1 substitutions conversion of this threonine to a more hydrophobic residue was sufficient to restore a moderate level of activity.

The results for the single mutations in PvdDCP1 at the +24 position were not so straightforward. Twice as many amino acids (12 in total) substituted at this position were found to result in functional variants. Previously Ile and Leu mutations had been identified via error prone PCR, so again assumed that hydrophobic residues were necessary for interaction with the TE interface. While the top variants identified in this chapter (Ile, Val, Leu and Met, as in the liquid culture assay) are all hydrophobic residues, there were other residues at this position also capable of giving rise to functional variants, including the hydrophilic Ser, Gln and Thr. Also, the PvdDCP1 WT residue at this position is hydrophobic (Phe), so if a hydrophobic residue at the +24 position was all that was required to yield a functional variant then the WT PvdDCP1 substitution should have been active. The main difference between Phe and the residues found in the improved mutants is that Phe is a bulky aromatic residue, while all the other residues are smaller and non-aromatic. In the WT BpsA CP, this position is occupied by a Leu, a smaller hydrophobic residue. We can conclude that a small, hydrophobic residue at the +24 position which is similar to the BpsA WT residue (Leu) is better able to interact with the BpsA TE domain. It appears that even a small, neutral residue that is hydrophilic, is still able to interact more successfully with the TE domain than the bulky aromatic residue Phe.

While a single mutation at the +4 or +24 positions in PvdDCP1 was in some cases sufficient to convert a completely non-functional CP into a functional one, combining beneficial mutations at these positions showed an even greater effect. There were a larger number of +4/+24 substitution variants found, even though the library was not screened to saturation (a total of 30 unique active clones were found from a total of 39 improved variants sequenced). While many of these had relatively low activity and looked to have passenger or antagonistic mutations present, overall there was a much larger number of mutants with highly improved activity than was seen for mutations at each site alone (e.g., fourteen +4/+24 clones turned blue after 1 day agar, compared to three sole +24 site variants and none of the sole +4 site variants). Many of these top active variants had synergistic mutations at this position; mutations that showed marginal improvement on their own yet reached almost WT levels of activity in combination (as measured in a liquid

culture assay). When kinetics of the top variants were tested most were found to have similar efficiency levels to the WT BpsA enzyme, the two +4/+24 variants tested (I/I and L/M) had better or close to the efficiency of the WT enzyme.

The +24 site of the CP domain had also been identified as being involved in domain interactions via scanning alanine mutagenesis studies, in both the EntB and EntF CP domains, where a mutation to Ala at the +24 position was found to be detrimental to activity (Lai et al. 2006; Zhou, Lai, and Walsh 2006). One of the limitations of the scanning alanine mutagenesis approach is that alanine may be tolerated in a position, even if that position is important for function. In this case we have shown that small hydrophobic residues at the +24 position may improve the function of a CP-substitution, indeed Ala was one of the mutations found in this study that gave rise to a functional variant. This may be why Lai et al. (2006) found the +24 site to only be of intermediate importance compared to the other sites identified in that study.

In another CP substitution study (where the CP of the NRPS VibB, required for vibriobactin synthesis in *Vibrio cholera*, was substituted into the *E. coli* enterobactin synthase EntB) (Zhou et al., 2007b)), a mutation at the +24 position was also identified as being able to restore function to CP-substituted protein. In this case the improved variant contained the charged, hydrophilic residue Lys at +24. The downstream domain of the CP here was a C domain. It may be that the +24 position is a key residue for the interactions between CP and their downstream domains, and that the type of residue required may be dependent on what the downstream domain is. If this is the case, then site directed mutagenesis may be able to be used instead of a completely random approach to optimise the activity of future CP domain substitutions. This would have the advantage of generating a much lower number of possible clones to screen.

Some inconsistencies were seen in the way indigoidine production activity levels were measured, both when comparing solid media *vs*. liquid media of *E. coli* expressing the BpsA variants, and when comparing the kinetics of the purified enzymes. Some variants that were slower to turn blue on an agar plate were faster in liquid assay. For example, the +24 Ile-substituted variant was faster than the Leu, Val and Met-substituted variants in the liquid culture assay but a day slower on the solid media assay. Each of these methods of measuring indigoidine production is appropriate for different objectives. The solid media plate assay is simple and gives a quick and easy estimation of activity. This is excellent for screening as it has a high capacity, and it is easy to spot active mutants. This assay gives a rough estimation of activity, with variants that show activity within the first 48 h usually sensitive enough to then test in a liquid assay. However, the liquid assay is more appropriate for second tier screening of hits, as it is a lot more sensitive and quantitative. It is

more of a medium throughput approach so is best used when the most active hits have already been identified on agar plates. One limitation of the liquid culture assay is the comparison with the WT enzyme. The WT reaches peak levels of 590 nm absorbance faster than the mutants tested, and when measurements are taken at either 24 or 48 h (depending on activity of variants) the WT absorbance has likely already tapered off. This means a variant may look to have a similar activity to WT at one time point, but if measured earlier it would have appeared far less active comparatively. This assay is effective however for comparing mutants with similar levels of activity. Finally, measurement of Michaelis-Menten enzyme kinetics provided a much more detailed analysis of enzyme activity, however is much more time consuming so was only used for top mutants. Results were found to be mostly consistent with liquid assay results, in that when looking at enzyme efficiencies the top mutants were comparable to the WT. It is important to note however that the WT BpsA enzyme had the highest $V_{max}$ of any of the variants tested, which is consistent with this enzyme performing most effectively in the liquid assays, where the glutamine concentration was not limiting. Thus, even though the +4/+24 I/I clone had a higher efficiency than the WT, it did not appear to be more active in the liquid culture assay, where its substantially lower $K_m$ for L-glutamine would not have provided a discernible rate advantage.

As this work was being performed, a CP substitution study was published using the BpsA homologue from *Photorhabdus luminescens*, IndC. Beer et al. (2014) substituted eight natural and seven synthetic CP domains into IndC. Five of the eight natural CP domains were located upstream from TE domains in their native contexts, yet only one of these substitutions was functional. This was inconsistent with the results seen in our study, but these other substitutions may have failed due to incorrectly defined domain boundaries. As CP domains are part of larger enzyme complexes, the boundaries of domains need to be carefully defined. Beer et al. (2014) only optimised their boundary positions after their initial substitutions were completed, going on to show that once these had been optimised a previously non-functional substitution (the CP from BpsA) was now functional. As the other CPs were not retested with the new boundaries, it is not possible to determine whether the other CP substitutions would have been functional had the correct boundaries been selected. In contrast, the boundaries for the BpsA CP substitutions described in this chapter had previously been optimised (Owen, 2010), and are similar to the final optimised boundaries selected by Beer et al. Of note, Beer et al. did not highlight any residues as being particularly important. All of the substitutions that were functional had Val at the +4 position and Phe at the +24 position, and were identical to the WT IndC residues.

The implications of the work described in this chapter are that there are several options for engineering CP-substituted variants in order to improve the activity of the substituted enzymes. Firstly, if you take into account the location of the CP domain in its original enzyme (e.g., upstream of a C domain *vs.* upstream of a TE domain) and transport it into an equivalent location, it seems that this would increase the likelihood of it being functional. This has been a consistent result seen with the BpsA enzyme and may translate to other NRPS enzymes as well. It has also been shown that targeted mutagenesis of just one or two important residues can be successful in improving the activity of a substituted variant to almost WT levels. In particular the +24 position has shown to be important for function both in BpsA substitutions as well as substitutions into another NRPS enzyme (EntB), implying that this might be a general rule that could be applied to improve CP domain activity within recombinant NRPS constructs.

# Chapter 4: Metagenomic screening for PPTases and secondary metabolite genes

## 4.1 Introduction

The previous chapter describes how BpsA can be used to investigate interactions between CP domains and their neighbouring domains. As part of this collaborative research, Dr Jeremy Owen had earlier evolved several CP-substituted variants of BpsA. These contained CP domains from foreign NRPS enzymes whose activity was improved via directed evolution to maximise their activity within BpsA. The research described in this chapter utilised some of these alternative BpsA enzymes in a different screening process, the discovery of novel PPTase and secondary metabolite genes from metagenomic DNA samples. BpsA is inactive when expressed in *E. coli* due to the lack of any activating endogenous PPTase (Owen et al., 2011); co-expressing the *bpsA* gene with a length of DNA that includes an activating PPTase gene results in production of indigoidine. Thus, we reasoned that we would be able to use this BpsA based screen as a means to discover novel PPTases and, by association, the secondary metabolite genes that are often found adjacent them. This research focused on the further development of this screen, both by improving the *E. coli* screening strain by reducing background activation of BpsA and also by using the previously engineered CP-substituted BpsA variants to broaden the screen.

In the current work, the screening strain was first improved via deletion of the native *E. coli* PPTase, *entD*. This reduced background activation of BpsA, which had previously resulted in a small level of blue pigmentation seen in the centre of all colonies after two days. This was particularly important for utility of one of the evolved BpsA variants, which contained a native CP recognised by EntD. Expanding the screening method to include use of the evolved BpsA variants was found to broaden the range of PPTase-containing hits recovered. Subsequently a detailed proof of principle was performed, including the complete screening of two separate metagenomic libraries with three different *bpsA* variants (including WT *bpsA*). The results demonstrated that alternative CP domains integrated into *bpsA* can recover PPTase enzymes that have differing specificities. It was also shown that this screen can serve as a very effective enrichment method for discovering secondary metabolite biosynthesis genes (NRPSs and PKSs), with nearly 50% of the hits discovered containing evidence of these biosynthetic gene clusters. The majority of results presented in this chapter were published in the peer-reviewed journal Environmental Microbiology (Owen et al., 2012), a publication for which I was co-first author.

### 4.1.1 Metagenomic screening to discover previously inaccessible enzymes and natural products

Microorganisms have always been a rich source of enzymes and natural products, and historically discoveries of such products have come from microorganisms which are able to be cultured under standard laboratory conditions. Recent metagenomic analyses of various microbial environments has shown that they contain much more variety than previously thought, with over 99% of these microorganisms seemingly unable to be grown in the lab (Keller and Zengler, 2004; Rappé and Giovannoni, 2003; Torsvik et al., 2002). The field of metagenomics has been expanding and many methods for tapping into this diversity have been developed (Banik and Brady, 2010). Such methods involve isolating the DNA directly from environmental samples known to be rich in microbial DNA, such as soil or sponges, and then screening these via various methods to uncover new genes of interest. These screening methods are traditionally split into either sequence or function based screening, both of which have advantages and disadvantages. Sequence based screening relies on knowing what you are looking for and isolating genes whose sequence or conserved sequence regions are already known; this can be a powerful technique generating huge amounts of data, however you will not find anything unless you know what to look for. The basis for our screen is function based, which has the advantages of being able to screen for something truly novel, as long as you have the right activity screen to find it. This approach also ensures that the genes found will be able to be expressed as functional proteins in *E. coli* – an important consideration for downstream applications of the recovered genes. Function based screens can also be cheaper than large scale sequence based efforts, however are dependent on being able to clearly detect an activity of interest.

### 4.1.2 Use of PPTases in metagenome screening

PPTases have been utilised before in a metagenomic screening context to discover new secondary metabolite genes, albeit in quite a different context. In an early instance of this Sfp, a broad range PPTase, was used to identify CP domains within a metagenomic sample (Zhang et al., 2009). This study used a phage display method to display expressed proteins from a metagenome sample on the phage's surface, and subsequently any carrier proteins present were labelled with CoA-attached biotin using Sfp (Yin et al., 2007). Any biotin labelled CPs were then isolated through streptavidin binding and then identified, with the goal being to recover the natural product biosynthetic gene clusters that expressed the labelled CPs. This was initially shown to be effective in a proof of principle study that used a library created from *B. subtilis* DNA (Yin et al., 2007), with

almost half of the known CPs in *B. subtilis* recovered in this screen. Another group developed a similar method which also involved using Sfp to label CPs, this time in a proteomic sample instead of a phage display (Meier et al., 2009). This research again employed a CoA-biotin derivative, and isolation of labelled proteins through binding with streptavidin. Proteins were then identified through mass spectroscopy methods. These approaches both used one PPTase (Sfp) to identify many CPs from a genome. This has the advantage of finding NRPS or PKS genes directly, although many false positives were also recovered in both cases. The concept of targeting PPTases as a means of identifying unknown antibiotic biosynthesis clusters was first suggested by Mootz et al (2002b), by using complementation of a *lys5* mutation in *Saccaromyces cerevisiae*. The *lys5* gene encodes an essential PPTase in the lysine biosynthesis pathway. Foreign PPTases expressed in a *lys5* mutation strain allow for growth on media lacking lysine. However, there is no evidence of proof of principle for their idea having been demonstrated. Our aim was to screen for PPTase enzymes using the CP domain of BpsA as a target. As the genes for PPTases and the secondary metabolite enzymes they activate appear to often be found nearby to each other, it was hoped that the discovery of these PPTases would lead to the discovery of new secondary metabolite gene clusters.

### 4.1.3 BpsA as a tool in metagenome screening

Many of the properties that made BpsA an ideal candidate for screening of improved CP domain swaps in the previous chapter also lend themselves to this metagenomic screening strategy. Still a key property is its ability to produce a coloured product when activated, with any blue colonies easily visualised on relatively high density agar plates that contain L-glutamine. Another property of BpsA that makes it particularly useful for metagenome screening is the fact that it is expressed in its inactive form when expressed in *E. coli.* This is due to the fact that the native PPTases of *E. coli* do not recognise the CP domain of BpsA. When BpsA is expressed alongside environmental DNA (eDNA) fragments, the presence of any expressed PPTases that recognise BpsA will cause the enzyme to be activated. This idea was first tested using a small insert (2-6 kb) soil library, kindly provided by Nadia Parachin of Lund University (hereafter known as the “Swedish soil library”) (Owen, 2010). A total of nine hits were found, representing three unique sequences, all containing genes coding for PPTases (Figure 4.1). One of these hits also had fragments of natural product biosynthetic genes clustered with it. The annotations for all genes found are in Table 4.1. These promising results led to the further development of this screen, in the research that will be described in the remainder of this chapter.

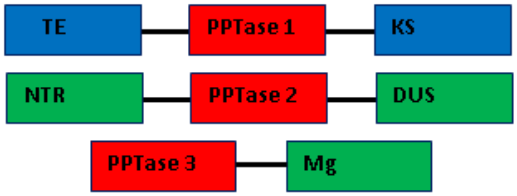



**Figure 4.1 eDNA fragments isolated from preliminary work. Adapted from Fig 7.10 (Owen, 2010).** Three hits found from the Swedish soil library in an initial screen. Number of PPTase corresponds to number of the clone. Red indicates a PPTase; blue indicates biosynthetic associated genes and green indicates all other genes. Table 4.1 gives a description of the genes.

**Table 4.1 Top BlastX matches for ORFs discovered in eDNA fragments**

| Clone | Gene | Top described match from BLAST search | Species within which top match was found | Percentage identity |
|---|---|---|---|---|
| **1** | **TE** | microcystin synthetase associated thioesterase | *Planktothrix agardhii* | 46% |
| | **PPTase 1** | 4'-phosphopantetheinyl transferase | *Beijerinckia indica subsp. indica* | 43% |
| | **KS** | polyketide synthase | *Streptomyces coelicolor* | 59% |
| 2 | NTR | Nitroreductase | *Nitrococcus mobilis* | 50% |
| | **PPTase 2** | 4'-phosphopantetheinyl transferase | *Micromonospora aurantiaca* | 39% |
| | DUS | Dihydrouridine synthase | *Amycolatopsis mediterranei* | 63% |
| **3** | **PPTase 3** | 4'-phosphopantetheinyl transferase | *Bradyrhizobium japonicum* | 43% |
| | Mg | Magnesium Transporter | *Marichromatium purpuratum* | 55% |

#### 4.1.3.1 Demonstration that *bpsA* constructs containing alternative carrier protein domains expand the screening capabilities

One of the limitations of only using WT BpsA is that it can only pick up those PPTases that would recognise and activate the CP domain of BpsA. We reasoned that there may be many PPTases that are incapable of such recognition (e.g., as is in the case with the native PPTases of *E. coli*). If so, then it should be possible to improve upon the initial screen by introducing different CPs from foreign NRPS enzymes into BpsA and using these to access a wider variety of PPTase and therefore secondary metabolite genes. As discussed in the previous chapter, the evolution of CP domain substitutions within BpsA had previously been performed in the context of gaining better understanding of interactions between different NRPS domains. As a result of this work, several different BpsA CP-substitutions were created, having been evolved for increased activity (Owen, 2010). In this chapter we aimed to take advantage of this unique resource, and use these evolved

forms of *bpsA* as alternative reporter genes for PPTase activity - thereby expanding number of potential PPTase recognition sites and thus hits found in this metagenomic screen. The first alternative CP domain that was used in this study was the terminal CP domain from *pvdD*, a pyoverdine producing NRPS of *P. aeruginosa* PAO1. The second alternative CP domain used was from the *entF* NRPS from the enterobactin synthesis cluster in *E. coli*. Firstly the native PPTase of *E. coli*, *entD*, was knocked out of the BL21 *E. coli* strain used for screening, in order to eliminate background activation by EntD. The strains of *E. coli* with the *entD* gene deletion which carried the improved BpsA CP domain substitution constructs were called EcoBlueP (BpsA with PvdD CP substitution) and EcoBlueE (BpsA with EntF CP substitution), with the WT BpsA enzyme containing strain referred to as EcoBlue1.

### 4.1.4 Metagenomic libraries used in this study

The two metagenomic libraries used in this study were both previously synthesised small insert plasmid libraries, limiting the size of any genes found. This is not a limiting factor for finding PPTase genes, the largest known being around 700 bp long. It did however mean that while this screen could find fragments of secondary metabolite genes, it would likely not recover any complete gene clusters (which are frequently as big as 100 kb) and so would only be used as a proof of principle screen in this regard. The Swedish soil library used in the preliminary screening (Owen, 2010) was also used here, both in order to screen it to saturation with the WT *bpsA* and to test it with the two new alternative CP-substituted *bpsA* constructs. This library contained inserts ranging from 2 to 6 kb in size. Another metagenomic library was also used - a larger library created, but not screened by, Dr Owen. This was created using soil from Wellington, New Zealand, and will be referred to as the “New Zealand soil library”. This second library had inserts ranging from 1 to 3 kb in size.

### 4.1.5 Aims

- Improve screening strains via knock out of the native *E. coli* PPTase, *entD*
- Rescreen Swedish soil library to saturation using EcoBlue1, EcoBlueP and EcoBlueE
- Screen a newly created NZ soil library using EcoBlue1, EcoBlueP and EcoBlueE

# 4.2 Results

## 4.2.1 Creation of optimised *ΔentD* screening strain

The first small scale proof of principle screen described in section 4.1.3 was performed in BL21, an *E. coli* strain that contains all native *E. coli* PPTases. However a level of background activity was seen as a result of BpsA being activated to a small degree by the native *E. coli* PPTases. This and the fact that EcoBlueE contains an *entF*-substituted *bpsA* variant, *entF* being a native substrate for the *E. coli* PPTase *entD,* motivated us to knock out *entD*, to eliminate any background activation. This PPTase is non-essential, involved in secondary metabolism including the production of enterobactin (Lambalot et al., 1996). The knockout of *entD* was performed using the method of Datsenko and Wanner (Datsenko and Wanner, 2000), as described in more detail in section 2.7. The successful removal of the *entD* gene eliminated the background activation of BpsA, as can be seen in Figure 4.2. Without the knockout, after two days small levels of background activity are seen with WT BpsA (small blue dots in the centre of colonies, Figure 4.2, upper left corner). Much darker blue can be seen in the strains with the two CP-substituted BpsA reporters expressed, showing that it was indeed necessary to knock out *entD* in order to screen with these alternative reporters. After *entD* knock out (lower panels) no blue pigmentation can be seen with any of the BpsA variants.

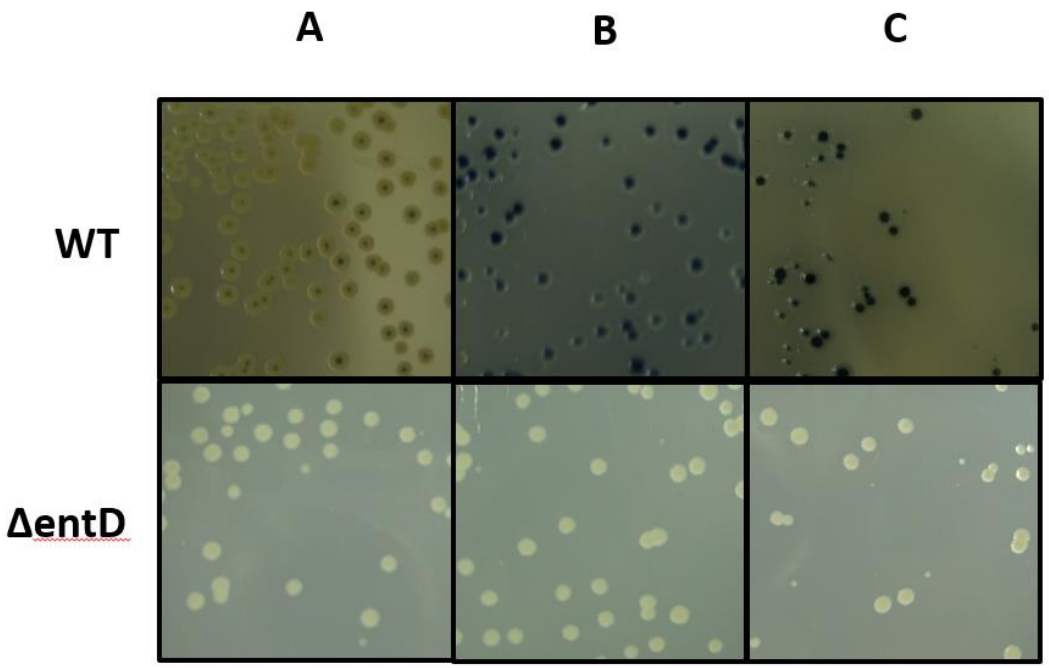



**Figure 4.2 BpsA variant activation in WT strain vs *entD* knock out.** A) WT *bpsA*, B) *pvdD* substituted *bpsA* and C) *entF* substituted *bpsA*, were each expressed in either BL21 (top panels) or BL21 *ΔentD* (bottom panels). All strains were grown on identical pigment producing media Photos taken after two days incubation at room temperature.

### 4.2.2 Screening the Swedish soil eDNA library with three alternative *bpsA* reporter strains

In previous work the Swedish soil eDNA library was screened for PPTase activity using WT BpsA. This resulted in the discovery of three unique clones, each containing one PPTase gene as shown in Figure 4.1. Here this same library was screened again, first using the WT BpsA reporter to ensure the library was screened to saturation. Approximately 3 x $10^6$ clones were screened (an approximate 20 fold coverage given the estimated library size of 1.3 x $10^5$ unique clones (Parachin and Gorwa-Grauslund, 2011)), with approximately 5-10 x $10^3$ clones per plate. It was found that this cell density was high enough to enable genuinely high throughput screening while still allowing for selection of individual colonies from a plate, though occasionally hits had to be re-streaked onto a second plate in order to obtain single colonies. Plates were incubated at 21 °C for a total of seven days to ensure all PPTases, even slow acting ones, were found. Individual blue colonies were easily identified on even relatively high density plates due to the clear dark blue colour, e.g. Figure 4.3. Both plasmids (*bpsA* containing and eDNA insert containing) were then isolated in a mixed miniprep and transformed into another *E. coli* strain, enabling the eDNA containing plasmid (pRSETB) to be selected in isolation via ampicillin selection, in order to isolate

the single eDNA containing plasmid. Transformants were tested to confirm the loss of spectinomycin resistance and therefore elimination of the *bpsA* containing plasmid (pCDFDuet). The eDNA inserts were then sent for sequence analysis. A total of 61 clones were identified from the Swedish soil library when screened in EcoBlue1. These 61 clones comprised only four unique inserts, the three that had been previously found (Owen, 2010) and a fourth previously undiscovered insert. This indicated that this library had not previously been screened to saturation with the WT BpsA enzyme, but likely now was. Given the relatively large size of the soil library and the fact that most microorganisms should contain at least one PPTase, it seemed unlikely that there would be only four PPTase genes present. We concluded that this screen was only picking up PPTases that recognise the BpsA CP domain, which led us to conduct additional screens using our evolved *bpsA* variants, as described in section 4.1.3.1. Screening the library in EcoBlueP and EcoBlueE resulted in three additional clones being recovered, as well as the first four that had been recovered using the EcoBlue1 strain. For EcoBlueP, screening approximately the same number of colonies as with EcoBlue1 we recovered 51 blue colonies. When the inserts from these were sequenced these were identified as the four previous clones, along with one new clone (clone 5, Figure 4.4). Screening the library with EcoBlueE yielded 60 blue colonies. Sequencing revealed these to be each of the previously discovered five clones together with two additional clones (clones 6 and 7, Figure 4.4). In this case the evolved versions of *bpsA* were necessary to find other PPTases present in the eDNA library, which may not activate the WT BpsA as well. Figure 4.4 gives a summary of the hits found in this library, as well as the strain they were found in. Table 4.2 gives the full annotations for each ORF found in each of the eDNA inserts.

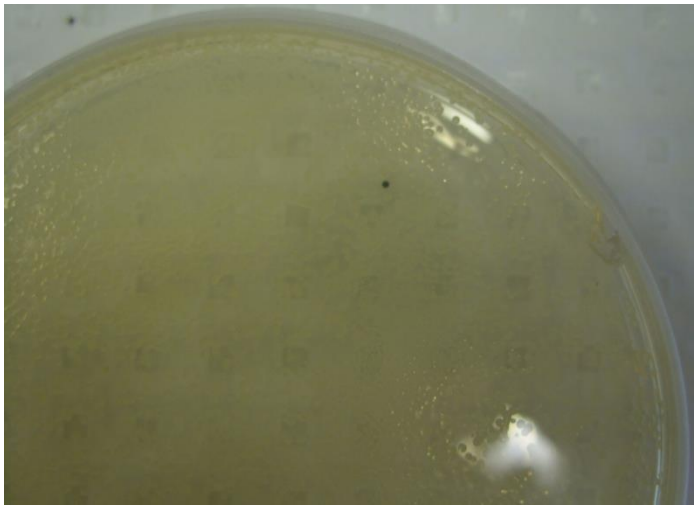
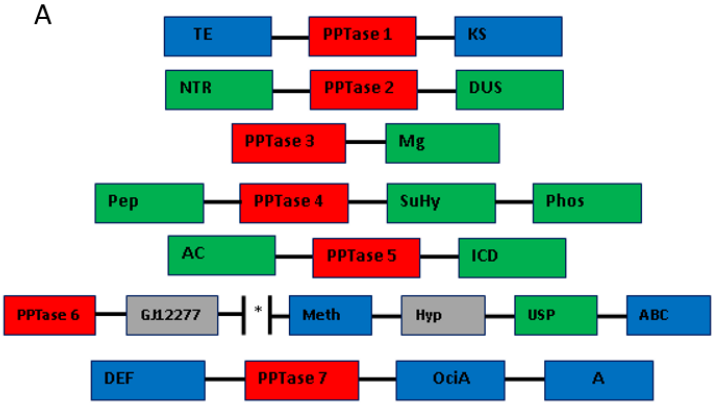


**Figure 4.3 Identification of a blue PPTase containing colony through eDNA screening.** Blue colony containing a PPTase 'hit' easily visible on an eDNA screening plate.

A




B

| Clone | Found in? | | |
|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE |
| 1 | Y | Y | Y |
| 2 | Y | Y | Y |
| 3 | Y | Y | Y |
| 4 | Y | Y | Y |
| 5 | N | Y | Y |
| 6 | N | N | Y |
| 7 | N | N | Y |

**Figure 4.4 Summary of all eDNA hits in Swedish soil library found with EcoBlue1, EcoBlueE, and EcoBlueP.** A) Summary of all the hits found in the Swedish soil library. Number of PPTase corresponds to number of clone. Red indicates a PPTase, blue indicates genes associated with biosynthetic gene clusters and green indicates all other genes. Grey indicates hypothetical gene or proteins with unknown function. B) Table indicates which of the three reporter strains each hit was found with, with a Y indicating it was found using that strain and an N indicating it was not found using that strain. * indicates that sequencing reactions from either direction were unable to proceed past that point. The full annotation details for each of the ORFs is given in Table 4.2

Out of the seven PPTase-containing clones found, three were found alongside genes that indicated the likely presence of a secondary metabolite gene cluster. Insert one contained partial ORFs whose top BLAST matches were a thioesterase domain (from a PKS) and a polyketide synthase gene (Table 4.2). Insert seven contained two genes where the top BLAST matches were to NRPS genes, as well as one ORF where the top match was for drug efflux transporter, also commonly associated with these types of biosynthesis genes, for exporting products from the producing cell. Insert six contained an ORF whose top BLAST match was for a methyltransferase. These are a class of common secondary metabolite tailoring domains, which add methyl groups to their associated products. Overall the BLAST scores for each of these genes was relatively low, with the highest being 59% but all the rest being under 50%, indicating that the biosynthesis genes found in such metagenomic libraries may be substantially different from other known biosynthesis genes. This highlights the advantage of functional based screening, as given the low similarity of sequences that were found it is unlikely these genes could have been found with sequence based screening (which relies on searching for conserved sequences).

**Table 4.2 Top BLASTX matches for ORFs within the Swedish eDNA library clones**

| Clone | Gene | Top described match from BLAST search | Species from which top match was found | Percentage identity |
|---|---|---|---|---|
| 1 | **TE** | microcystin synthetase associated thioesterase | *Planktothrix agardhii* | 46% |
| | **PPTase 1** | 4'-phosphopantetheinyl transferase | *Beijerinckia indica subsp. indica* | 43% |
| | **KS** | polyketide synthase | *Streptomyces coelicolor* | 59% |
| 2 | NTR | Nitroreductase | *Nitrococcus mobilis* | 50% |
| | **PPTase 2** | 4'-phosphopantetheinyl transferase | *Micromonospora aurantiaca* | 39% |
| | DUS | Dihydrouridine synthase | *Amycolatopsis mediterranei* | 63% |
| 3 | **PPTase 3** | 4'-phosphopantetheinyl transferase | *Bradyrhizobium japonicum* | 43% |
| | Mg | Magnesium Transporter | *Marichromatium purpuratum* | 55% |
| 4 | Pep* | peptidase m14 carboxypeptidase a | *Micromonospora sp. L5* | 37% |
| | **PPTase 4** | 4'-phosphopantetheinyl transferase | *Agrobacterium vitus* | 40% |
| | SuHy | Putative secreted sugar hydrolase | *Stigmatella aurantiaca* | 26% |
| | PHOS | type I phosphodiesterase/nucleotide pyrophosphatase | *Candidatus Solibacter usitatus Ellin6076* | 39% |
| 5 | Ac | adenylyl cyclase class-3/4/guanylyl cyclase | *Ralstonia eutropha JMP134* | 46% |
| | **PPTase 5** | 4'-phosphopantetheinyl transferase | *Microcoleus vaginatus FGP-2* | 39% |
| | ICD | putative integrase core domain protein | *uncultured marine microorganism* | 45% |
| 6 | **PPTase 6** | phosphopantetheinyltransferase family protein | *Gloeobacter violaceus* | 40% |
| | GJ12277 | GJ12277 | *Drosophila virilis* | 28% |
| | **Meth*** | type 11 methyltransferase | *Desmospora sp. 8437* | 34% |
| | HYP | hypothetical protein | *gamma proteobacterium IMCC1989* | 27% |
| | USP* | similar to universal stress protein | *Candidatus Kuenenia stuttgartiensis* | 65% |
| | **ABC** | ABC Transporter | *Syntrophobotulus glycolicus DSM 8271* | 45% |
| 7 | **DEF*** | drug efflux transporter | *Pseudomonas aeruginosa PAO1* | 26% |
| | **PPTase 7** | 4'-phosphopantetheinyl transferase | *Hyphomicrobium denitrificans ATCC 51888* | 51% |
| | **OciA** | OciA (an NRPS) | *Planktothrix agardhii NIVA-CYA 116* | 47% |
| | **A** | amino acid adenylation domain protein | *bacterium Ellin514* | 38% |

ORF names are given on the basis of top previously characterised match (lowest e value) as identified in BLASTX alignments. *Indicates ORFs where the top match was a hypothetical protein.

#### 4.2.2.1 PPTase activity of hits from Swedish library within EcoBlue1, EcoBlueP, and EcoBlueE

As some clones were only recovered using CP-substituted BpsA variants we wanted to check explicitly whether the PPTase in that particular clone was only able to recognise that particular CP domain. To test this each of the seven unique clones was transformed into EcoBlue1, EcoBlueP,

and EcoBlueE. These were then all plated out onto auto-induction agar plates so that the density of cells on each plate was approximately 100 colonies. After initial overnight growth at 37 °C plates were then incubated at room temperature for seven days and the time for blue colour to develop was observed. The results of this are summarised in Table 4.3. The number of ticks indicates the time taken for blue colour to develop. Three ticks indicates the blue colour developed on the first day, two ticks indicates blue colour development after two days, and one tick indicates a faint blue colour developed within three to seven days. A cross indicates no blue developed within the seven days. Figure 4.5 shows an example of representative plates, for hit seven. No blue colour was seen in EcoBlue1, and a darker blue was seen within EcoBlueP rather than EcoBlueE.

**Table 4.3 Cross reactivity of all Swedish clones in each reporter strain**

| Clone | Found in | | | Level of activity in | | |
|---|---|---|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE | EcoBlue1 | EcoBlueP | EcoBlueE |
| 1 | Y | Y | Y | ✓✓✓ | ✓✓ | ✓ |
| 2 | Y | Y | Y | ✓✓✓ | ✓✓✓ | ✓✓ |
| 3 | Y | Y | Y | ✓✓✓ | ✓✓ | ✓ |
| 4 | Y | Y | Y | ✓✓ | ✓✓✓ | ✓✓ |
| 5 | N | Y | Y | ✓ | ✓ | ✓ |
| 6 | N | N | Y | x | x | ✓ |
| 7 | N | N | Y | x | ✓✓ | ✓ |

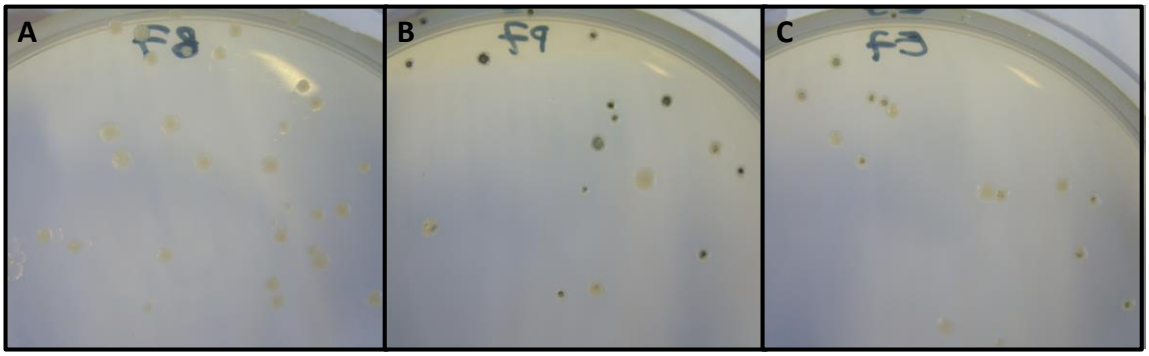



**Figure 4.5 Clone 7 expressed EcoBlue1, EcoBlueP, and EcoBlueE three days after plating on auto-induction media with incubation at 37 °C.** Clone 7 from the Swedish soil library in A) EcoBlue1 B) EcoBlueP and C) EcoBlueE. After three days incubation at room temperature, the darkest colour was seen in EcoBlueP, followed by EcoBlueE, with no blue colour seen in EcoBlue1. The colour of colonies on each of these plates did not alter substantially after an additional four days of incubation.

Table 4.3 shows that when activity was directly tested in this manner, results were generally consistent with the strain each PPTase bearing clone had been recovered in. There were however two discrepancies. Clone 5 was only found using the EcoBlueP and EcoBlueE reporter strains, yet this clone showed low level activity in each reporter strain (including EcoBlue1). Clone 7 was only

found using EcoBlueE yet showed a higher activity in EcoBlueP when transformed into each reporter strain. These discrepancies could be due to the frequencies with which each of these clones were found. Both clones were found at a low abundance within the library, clone 5 twice and clone 7 once. They would both have been easy to miss because they were found in such low abundance. The generally low activity of clone 5 in each of the reporter strains may have also been a contributing factor. It is still clear however that screening with the alternative strains EcoBlueP and EcoBlueE was necessary to recover all of the seven clones that were found, as if the library had only been screened with EcoBlue1 then we could never have recovered clones 6 and 7, which showed no activity at all in this strain. Nonetheless, every hit found activated the EntF-substituted BpsA variant, suggesting that all hits could have been found using the EcoBlueE strain alone.

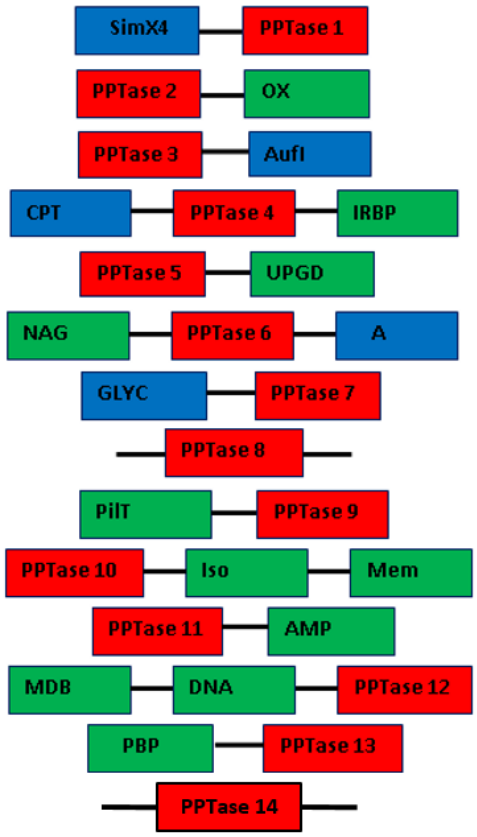
This result supports the fact that as well as discovering new secondary metabolite biosynthetic genes, this screen is also capable of finding PPTase genes with different CP specificities, which may also have useful biotechnological applications (indeed, development of more specific PPTase enzymes became a primary focus of research that is described later in this thesis). In terms of generating dark blue colonies, the evolved variants of *bpsA* were generally less active overall then the WT enzyme. Nevertheless, some PPTases still showed more activity with these "impaired" variants (e.g. PPTases 4 and 7, both showing greater activity with EcoBlueP) perhaps indicating a much greater affinity for that particular CP as a substrate.

### 4.2.3 Screening the New Zealand soil library with EcoBlue1, EcoBlueP, and EcoBlueE

In addition to the Swedish soil library, Dr Owen also created a library made from soil collected from a residential property in Wellington, New Zealand (New Zealand soil library or NZ soil library). In the current work this was also screened for PPTase activity, using EcoBlue1, EcoBlueP and EcoBlueE. The size of this library was not quantified, but it was assumed to be a much larger and more diverse library as double the number of unique hits were found, even though it did not appear to yet be screened to saturation. For this library, 2.5-3 x $10^6$ colonies were screened using each reporter strain. For EcoBlue1, 40 blue colonies were found, which when sequenced proved to be nine unique clones. When the library was screened in EcoBlueP 34 blue colonies were found, which when sequenced proved to be ten unique clones. In EcoBlueE only five blue colonies were found, which when sequenced proved to be three unique clones. Figure 4.6 gives a graphical summary of all hits found in the NZ soil library, as well as in which strain each hit was found. Table 4.4 gives the full annotations for each ORF found in each of the eDNA inserts. Again the two

alternative reporters were required to find every hit, with some being only found in either EcoBlueP (clones 1, 6, 7) or EcoBlueE (clone 13).

A




B

| Clone | Found in? | | |
|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE |
| NZ 1 | N | Y | N |
| NZ 2 | Y | Y | Y |
| NZ 3 | N | Y | N |
| NZ 4 | Y | Y | Y |
| NZ 5 | Y | Y | N |
| NZ 6 | N | Y | N |
| NZ 7 | N | Y | N |
| NZ 8 | Y | Y | N |
| NZ 9 | Y | Y | N |
| NZ 10 | Y | N | N |
| NZ 11 | Y | N | N |
| NZ 12 | Y | N | N |
| NZ 13 | N | N | Y |
| NZ 14 | Y | Y | N |

**Figure 4.6 Summary of hits found in NZ soil library.** A) Summary of all the hits found in the NZ soil library. Number of PPTase corresponds to number of clone. Red indicates a PPTase, blue indicates biosynthetic associated genes and green indicates all other genes. B) Table indicates which of the three reporter strains each hit was found with, with a Y indicating it was found in that strain and an N indicating it was not found using that strain. The full annotation details for each of the ORFs is given in Table 4.4.

Within the NZ soil library, five out of the 13 clones were found to also contain genes that are indicative of natural product gene clusters. Insert one contained a gene that matched to a simocyclinone biosynthesis gene, a PKS whose product has anti-tumour properties (Schirmer et al., 2005). Insert two contained a match to another PKS, AufI, involved in the production of the secondary metabolite aurafuron (Frank et al., 2007). Insert four contains a cyclic peptide transporter, commonly found in these gene clusters for use in the export of biosynthetic products (von Döhren, 2009). Insert six was a match to an NRPS A domain. Insert seven contains a glycosyltransferase, which is associated with post-translational modification of both PKS and NRPS

enzymes, for example in vancomycin production (Baltz, 2002). The gene isopropylmalate isomerase from insert ten has also been found in a gene cluster containing NRPS and PKS genes (Zhu et al., 2009).

**Table 4.4 Top BLASTX matches for NZ library clones**

| Clone | Gene | Top described match from BLAST search | Species from which top match was found | % identity |
|---|---|---|---|---|
| 1 | **SimX4*** | SimX4-like protein | *Streptomyces sviceus* | 79% |
| | **PPTase 1** | Phosphopantetheinyl transferase | *Streptomyces avermitilis* | 67% |
| 2 | **PPTase 2** | 4'-phosphopantetheinyl transferase | *Nostoc punctiforme* | 39% |
| | Ox | oxidoreductase | *Sorangium cellulosum* | 33% |
| 3 | **PPTase 3** | 4'-phosphopantetheinyl transferase | *Hyphomicrobium denitrificans* | 45% |
| | **AufI*** | Polyketide Synthase AufI | *Stigmatella aurantiaca* | 37% |
| 4 | **CPT** | cyclic peptide transporter subfamily | *Synechococcus sp.* | 64% |
| | **PPTase 4** | 4'-phosphopantetheinyl transferase | *bacterium Ellin* | 58% |
| | IRBP | interphotoreceptor retinoid-binding protein | *Citromicrobium sp. JLT1363* | 45% |
| 5 | PPTase 5 | 4'-phosphopantetheinyl transferase | *Verrucosispora maris* | 41% |
| | UPGD | undecaprenyl-phosphate glucose phosphotransferase | *Lautropia mirabilis* | 61% |
| 6 | NAG | Nacetylglucosaminyldiphosphoundecaprenol N-acety l-beta-D-mannosaminyltransferase | *Streptosporangium roseum* | 55% |
| | PPTase 6 | 4'-phosphopantetheinyl transferase | *Thiomonas intermedia K12* | 41% |
| | **A** | amino acid adenylation domain protein | *bacterium Ellin* | 50% |
| 7 | **GLYC** | glycosyltransferase | *Paenibacillus vortex* | 58% |
| | **PPTase 7** | 4'-phosphopantetheinyl transferase | *Sorangium cellulosum* | 41% |
| 8 | **PPTase 8** | 4'-phosphopantetheinyl transferase | *Microcoleus vaginatus* | 47% |
| 9 | PilT | twitching motility protein PilT | *Heliobacterium modesticaldum* | 51% |
| | **PPTase 9** | 4'-phosphopantetheinyl transferase | *Stigmatella aurantiaca* | 40% |
| 10 | **PPTase 10** | 4'-phosphopantetheinyl transferase MtaA | *Acidobacterium capsulatum* | 41% |
| | **Iso*** | isopropylmalate isomerase large subunit | *Roseovarius sp. 217* | 41% |
| | Mem | membrane protein | *Ramlibacter tataouinensis* | 27% |
| 11 | **PPTase 11** | 4'-phosphopantetheinyl transferase | *Nostoc punctiforme PCC 73102* | 46% |
| | Amp | AMP-dependent synthetase and ligase | *Bacillus tusciae* | 47% |
| 12 | MDB* | molybdopterin binding aldehyde oxidase and xanthine dehydrogenase | *Desulfotomaculum reducens* | 44% |
| | DNA | replicative DNA helicase domain protein | *Borrelia afzelii PKo* | 33% |
| | **PPTase 12** | phosphopantetheine-protein transferase | *Microcoleus vaginatus* | 37% |
| 13 | PBP | polysaccharide biosynthesis protein | *Rhodoferax ferrireducens T118* | 43% |
| | **PPTase 13** | putative 4'-phosphopantetheinyl transferase | *Candidatus Nitrospira defluvii* | 36% |
| 14 | **PPTase 14** | phosphopantetheinyl transferase | *uncultured gamma proteobacterium HF4000_48E10* | 40% |

ORF names are given on the basis of top previously characterised match (lowest e value) as identified in BLASTX alignments. *Indicates ORFs where the top match was a hypothetical protein.

### 4.2.3.1 PPTase activity of hits from NZ library within the EcoBlue1, EcoBlueP and EcoBlueE

As with the Swedish library, we wanted to see whether certain clones could only have been recovered using specific *bpsA* reporter strains. The NZ soil library was not thought to be screened to saturation (the NZ library was a larger size than the Swedish library, with fewer redundant hits found overall), and it was reasoned that testing the activity of each PPTase in each strain might provide an estimate of how close to being screened to saturation the NZ library was. For instance only three unique hits were found in EcoBlueE, so this was a way of testing how many more of the 14 total hits could be recovered using EcoBlueE.

To test this, each of the 14 unique clones found in the NZ soil library were used to transform into each of EcoBlue1, EcoBlueP, and EcoBlueE. These were then plated out onto identical auto-induction plates so that the density of cells on each plate was approximately 100 colonies. After initial overnight growth at 37 °C, plates were incubated at room temperature for seven days and the time for blue colour to develop was observed. Results are summarised in Table 4.5. The number of ticks indicates time for blue colour to develop. Three ticks indicates the blue colour developed on the first day, two ticks indicates blue colour development after two days, and one tick indicates a faint blue colour developed within three to seven days. A cross indicates no blue developed within the seven days.

**Table 4.5 Cross reactivity of all NZ clones in each reporter strain**

| Hit | Found in | | | Level of activity | | |
|---|---|---|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE | EcoBlue1 | EcoBlueP | EcoBlueE |
| 1 | N | Y | N | ✓✓✓ | ✓✓✓ | ✓✓✓ |
| 2 | Y | Y | Y | ✓✓✓ | ✓✓✓ | ✓✓✓ |
| 3 | Y | Y | N | ✓✓✓ | ✓✓✓ | ✓✓ |
| 4 | Y | Y | Y | ✓✓✓ | ✓✓✓ | ✓✓✓ |
| 5 | Y | Y | Y | ✓✓✓ | ✓✓✓ | ✓✓ |
| 6 | N | Y | N | x | ✓✓ | x |
| 7 | N | Y | N | x | ✓✓✓ | ✓ |
| 8 | Y | Y | N | ✓✓ | ✓✓✓ | x |
| 9 | Y | Y | N | ✓✓✓ | ✓✓ | x |
| 10 | Y | N | N | ✓✓ | ✓ | x |
| 11 | Y | N | N | ✓✓✓ | ✓✓ | ✓✓ |
| 12 | Y | N | N | ✓✓ | ✓ | ✓✓ |
| 13 | N | N | Y | ✓✓✓ | ✓✓ | ✓✓ |
| 14 | Y | Y | N | ✓✓✓ | ✓✓ | ✓✓ |

The results seen here (Table 4.5), indicating which clones are capable of giving rise to indigoidine production when expressed in each of the *bpsA* variant containing strains, were not consistent with which strains these clones were actually found in (Figure 4.6). While there were two instances of such inconsistencies when the same experiment was run for the Swedish soil library, there were many more such instances here, where a clone was not found in a particular strain yet was still able to activate the *bpsA* variant within that strain. It was suspected that this library had not been screened to saturation, and these results confirmed that. Nonetheless, it was demonstrated that (similar to the Swedish library) the complete set of hits could not be found with EcoBlue1 alone, with two out of 14 clones showing no activity in this strain. Interestingly, all clones in this instance had activity in the EcoBlueP strain, meaning we could have screened with this reporter alone and still have recovered all of the hits. As this library was not screened to saturation however, we cannot state that this is the case within the whole library.

## 4.3 Discussion

The discovery that the vast majority of bacteria cannot be cultured effectively in the lab has prompted widespread efforts to find other ways of accessing the natural product biosynthetic genes of uncultivable microorganisms, as bacteria have been such a rich source for drug discovery from natural products in the past. Here we describe a novel method, a relatively simple screen based on the function of PPTase enzymes. PPTases activate both NRPSs and PKSs - secondary metabolite genes responsible for many bioactive natural products. We hypothesised that PPTase genes would co-localise with their biosynthetic counterparts, and that by finding PPTase genes we could then pull out associated NRPS or PKS genes. The functional screening approach described here has several advantages, including the relative ease and high throughput nature of the screen which means a large number of potential hits can be screened quickly and easily. Another advantage of a function based screen over a sequence based screen is that it doesn't require prior knowledge of exact sequences or a specific product. As long as a biosynthetic gene cluster is found alongside a PPTase then the gene cluster itself could code for anything, making it more likely to discover clusters that encode entirely new structures. Indeed, many of the secondary metabolite genes we did find shared very low sequence identity with any known genes.

This novel metagenome screening method was successful in finding 21 previously undiscovered PPTase genes, including several that were associated with NRPS and/or PKS biosynthetic gene clusters. As PPTases are not always found directly near the genes for their substrates we did not expect to find evidence of these secondary metabolite genes in every hit. Still, this method has

proved to be a powerful enrichment tool, with a minimum of eight out of the total 21 clones found from both libraries containing evidence of these secondary metabolite genes, around 40% of the hits If we take in to account that these are small inserts, where we can't always see sufficient genes present upstream or downstream of the PPTase gene to be confident that they are not within a biosynthetic cluster, the actual hit rate is likely even higher.

The types of biosynthetic genes recovered included not only core NRPS and PKS genes (for example the A domain of an NRPS (Clone 7 from the Swedish soil library) or the "polyketide synthase" gene (Clone 1 from the Swedish soil library) found in the Swedish soil library), but also NRPS or PKS tailoring domains or other associated genes. Examples of these tailoring domains include glycosyl transferase (Clone 7 from the NZ soil library) or methyltransferase (Clone 6 from the Swedish soil library) domains.

While this experiment has shown this approach to be a successful one on the small scale, we were only using very small insert DNA libraries, the largest insert recovered being approximately 5 kb (Clone 7 in the Swedish soil library) whereas most were much smaller (1-2 kb). Secondary metabolite gene clusters can be up to 100 kb or larger. In order to find whole secondary metabolite gene clusters, libraries with a much larger insert size need to be screened. For this, specialised plasmids such as cosmids, fosmids and bacterial artificial chromosomes (BACs) could be used. These allow much larger lengths of DNA to be screened, around 40 kb for fosmids and cosmids or 150-350 kb for BACs. These types of libraries tend to be much harder to make and might lead to lower hit rates, e.g. due to potential problems with expression of genes (if a PPTase gene was in the in middle of a large insert and not near any promoters that *E. coli* can recognise). However, with these larger insert libraries there is the potential to discover entire secondary metabolite gene clusters. An advantage of this would be that if the PPTase is expressed well and functional in *E. coli* (as it must be, to be found in the screen) then this may increase the likelihood of any secondary metabolite genes being expressed in a functional form also.

As well as a method for discovering new natural product biosynthetic genes, this screen is also able to uncover new PPTase enzymes. There has been no direct screen for these enzymes in the past. This is of interest because of the many biotechnological applications of PPTases. The further characterisation of the PPTases, initiated in this chapter, was expanded in research described in the next chapter of this thesis. As well as allowing the discovery of totally new PPTases, the nature of the screen whereby we use different evolved variants of *bpsA* that contain alternative carrier protein domains means that we may also be able to pre-select for PPTases that have different carrier protein specificities. Table 4.3 and Table 4.4 suggest that each PPTase has a range of

activities with the different reporter strains. This is of interest for several applications of PPTases, for example site specific protein labelling, where more specific PPTase CP pairs may broaden the applications of the current technology. This property will be explored further in chapter Five.

An unusual and unexplained result is the small number of blue colony hits found in EcoBlueE compared with the other two strains using the NZ soil library, where only five blue colony hits were found in EcoBlueE, compared to 40 in EcoBlue1 and 34 in EcoBlueP. When screening the Swedish soil library, a similar number of hits (50-60) were found in all three strains. This suggests that we can rule out the *entF*-substituted *bpsA* being less active in general.

Screening both of the eDNA libraries with CP-substituted *bpsA* variants in addition to the WT *bpsA* gene proved to be a valuable strategy for uncovering an even greater range of PPTases and thus biosynthetic gene clusters. A total of five of the hits found from both libraries would not have been discovered with WT *bpsA* alone, due to the apparent inability of the PPTase enzyme to recognise and activate the native CP from *bpsA*. It seemed likely that there were many more PPTases present in the eDNA sample than the screen with WT *bpsA* was picking up. Indeed, it is probably still the case that there are even more PPTase containing inserts in both libraries that remain undiscovered. Possible reasons for this include that the three CPs used may not be sufficiently diverse, and screening with even more still may increase the numbers of hits recovered. The PPTase enzymes may also have had trouble being expressed in *E. coli*.. This however can also potentially be an advantage, as we may be selecting for enzymes that express easily in *E. coli* and so by not finding any poorly expressing enzymes we will avoid downstream expression and solubility issues. Another potential reason for not finding hits is that a relatively large proportion of eDNA may have contained incomplete PPTase genes.

Since publication of our screen (Owen et al., 2012) , other labs have adopted a similar approach. One screen for PPTase activity inspired by our research involves using a PPTase knock out strain of a bacterium that employs NRPS enzymes to make a siderophore; i.e., a secondary metabolite used to scavenge iron from iron restricted environments (Charlop-Powers et al., 2013). Here, libraries were screened on iron-restricted agar plates (to which a chelating agent had been added), and the presence of a PPTase within the eDNA meant siderophore production pathways could be activated and bacteria were able to grow. eDNA from these colonies was then isolated and PPTases were identified along with genes for potential secondary metabolite products. Proof of principle for this was demonstrated in two systems, *E. coli* with *entD* knocked out and *Pseudomonas aeruginosa* with *pcpS* knocked out. This screen can only find PPTases capable of activating both *entF* and *entB* (*E. coli* enterobactin CPs) in *E. coli* or all the PCP domains of the *P.*

*aeruginosa* pyoverdine synthetases. The *P. aeruginosa pcpS* gene is essential and for this screen to work a special strain needs to be used which has *pcpS* knocked out and replaced with *acpS* (Barekzi et al., 2004). This knockout is still viable, as *acpS* activates primary metabolism pathways, but is unable to activate the pathways of secondary metabolism. Overall this knock out screen proved to be a powerful enrichment tool, resulting in 20-50 fold enrichment for NRPS and PKS genes (Charlop-Powers et al., 2013). It should be noted however that this screen is dependent on having a PPTase knockout strain available, and will only find PPTases capable of activating all CP domains of the strain used.

While all of the PPTase enzymes discovered in this screen had homology to other known PPTases, this method could theoretically be used to find other genes with PPTase activity that may not share homology with known PPTases, i.e. a presently unknown class of PPTase. There has never been in the past a method to screen directly for and discover new PPTase activity like this. One of the big advantages of function based screens when compared to sequence based is that as long as a gene expresses a certain function, it does not matter if the sequence is known or not. During the course of this project it was thought that this screening method had uncovered a new type of PPTase, with a clone that was recovered three times and had a very small insert size (only 942 bp long), the top BLASTX match for which was a hypothetical protein (which contained a “domain of unknown function 1257”), this hit was not pictured in Figure 4.6. It appeared that this plasmid was conferring a PPTase activity and being such a small ORF it seemed obvious (at the time) that this must encode a PPTase enzyme. Extensive time and effort was spent on analysing this clone. Firstly it was attempted to clone any possible ORFs present in the insert (two small ORFs were 400-450 bp each), but when this was unsuccessful in producing a functioning PPTase product, progressively smaller sections of the ORF were also cloned, to try and determine minimum sequence which would give PPTase activity. This eDNA plasmid was then tested for its ability to complement a *ΔentD* strain, when grown on iron restricted (i.e. to show a different measure of PPTase activity- not just the activation of BpsA). It did not show the ability to complement *entD*. Two cyanobacterial genes with the same “domain of unknown function 1257” were also cloned and tested for the ability to activate BpsA, but they were also unable to. Nonetheless, when this hit was tested for activity from the original glycerol stock, indigoidine production was still seen. These cells were then passaged in order to lose the eDNA containing plasmid to test whether there was something strange about the strain that had caused auto-activation of BpsA. Eventually after multiple time resequencing the eDNA insert, it became apparent that there had been two eDNA insert containing clones present in the cell of the original hit strain- in two of the hits this was Clone 2 and in one Clone 4 (from the NZ soil library). It appears that the plasmid containing the

gene for the hypothetical protein must have been transformed into the same cell as each other originally in a double transformation event. Upon further research it was discovered that this kind of double transformation event can actually be quite common, with one test showing double transformants present when as little as 5 ng total DNA was used (Goldsmith et al., 2007). In the transformation of libraries for this study, between 50 and 100 ng of DNA was used, so this explains the presence of these double transformants. In the future this is something that needs to be taken into account when preparing such libraries, and much smaller concentrations of DNA should perhaps be used.

# Chapter 5: BpsA as a tool to characterise PPTase enzymes

## 5.1 Introduction

As a consequence of the metagenomic screening method employed in Chapter 4 for the discovery of new biosynthetic gene clusters, a number of novel PPTase genes were also discovered. These are referred to here as the "eDNA PPTases", with the PPTases from Clones 1-7 of the Swedish soil library referred to as PPTases Sw 1-7 and the PPTases from Clones 1-14 of the NZ soil library referred to as PPTases NZ 1-14. PPTases themselves are of interest for a number of different biotechnological applications, e.g. for use in the site specific labelling of proteins. In particular, PPTases that have a high degree of specificity for a CP domain substrate hold potential for many applications. The 21 eDNA PPTases discovered in Chapter 4 were of particular interest as it appeared that they intrinsically possessed a relatively high degree of specificity for the CP domains with which they were discovered. To further test this, in the research described in Chapter 4 each PPTase-containing eDNA insert was tested for the ability to activate the three different BpsA variants (Tables 4.3 and 4.6). Some of the PPTases appeared to be highly specific, having the ability to activate one or two of the BpsA substitution variants but not all of them (for example, for PPTase Sw 6 indigoidine production was only seen in the EcoBlueE strain; while for NZ 6 indigoidine production was only seen in the EcoBlueP strain).

In order to further examine the activity of these eDNA PPTases, the individual genes for each of the PPTases were cloned as part of work presented in this chapter, and further characterised using BpsA-coupled assays. Historically, methods for characterising PPTases have been both time consuming and laborious, which has meant very few of the PPTases described in published literature have been characterised *in vitro* at a kinetic level. As of the recent review by Beld et al (2014), only seven PPTases had been described kinetically. Methods more recently developed in our lab which involve the monitoring of the PPTase reaction by coupling it to the BpsA indigoidine synthesis reaction provide a less cumbersome means of measuring PPTase activity. These BpsA based assays were used here to further characterise the 21 eDNA PPTases described in Chapter 4.

Under this closer examination, it appeared that many of the eDNA PPTases had relatively low activities overall. They were also far less specific (in terms of their preferences for certain CP domains) than had been anticipated based on the work described in Chapter 4. The scope of the

current chapter was therefore broadened to include other, already known PPTase enzymes. These had been characterised before to varying extents, and the BpsA based assays were used here to expand on the existing knowledge of these enzymes.

### 5.1.1 Traditional methods of PPTase characterisation

Traditionally, methods for measuring the kinetic activity of PPTases have been technically challenging and time consuming. The formation of the activated holo-CP product by a given PPTase could be monitored in one of two ways, either through the incorporation of radiolabelled [$H^3$-pantetheine]-CoA onto the activated holo-CP, followed by quantification using liquid scintillation counting (Gehring et al., 1997; Quadri et al., 1998; Sanchez et al., 2001), or through HPLC analysis of the CP domains to quantify the relative levels of apo and holo CP domain based on their difference in size (Finking et al., 2002; Huang et al., 2006). Both methods have also been used to determine simply whether or not a PPTase is capable of activating a particular CP domain. Where a range of substrate concentrations are tested these methods can be used to determine kinetic parameters. However, these assays are all end point assays, and the PPTase reaction cannot be monitored in real time.

### 5.1.2 BpsA as a method for determining PPTase kinetic parameters

In our lab a new method of measuring PPTase kinetics, based on the enzyme BpsA, was previously developed (Owen et al., 2011). This assay is based on the fact that the rate of indigoidine synthesis is directly proportional to the amount of holo-BpsA present. The rate at which indigoidine synthesis increases over time is therefore proportional to the rate at which a PPTase activates BpsA. This can be measured in a simple assay by measuring the absorbance at $OD_{590}$ (i.e., levels of indigoidine synthesis) over time. The data for the reaction velocity was found to fit a Michaels-Menten model, allowing kinetic parameters to be determined. This method was used to determine the kinetic parameters for three PPTases: Sfp, PcpS, and PP1183 (Owen et al., 2011). The data obtained for Sfp and PcpS (with respect to CoA as the variable substrate), was in reasonable agreement with previously determined values. This method has since been used in the characterisation of several other PPTases (Vickery et al., 2014; Zimhony et al., 2015).

### 5.1.3 Competition assay with BpsA

While measuring kinetic parameters in a BpsA coupled assay gives data for a PPTase in relation to its ability to activate BpsA (in particular the CP domain of BpsA), it does not provide any

information about the ability of a PPTase to activate other CP domains. For this an additional assay also developed in our lab, the BpsA competition assay, can be performed (Owen et al., 2011). In this assay a PPTase is incubated with BpsA, a limiting amount of CoA, and varying amounts of the target CP domain, which then "compete" with BpsA for the limited pool of CoA. The amount of holo-BpsA at each concentration of competing CP is determined via analysis of the indigoidine synthesis rate (Figure 5.1). An $IC_{50}$ can be calculated, which represents the CP domain concentration at which the CP is able to compete for 50% of the available CoA, i.e. where the rate of activation of both BpsA and the competing CP by the PPTase is equal. Where a PPTase has a greater preference for the competing CP domain over BpsA, then the $IC_{50}$ value will be lower. If a PPTase was unable to activate the competing CP domain then this would manifest as the velocity of the indigoidine synthesis reaction remaining constant, independent of the concentration of competing CP domain.

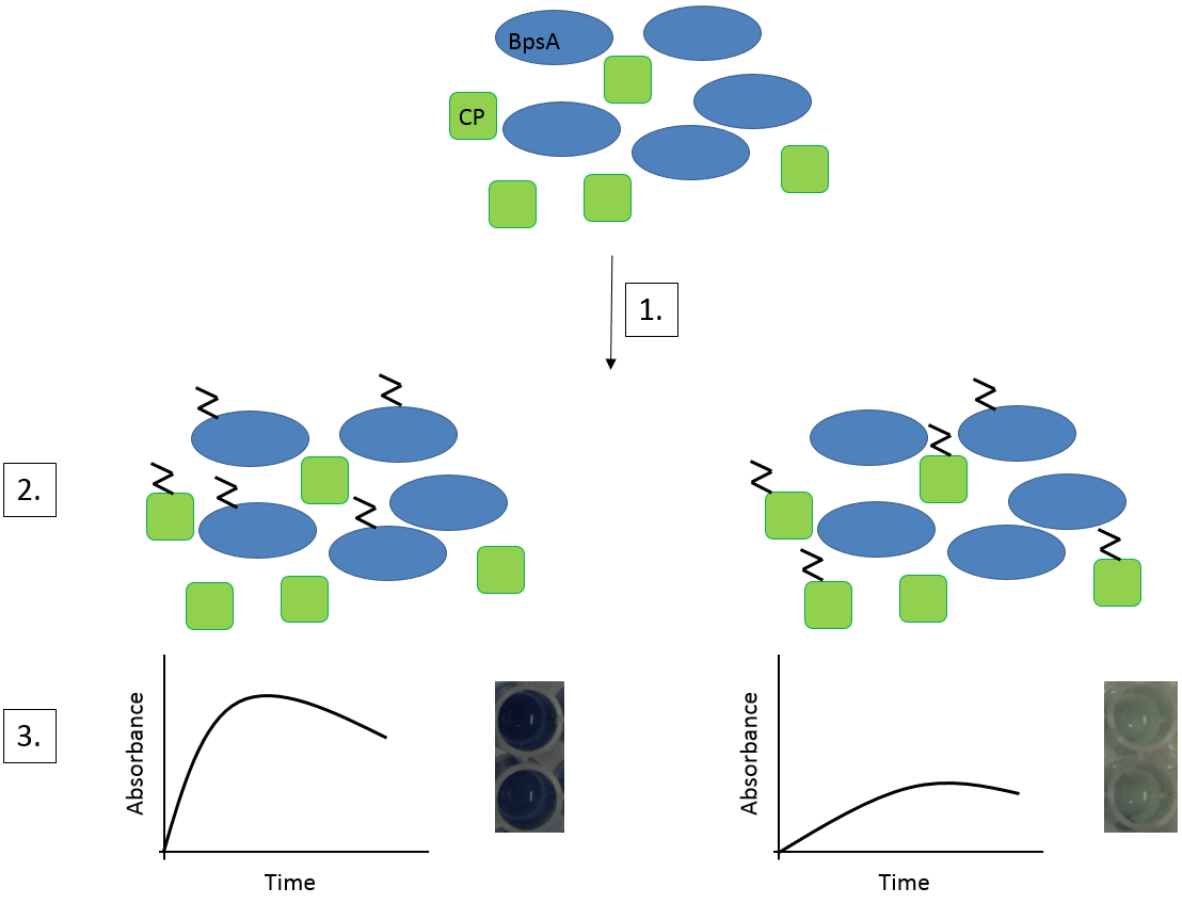



**Figure 5.1 BpsA competition assay scheme.** (1) A solution containing a mixture of BpsA (Blue) and a foreign CP (Green) is incubated with a PPTase and a limiting amount of CoA. (2) The PPTase transfers all available phosphopantetheine from the limited CoA to the BpsA enzyme and/or free CP, the amounts of each depend on the CP specificity preference of the particular PPTase (e.g. a preference for BpsA (left), or for the free CP (right)). (3) Amount of activated BpsA present is determined via the rate of indigoidine synthesis, from which the preference of the PPTase for any CP in comparison to BpsA is derived.

### 5.1.3.1 Candidate CP domain section for competition assays

The competition assay allows assessment of PPTase activity with different CP substrates. For the work described in this chapter, a range of CP domains were selected for use in these assays: both type I and type II ACPs and PCPs. CP domains were also selected on the basis of having been purified before, and found to remain primarily in the apo-form when expressed in *E. coli*. Already available as cloned individual CPs in the lab were the individual CP domain from BpsA (BpsA_CP) and the first CP domain from PvdD (PvdDCP1), both type I PCPs. Additionally, the CP from EntF (type I PCP) was cloned as well. As these were all type I PCPs, two type II PCPs were also selected; BlmI, a type II PCP from the bleomycin synthetase NRPS (Du and Shen, 1999) and SgcC2,a type II PCP from PKS responsible for the synthesis of the enediyne C-1027 (Van Lanen et al., 2005). Two ACP domains (a type I and a type II) were also chosen. MtaF is a type I ACP from a hybrid NRPS/PKS involved in myxothiazol synthesis (Gaitatzis et al., 2001) and TcmM is a type I ACP from the PKS responsible for tetracenomycin synthesis (Shen et al., 1992). BlmI and TcmM had been used as substrates for PPTase kinetic analyses before (Huang et al., 2006; Sanchez et al., 2001) whereas SgcC2 and MtaF had not. All CP domains used in this study are summarised in Table 5.1.

**Table 5.1 Carrier protein domains used in this study**

| Carrier protein | Type | Species of origin |
|---|---|---|
| BlmI | Type II PCP | *Streptomyces verticillus* |
| SgcC2 | Type II PCP | *Streptomyces globisporus* |
| MtaF | Type I ACP | *Stigmatella aurantiaca* |
| TcmM | Type II ACP | *Streptomyces glaucescens* |
| EntF | Type I PCP | *E. coli* |
| PvdDCP1 | Type I PCP | *P. aeruginosa* |
| BpsA_CP | Type I PCP | *S. lavendulae* |

## 5.1.4 Additional PPTases selected for characterisation

A number of additional PPTases were selected for characterisation using the BpsA based assays once it became apparent that the majority of the eDNA PPTases had very low activity (too low to perform the competition assay with). Sfp was chosen as an internal control, as the most comprehensively analysed of the previously studied PPTases. Sfp has a broad CP substrate specificity, and has been shown to be active with ACPs, PCPs, and ArCPs from a number of different species. It is the PPTase for which the most kinetic studies have been performed (Finking et al., 2002; Mootz et al., 2001; Quadri et al., 1998; Sanchez et al., 2001). The next PPTase chosen for further characterisation was Svp, from *Streptomyces verticillus*. This has been shown to be another broad acting enzyme that has activity with both PCP and ACP domains, and is also one of the few

PPTases for which a comprehensive kinetic analysis has been performed (Huang et al., 2006; Sanchez et al., 2001), which found it to have 346 fold greater activity than Sfp with a particular ACP substrate (Sanchez et al., 2001). Svp has since become more widely used for homologous expression of NRPS and PKS enzymes, and was used for the original homologous expression of BpsA (Takahashi et al., 2007). MtaA, from *Stigmatella aurantiaca* is also a broad specificity candidate, activating the NRPS/PKS hybrid (which contains both ACP and PCP domains) responsible for the production of the antibiotic myxothiazol. It has been shown to be able to activate both ACP, ArCP and PCP domains, although it has not been kinetically characterised before (Gaitatzis et al., 2001). Its broad range of activity has also led to its use in heterologous expression applications, including recently for the heterologous expression of the BpsA homologue, IndC, in *E. coli* (Brachmann et al., 2012). Gsp, from *Bacillus brevis* was discovered around the same time as Sfp, and has been found to complement Sfp (Borchert et al., 1994), as well as Lys5 (Mootz et al., 2002b) and AcpS (Mootz et al., 2001). The fact that it can complement type I and type II PPTases means it is regarded as having a broad substrate specificity, though it has never been tested directly for activity with a range of CP domain substrates, or kinetically characterised. It has however been used for the heterologous expression of NRPSs in *E. coli* (Ku et al., 1997). The final PPTase selected for characterisation in the present work was EntD, the PPTase involved in enterobactin synthesis in *E. coli*, and which was also one of the first classified PPTases (Lambalot et al., 1996). Lambalot et al. showed that *in vitro* EntD activated the PCP from EntF, and had a very low activity level with ACP, despite being unable to complement an AcpS deletion (Barekzi et al., 2004; Flugel et al., 2000). However, EntD has not previously been tested directly for activity with a wide range of CP substrates nor characterised kinetically. Additionally we examined the PPTase from *Pseudomonas syringae pv. phaseolicola* 1448A (PPT1448a), which had not been characterised before at all. The PPTases selected for study in this chapter are listed in Table 5.2.

**Table 5.2 PPTases used in this study**

| PPTase | Species of origin | Reference |
|---|---|---|
| Sfp | *B. subtilis* | (Quadri et al., 1998) |
| Svp | *S. verticillus* | (Sanchez et al., 2001) |
| Gsp | *B. brevis* | (Borchert et al., 1994) |
| MtaA | *S. aurantiaca* | (Gaitatzis et al., 2001) |
| EntD | *E. coli* | (Lambalot et al., 1996) |
| PPT1448a | *P. syringae* | This study |

### 5.1.5 Aims

- Clone each of the eDNA PPTases that was described in Chapter 4 into pET28a(+) and evaluate activity levels with available CP-substituted BpsA variants.
- Establish conditions for the purification of eDNA PPTases and characterise them *in vitro* using the BpsA-coupled kinetic assay and competition assay.
- Use BpsA based assays to further characterise the activity of a selection of other known PPTase enzymes with a range of CP domain substrates.

## 5.2 Results I: Characterisation of the activity and specificity of the eDNA PPTases

### 5.2.1 Activity of eDNA PPTases with three BpsA variants, as tested on solid media

In the research described in Chapter 4, the specificity of each of the eDNA PPTases was assessed by expressing the plasmids containing the entire eDNA insert (which had been identified as containing a PPTase gene) in each of the EcoBlue1, EcoBlueP and EcoBlueE strains. To objectively test the activity of each of the PPTases, the individual PPTase genes (hits Sw 1-7 and NZ 1-14) were isolated and cloned from their respective eDNA inserts into the pET28a(+) expression vector. For three of the PPTases (Sw 4 and 7, and NZ 3), it was unclear where the exact ORF for the gene started, with each of these three hits having two potential start codons. These PPTases were cloned twice, with each potential start codon, resulting in a full length gene and either a potentially truncated or extended version (with the longer ORF referred to as “a” and the shorter as “b”).

The pET28(+) constructs containing each of the eDNA PPTase genes were then used to transform each BpsA variant strain (EcoBlue1, EcoBlueP, and EcoBlueE). Each of the resulting strains was then plated on auto-induction plates at a density of approximately 100 colonies per plate, and the amount of time taken for pigment production was recorded. The results for this experiment are shown in Figure 5.2, with the shade of blue indicating the activity levels of each PPTase within each BpsA variant strain, with darker blue indicating faster indigoidine production and therefore greater activity of a particular PPTase with the corresponding CP domain substrate.

| PPTase | Relative level of activity in | | |
|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE |
| Sw 1 | | | |
| 2 | | | |
| 3 | | | |
| 4a | | | |
| 4b | | | |
| 5 | | | |
| 6 | | | |
| 7a | | | |
| 7b | | | |
| | | | |
| NZ 1 | | | |
| 2 | | | |
| 3a | | | |
| 3b | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |

**Figure 5.2 Analysis of relative activities for cloned eDNA PPTases in the three CP-substituted BpsA reporter strains EcoBlue1, EcoBlueP and EcoBlueE.** Each strain (EcoBlue1, EcoBlueP and EcoBlueE) containing each of the cloned eDNA PPTases was plated out at a density of approximately 100 colonies per plate on auto-induction pigment producing plates and incubated at room temperature after overnight growth. Dark blue shading indicates indigoidine synthesis was seen within the first 24 h, mid blue indicated it was seen at 24-48 h and for light blue indigoidine synthesis was seen after 48 h. White indicates no pigment production after four days of observation.

The levels of activity recorded here for each PPTase with the three different CP domain substrates were different from the levels of activity that were described in Chapter 4 when plasmids containing the entire eDNA insert were tested in a similar fashion. In most cases the activity level seen here for the cloned PPTases was either the same or improved when compared to the whole eDNA inserts. In some cases activity was seen here where none was observed for the corresponding construct previously. For example, when tested in the work described in Chapter 4, hits NZ 6 and 7 were completely inactive when expressed in EcoBlue1 and hits NZ 8, 9, and 10

were all completely inactive when expressed in EcoBlueE. Here, all of those combinations resulted in indigoidine synthesis. In fact, PPTase NZ 6 was so active with WT BpsA that indigoidine was produced in the EcoBlue1 strain within the first 24 h. The differences seen may reflect that some of the PPTases did not express well when present in the middle of a larger eDNA insert (perhaps due to inability of *E. coli* to effectively recognise the native gene promoter in the eDNA construct). However, possibly as a consequence of the increased levels of expression leading to heightened activity, none of the eDNA PPTases exhibited the same degree of specificity here that they had appeared to possess in the research presented in Chapter 4, with every eDNA PPTase clearly demonstrating at least some activity with every CP-substituted BpsA variant tested.

For each of the PPTases where two possible ORFs had been cloned, the shorter one in every case had less activity, indicating that the longer ORF was likely the correct gene sequence. For NZ 3 the truncated version (NZ 3b) did show a degree of specificity, having a small amount of activity when expressed in EcoBlueP alone; however this activity was so small overall it is unlikely to be of use in any downstream applications. For the other two PPTases (Sw 4 and 7), the longer (likely full length genes) did have more activity, but there was not a substantial difference between the two versions. The difference in lengths was greater for these two sets of genes than for the NZ 3 constructs (Sw 4a and 7a were 38 and 28 residues longer respectively, compared to NZ 3a being only 20 residues longer than NZ 3b).

### 5.2.2 Expression of eDNA PPTases

The eDNA PPTase pET28(a)+ constructs were used to transform BL21 (DE3) cells in order to enable purification of the PPTases for *in vitro* tests. PPTases are generally difficult enzymes to purify, with most expressed as insoluble proteins under standard protein expression conditions in *E. coli* (Huang et al., 2006; Rottier et al., 2013). Most PPTases require expression at low temperatures in order to obtain enough soluble protein. For each of the 21 eDNA PPTases (24 if including the truncated versions of Sw 4 and 7, and NZ 3) discovered in Chapter 4, multiple expression conditions were tested, in order to find a subset of PPTases that could be purified and studied *in vitro*. Initially several different chaperone strains were tested with three of the PPTases, however this approach did not result in greatly increased solubility compared to expression in BL21(DE3) at low temperatures and was not investigated further. Instead a variety of media conditions were tested for their ability to improve solubility at the standard low expression temperature of 18 °C. The first media condition investigated was the use of the media supplements glycine-betaine and sorbitol. High concentrations of sorbitol increase osmotic stress and promote uptake of glycine-betaine which acts as a chemical chaperone (Blackwell and Horgan, 1991; Oganesyan et al., 2007).

The use of auto-induction media was also tested; this method has been shown to greatly increase protein yields and can greatly increase the amount of soluble protein expressed (Studier, 2005).

Each of the eDNA PPTases was tested for expression levels on a small scale (in 10 ml cultures) with the following four expression conditions: LB media at 37 °C, LB media at 18 °C, LB media with 1 M sorbitol and 2.5 mM glycine-betaine at 18 °C (each using IPTG induction at a final concentration 0.5 mM), and auto-induction media at 18 °C. After expression, cultures were lysed by French-press and soluble and insoluble fractions were separated by centrifugation. The insoluble fractions were subsequently re-solubilised in 6 M urea and samples were analysed on SDS-PAGE gels. The media conditions that resulted in the highest amounts of soluble protein varied for each PPTase, though 18 °C was clearly the best temperature for expression with none of the proteins soluble at 37 °C. Figure 5.3 shows a representative example of the SDS-PAGE gels for three PPTases (Sw 3, NZ 1, and NZ 7) expressed under each of the four conditions. Sw 3 (Fig 5.3 A) was insoluble under all conditions, and was never able to be purified. Both NZ 1 and NZ 7 (Fig 5.3 B and C) were able to be expressed as soluble proteins in all media conditions at 18 °C, and were both later expressed on a large scale (in 400 ml) in LB at 18 °C for purification.

Ultimately, of the 24 eDNA PPTases tested (which included the three truncated genes), sufficient levels of soluble protein to continue with large scale expression and purification were observed for 13.

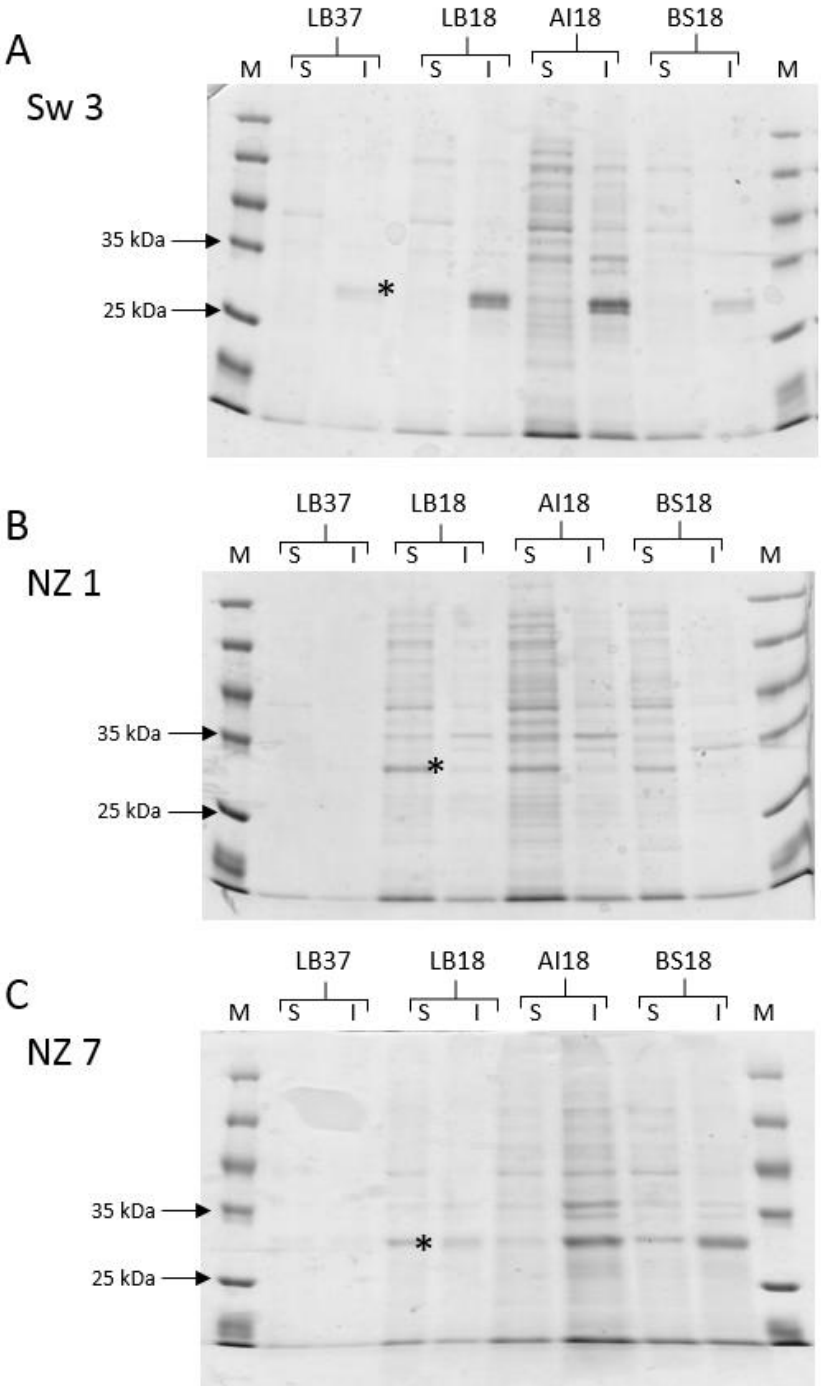



**Figure 5.3 SDS-PAGE gels showing relative soluble and insoluble protein expression levels for eDNA PPTases Sw 3, NZ 1 and NZ 7 under four different expression conditions.** The eDNA PPTases A) Sw 3, B) NZ 1 and C) NZ 7 were expressed in 10 ml cultures under four different conditions: in LB media at either 37 °C or 18 °C (LB37 and LB18), auto-induction media at 18 °C (AI18), and in LB media with glycine-betaine and sorbitol as supplements at 18 °C (BS18). With the exception of the auto-induction media, expression was induced by addition of IPTG to a final concentration of 0.5 mM. Cells were lysed via French press and soluble (S) and insoluble (I) fractions were separated via centrifugation. Resultant fractions were run on an SDS-PAGE gel. The predicted molecular weights (based on gene sequence) of each protein is: Sw 3- 30.4 kDa; NZ 1- 28.5 kDa and NZ 7- 30.4 kDa. Asterisks indicate the size of each protein.

### 5.2.3 Purification of PPTases

Where adequate levels of soluble protein were seen during small scale expression tests, the corresponding conditions were employed for the large scale (400 ml culture) expression of the protein, which was then purified using nickel affinity chromatography. Cells were lysed using the French press, as detergent methods for lysis were found to result in non-functional protein. Amendments to the standard nickel affinity chromatography method for the purification of PPTases are as described in section 2.8.5.2. The thirteen PPTases that were purified are displayed on an SDS-PAGE gel (Figure 5.4). Only one hit from the Swedish soil library (Sw 4 in both its full-length and truncated forms) could be purified. There were eleven hits from the NZ soil library that were purified. For all PPTases, relatively low amounts of protein were purified, with concentrations of final 4 ml protein preparations ranging from approximately 5 µM to 30 µM.

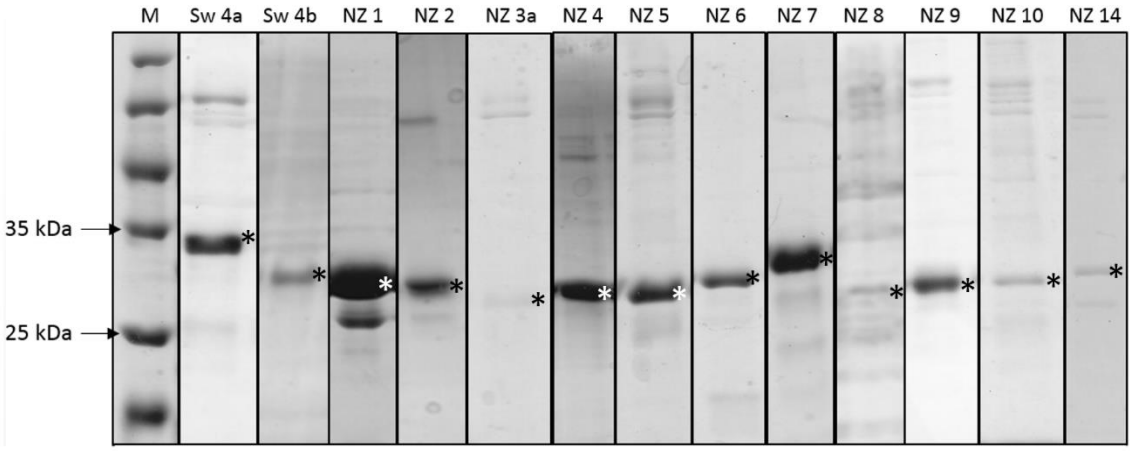



**Figure 5.4 SDS-PAGE analysis of purified eDNA PPTases.** SDS-PAGE analysis of eDNA PPTases immediately following purification and buffer exchange. NZ 1, 6, 7 and 14 were purified using LB18; NZ 2 was purified with BS 18, the rest were purified using AI18. The predicted molecular weights based on gene sequences are as follows: Sw 4a, 32.7; Sw 4b, 28.5; NZ1, 28.5; NZ 2, 28.4; NZ 3a, 27.8; NZ 4, 30.5; NZ 5, 27.6; NZ 6, 30.9; NZ 7, 30.4; NZ 8, 29.8; NZ 9, 33.9; NZ 10, 33.0; NZ 14, 29.7. Asterisks indicate the size of each protein.

### 5.2.4 Preliminary activity of purified PPTases

Before PPTases were evaluated in kinetic experiments, they were first tested for preliminary activity to ensure they were active enough for these further tests. In order to accurately measure kinetics for a PPTase using BpsA, there needs to be a change in indigoidine synthesis over time, i.e. an acceleration in the reaction (Owen et al., 2011). If the activity of a PPTase was too low, then the rate of indigoidine synthesis appeared to be steady (i.e. the $OD_{590}$ values as measured over

time appear as a straight line, with no acceleration increase towards a peak value) and accurate kinetics could not be determined.

Preliminary tests were carried out as described in section 2.11.3, with purified PPTase enzymes set up as a serial dilution series, mixed with BpsA and the reagents required for the indigoidine synthesis reaction, and then the $OD_{590}$ values were measured over time. An example of the results for three separate enzymes is shown in Figure 5.5. The decrease in absorbance values once a peak is reached is characteristic of the indigoidine synthesis curve, and may be due to a precipitation of indigoidine. Only the absorbance values before this peak is reached are used in any analysis. PPTase NZ 9 appears to be very active with the BpsA CP domain, resulting in a quick increase to a high maximum $OD_{590}$ value. The other two PPTases appear to be less active, resulting in much slower reactions. The results cannot be compared directly due to the enzyme concentrations being different (the highest concentration possible was used for each purified PPTase, to maximise the often low enzyme activities). For instance the concentration for NZ 10 was almost twice that of the other two PPTases, indicating that it may be an even weaker enzyme than it appears in Figure 5.5.

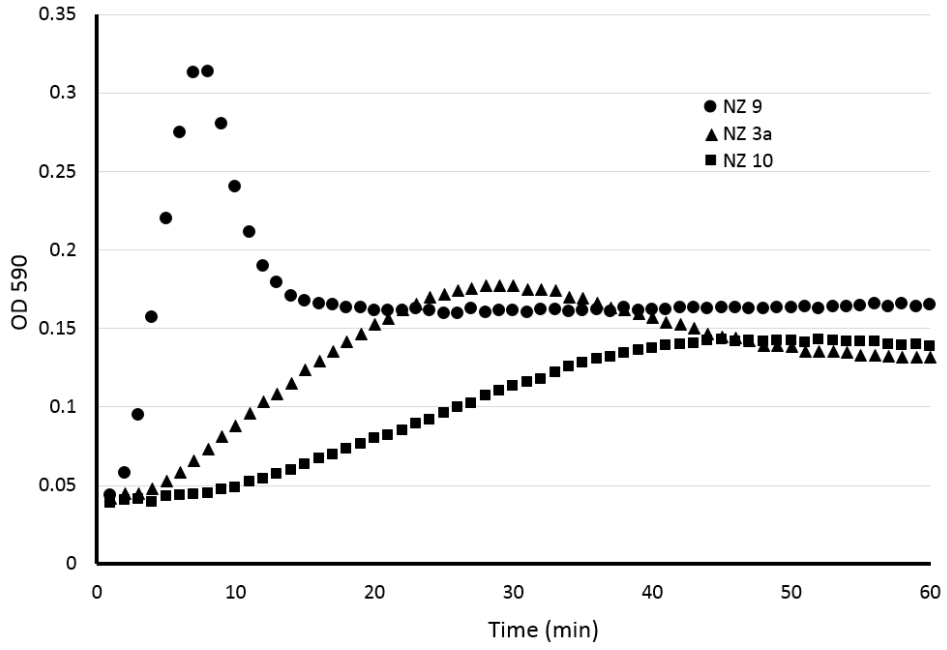



**Figure 5.5 Example of the preliminary activity assay for three eDNA PPTases (NZ 9, NZ 3a and NZ 10).** Preliminary activity assays were performed as described in section 2.11.3. The raw absorbance values at $OD_{590}$ as measured every min are displayed for three PPTases (NZ 9, NZ 3a and NZ 10). Data is shown for one replicate of a triplicate data set. The concentration of NZ 9 used was 1.8 µM, NZ 3a was 1.6 µM and NZ 10 was 3.5 µM.

There are several possible ways to quantify the activity of each enzyme in order to identify those with enough activity to perform a kinetic analysis (for example the peak $OD_{590}$ value, the slope of the curve, or the time taken to reach the peak $OD_{590}$ value). When each of these measures was compared they were found to give fairly consistent rankings from most active to least active PPTase for the tested enzymes (data not shown). As a simple measure for relative overall activity of the PPTase the "time to reach peak $OD_{590}$" was recorded. The time for each of the purified PPTases to reach their peak $OD_{590}$ as measured in the preliminary activity assay is listed in Table 5.3, ranked for activity from most active to least active. NZ 9 was the fastest of all the enzymes measured, with the times varying from 7.5 min (NZ 9) to 65 min (NZ 14). NZ 6, 7, and 8 are not displayed in this table as they displayed no activity over the time period measured even at the highest PPTase concentration.

**Table 5.3 Time for PPTase to reach peak $OD_{590}$ in preliminary activity assay**

| PPTase | Time at which peak was reached (min) |
|---|---|
| NZ 9 | 7.5 |
| NZ 1 | 9 |
| Sw 4a | 19 |
| NZ 2 | 24 |
| NZ 5 | 26 |
| NZ 3a | 28 |
| NZ 10 | 45 |
| Sw 4b | 59 |
| NZ 4 | 62 |
| NZ 14 | 65 |

### 5.2.5 Kinetics for the most active eDNA PPTases via BpsA assay

The top six hits as determined in section 5.2.4 were characterised further, using the BpsA assay, to derive kinetic parameters as described in section 2.11.4. The PPTase Sw 4b was also tested, in order to compare it to Sw 4a. The Michaelis-Menten curves for each of the top three enzymes are given in Figure 5.6, with separate graphs shown for the most active hit (NZ 9, Fig 5.6 A) and the next two most active hits (NZ 1 and 2, Fig 5.6 B), due to the large difference in velocity values. The remaining four enzymes tested had substantially lower velocities and are not illustrated here. The kinetic parameters for all seven enzymes (as measured with CoA as the variable substrate), are given in Table 5.4.

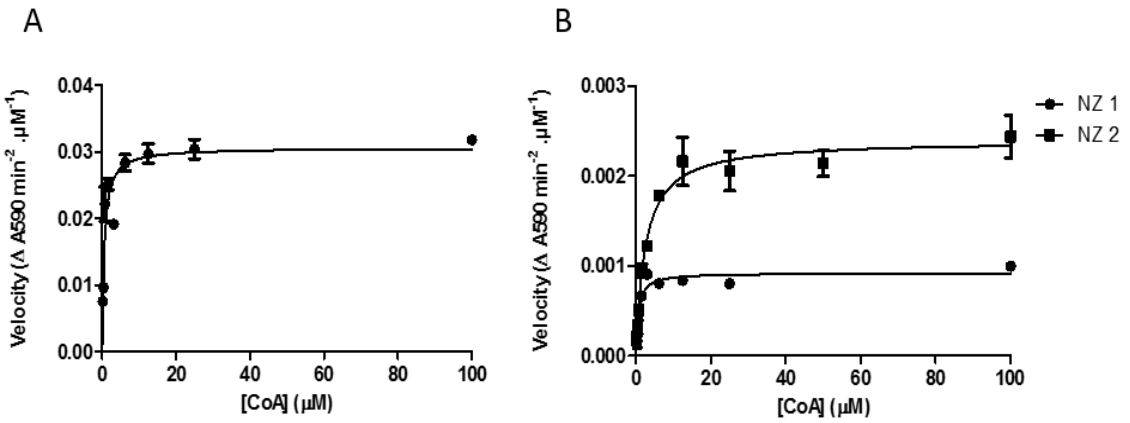



**Figure 5.6 Michaelis-Menten curves for the eDNA PPTases NZ 9 and NZ 1 and 2.** Plots indicate the maximum velocity of indigoidine production at different concentration of CoA for A) NZ 9 and B) NZ 1 and NZ 2. Curves were fitted using Graphpad Prism® and used to derive kinetic parameters for each PPTase. Data points were generated from three repeats and error bars are +/- SEM.

**Table 5.4 Kinetic parameters of the most active eDNA PPTases with CoA as variable substrate**

| PPTase | $V_{max}$ ($\Delta A_{590}.min^{-2}.mM^{-1}$) | $K_m$ (μM) | $V_{max}/K_m$ ($\Delta A_{590}.min^{-2}.\mu M^{-1}.mM^{-1}$) |
|---|---|---|---|
| NZ 9 | 30.6 ± 1.14 | 0.56 ± 0.11 | 54.5 ± 10.4 |
| NZ 1 | 0.11 ± 0.033 | 0.63 ± 0.11 | 1.5 ± 0.26 |
| NZ 2 | 2.40 ± 0.088 | 2.5 ± 0.40 | 0.96 ± 0.16 |
| Sw 4a | 0.045 ± 0.0027 | 1.28 ± 0.36 | 0.037 ± 0.010 |
| Sw 4b | 0.026 ± 0.0013 | 1.13 ± 0.28 | 0.024 ± 0.0060 |
| NZ 5 | 0.11 ± 0.015 | 21.89 ± 7.72 | 0.0052 ± 0.0020 |
| NZ 3a | 0.017 ± 0.00087 | 4.59 ± 0.91 | 0.0037 ± 0.00076 |

PPTase NZ 9, identified in the preliminary tests as potentially the most active of the eDNA PPTases, did have the highest $V_{max}/K_m$ (i.e., catalytic efficiency) compared to the other PPTases tested, and by this measure was 36 fold more active than the next most active PPTase (NZ 1) and more than 10,000 fold more active than the least active PPTase that was tested (NZ 5). In general relative comparisons of the catalytic efficiency to the initial method of ranking the most active enzymes (the time to reach a peak $OD_{590}$ value) were very consistent, and it appears that using this measure as an approximation of activity was justified. The main discrepancy is that Sw 4b had a higher catalytic efficiency than both NZ 5 and NZ 3a, when they had both reached the peak $OD_{590}$ value in half the time that Sw 4a did. This may be explained by the fact that NZ 5 and NZ 3a had relatively high $K_m$ values, and in the preliminary tests the CoA concentration used was well above both of these $K_m$ values. Somewhat surprisingly Sw 4a and 4b had very similar catalytic efficiencies, despite 4b being a truncated product and showing less activity with WT BpsA in the *in vivo* assay (section 5.2.1), taking an extra 24 h for indigoidine production when expressed in EcoBlue1.

### 5.2.6 Assessment of NZ 9 activity with additional CP domains via the competition assay

The eDNA PPTase that displayed the highest level of activity with BpsA, NZ 9, was examined further using the competition assay in order to identify which other CP substrates it could activate. NZ 9 was tested against the CPs from PvdD and EntF, and also four other CP substrates (see section 5.1.3.1). $IC_{50}$ values with each CP are given in Table 5.5. $IC_{50}$ values were derived for every CP except for TcmM, as the data for this assay did not fit an $IC_{50}$ curve.

The curve generated with TcmM is shown alongside the curve for the CP domain that gave the lowest $IC_{50}$ value (MtaF), in Figure 5.7 (the low $IC_{50}$ indicates that NZ 9 has a preference for the MtaF domain (which is an ACP domain) as a substrate as compared to the BpsA CP domain). This preference for the MtaF CP domain does not appear to be a general preference for ACP domains as substrates, as the ACP TcmM was the CP domain with which NZ 9 had the least activity (too low for an $IC_{50}$ value to even be calculated). Perhaps if higher still concentrations of TcmM had been available then the $IC_{50}$ could have been calculated. Nonetheless, every CP domain tested in competition with BpsA for NZ 9 resulted in a decrease of indigoidine synthesis velocity, indicating that NZ 9 is a PPTase that has a broad substrate range.

**Table 5.5 $IC_{50}$ values for NZ 9 with various CP domains**

| CP domain | $IC_{50}$ (μM) |
|---|---|
| BlmI | 5.7 (5.4-6.2) |
| SgcC2 | 1.2 (1.1-1.3) |
| MtaF | 0.34 (0.32-0.36) |
| TcmM | Undefined* |
| EntF CP | 4.8 (4.3-5.3) |
| PvdDCP1 | 1.4 (1.3-1.5) |

*Too low for accurate measurement
Parentheses indicate 95% confidence intervals

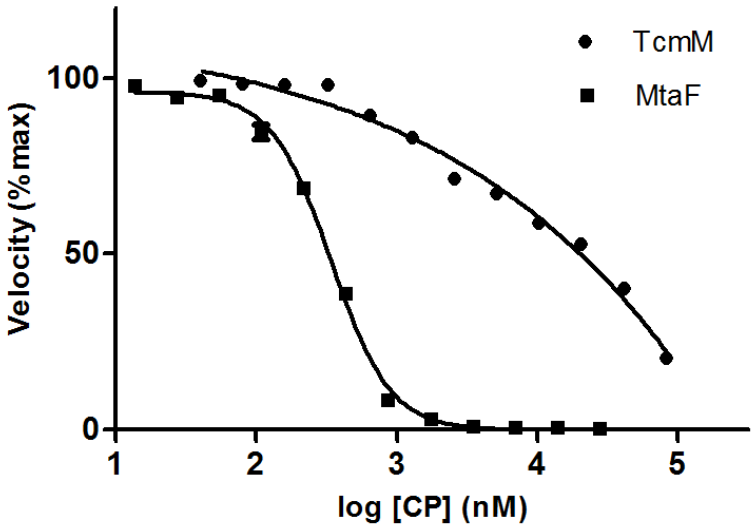



**Figure 5.7 $IC_{50}$ curves for NZ 9 with TcmM and MtaF as the competitive CP domains.** Plots shown indicate indigoidine synthesis velocity values as a percentage of the maximum velocity for each PPTase/CP pair, at varying concentrations of the competing CP domain (either TcmM or MtaF) for the PPTase NZ 9. Data points were generated from three repeats and error bars are +/- one SEM. Four-parameter dose-response curves were generated using Graphpad Prism®, and used to derive $IC_{50}$ values for each CP domain.

## 5.3 Results II: Characterisation of the activity and specificity of additional PPTases

Due to the inability to purify many of the eDNA PPTases and the generally poor activity levels of those that were purified, it was decided use the BpsA based PPTase characterisation assays on a selection of additional PPTases. The PPTases were mostly previously described enzymes that had had some degree of characterisation performed, as described in section 5.1.4. Four of the PPTases (Sfp, Gsp, Svp and MtaA) were ordered as codon optimised genes to minimise any problems with soluble protein expression. The other two genes (EntD and PPT1448a) were cloned from gDNA and so were not codon optimised.

### 5.3.1 Solid media activity test with three BpsA variants

Although most of the PPTases tested here (with the exception of EntD and PPT1448a) were known from previous work to have a broad substrate specificity, all were subjected to a rapid initial test for the substrate specificity by expressing each in the EcoBlue1, EcoBlueP and EcoBlueE strains and assessing indigoidine production on auto-induction agar plates as described in section 5.2.1. The results for this are seen in Figure 5.8. Most of the PPTases had high activity in all strains, with

PPT1448a (the only previously uncharacterised PPTase here) having the lowest activity, despite still showing some degree of activity across all strains.

| PPTase | Relative level of activity in | | |
|---|---|---|---|
| | EcoBlue1 | EcoBlueP | EcoBlueE |
| Sfp | | | |
| Svp | | | |
| Gsp | | | |
| MtaA | | | |
| EntD | | | |
| PPT1448a | | | |

**Figure 5.8 Cross reactivity of PPTases in the three BpsA variant strains, EcoBlue1, EcoBlueP and EcoBlueE.** Each strain (EcoBlue1, EcoBlueP and EcoBlueE) containing each PPTase was plated out at a density of approximately 100 colonies per plate on auto-induction pigment producing plates. The level of activity for each cloned PPTase is represented by the shade of blue presented. Dark blue indicates indigoidine synthesis was seen within the first 24 h, for mid blue indigoidine synthesis was seen at 24-48 h and for light blue indigoidine synthesis was seen after 48 h.

### 5.3.2 Purification and preliminary activity of PPTases

Purification of the additional PPTases was more straightforward than was encountered with the eDNA PPTases, with all successfully expressed in large scale cultures (400 ml) in LB media at 18 °C. Proteins were purified via nickel affinity chromatography as in section 2.8.5.2. and analysed by SDS-PAGE (Figure 5.9). PPT1448a is the one exception that did not appear to have purified very well; the concentration of this protein was relatively low, which may in part be due to the fact that the gene encoding this PPTase was not codon optimised for expression in *E. coli*.

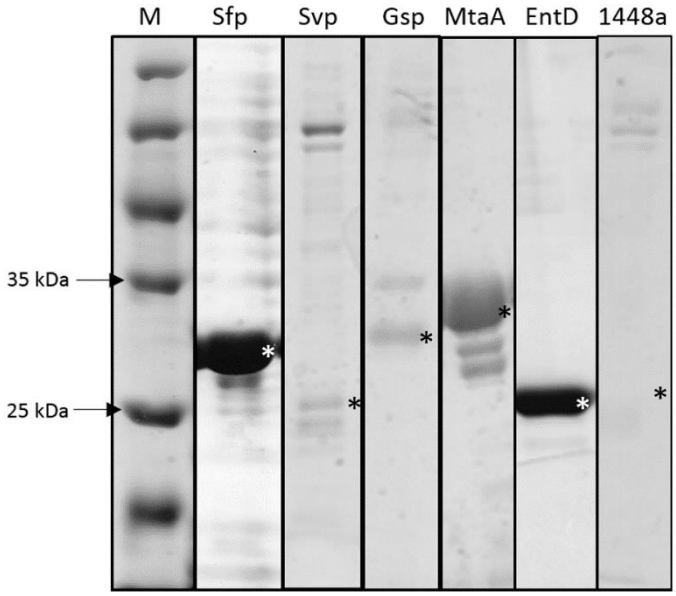



**Figure 5.9 SDS-PAGE analysis of purified PPTases.** SDS-PAGE analysis of PPTase proteins immediately following purification and buffer exchange. The predicted molecular weights based on gene sequences are as follows: Sfp, 28.3; Svp, 27.8; Gsp, 30.0; MtaA, 33.7; EntD, 25.8; PPT1448a, 25.6. Asterisks indicate the size of each protein.

For each of the purified PPTases, preliminary activity was tested as in section 5.2.4, and the time taken to reach a peak $OD_{590}$ value for each is given in Table 5.6. Three of the PPTases (MtaA, Sfp, Gsp) stand out as being substantially more active with the WT BpsA CP domain as substrate than the other three PPTases, reaching their peak $OD_{590}$ in under 10 min. Both EntD and PPT1448a were less active with WT BpsA in the *in vivo* test in section 5.3.1, so it is not unexpected that these were two of the less active enzymes. The seemingly low activity for Svp however was unexpected; Svp was the PPTase originally used in the heterologous expression of BpsA (Takahashi et al., 2007) and so has been shown to have activity with BpsA.

**Table 5.6 Time for PPTase to reach peak $OD_{590}$ value in preliminary activity assay**

| PPTase | Time at which peak was reached (min) |
|---|---|
| MtaA | 2 |
| Sfp | 3.5 |
| Gsp | 7.5 |
| EntD | 44 |
| Svp | 59 |
| PPT1448a | 89 |

### 5.3.3 PPTase kinetics measured via BpsA assay

Each of the PPTases was kinetically characterised for activity with WT BpsA, using CoA as a variable substrate. As expected the top three PPTases identified in the previous section had much higher activity levels than the remainder, and the Michaelis-Menten curves for these enzymes are presented in Figure 5.10. The corresponding kinetic parameters derived for each of the enzymes are listed in Table 5.7. For the enzyme PPT1448a, while activity was detected (i.e., slight increase in $OD_{590}$ values over a period of 90 min) the reaction was not fast enough to enable kinetic parameters to be calculated.

Sfp appeared to be the most active enzyme, with nearly four times the catalytic efficiency of the next most active enzyme, MtaA. The inability to derive kinetic parameters for PPT1448a was not surprising given the relatively low activity (Table 5.6) and the low amount of enzyme purified. This previously uncharacterised PPTase did however display the ability to activate BpsA, and perhaps if it had been possible to purify a greater amount of the protein then kinetics could have been measured.

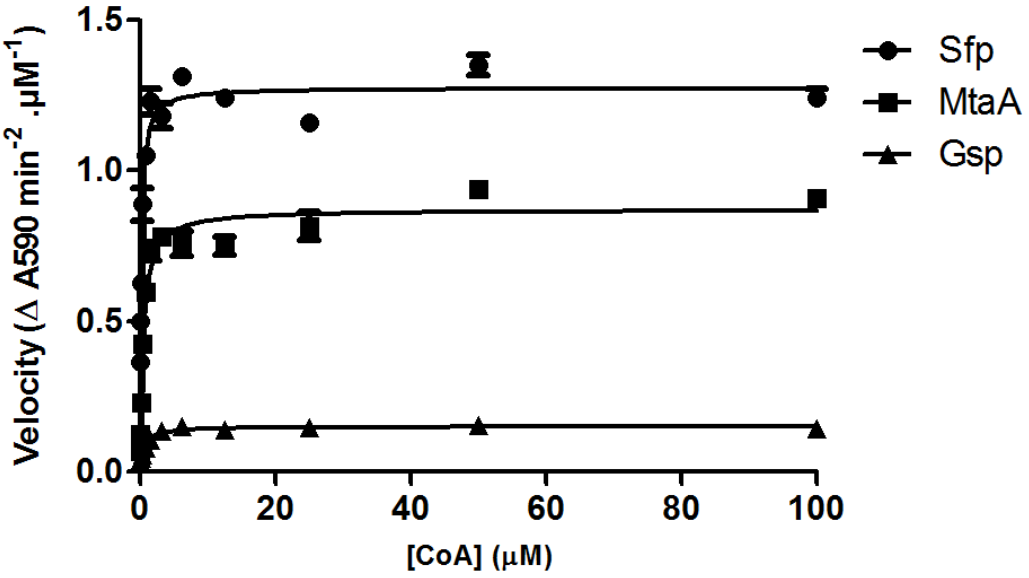



**Figure 5.10 Michaelis-Menten curves for Sfp, MtaA and Gsp.** Plots indicate the maximum velocity of indigoidine production at different concentrations of CoA for Sfp, MtaA and Gsp. Curves were fitted using Graphpad Prism® and used to derive kinetic parameters for each PPTase. Data points were generated from three repeats and error bars are +/- SEM.

**Table 5.7 Kinetic parameters for various PPTases with CoA as the variable substrate**

| PPTase | $V_{max}$ ($\Delta A_{590}.min^{-2}.mM^{-1}$) | $K_m$ (μM) | $V_{max}/K_m$ ($\Delta A_{590}.min^{-2}.\mu M^{-1}.mM^{-1}$) |
|---|---|---|---|
| Sfp | 1275 ± 17.3 | 0.161 ± 0.013 | 7919 ± 660 |
| Svp | 0.021 ± 0.0006 | 0.23 ± 0.053 | 0.094 ± 0.022 |
| Gsp | 150.4 ± 3.8 | 0.51 ± 0.069 | 292.7 ± 40.1 |
| MtaA | 870 ± 15.88 | 0.435 ± 0.043 | 2003 ± 200 |
| EntD | 5.3 ± 3 | 1.24 ± 0.28 | 4.3 ± 0.99 |
| PPT1448a | n.a | n.a | n.a |

### 5.3.4 Assessment of activity levels for Sfp, Gsp, MtaA and EntD with various CP domains via a competition assay

The four most active PPTases (having the highest catalytic efficiencies as described in Table 5.7) were each subjected to the CP domain competition assay to assess their activity with other foreign CP domains. Seven CP domains were tested for each PPTase, following the protocol in section 2.11.5. The $IC_{50}$ values for all possible PPTase/CP combinations are given in Table 5.8.

**Table 5.8 $IC_{50}$ values (μM) for PPTases as determined in the CP domain competition assay**

| CP | Sfp | Gsp | MtaA | EntD |
|---|---|---|---|---|
| BlmI | 2.3 (2.2-2.6) | 14.7 (10.9-20.0) | 6.5 (0.96-43.3) | 20.0 (6.6-60.1) |
| SgcC2 | 0.68 (0.65-0.71) | 0.55 (0.52-0.59) | 1.3 (1.1-1.4) | 2.0 (1.2-3.3) |
| MtaF | 0.40** | 0.76 (0.71-0.81) | 1.5 (1.3-1.7) | 18.8 (2.5-139.2) |
| TcmM | 39.4 (30.6-50.7) | 22.4 (19.6-25.6) | Undefined* | 10.2 (3.4-16.5) |
| EntF | 6.0 (5.5-6.7) | 5.2 (3.4-7.9) | Undefined* | 2.1 (1.5-3.1) |
| PvdDCP1 | 1.8 (1.6-2.0) | 0.92 (0.87-0.97) | 0.86 (0.69-1.08) | 1.9 (1.0-3.3) |
| BpsA CP | 0.51 (0.49-0.54) | 0.48 (0.44-0.52) | 0.68 (0.64-0.71) | 7.3 (3.6-15.0) |

* Too low for accurate measurement
**Confidence interval unable to be calculated
Parentheses indicate 95% confidence intervals

All of the PPTases tested were active to some degree with each of the CP domains, although the degree to which they were active with each did vary. In Figure 5.11, for each PPTase $IC_{50}$ curves are displayed for three representative competing CP domains to demonstrate the range of activities that were seen with different CP domain substrates. An unusual result was seen for Sfp and the CP domain MtaF (Fig 5.11 A). At low MtaF concentrations, BpsA activation was mostly complete, represented by 100% maximum velocity. Above the $IC_{50}$ value (at the increased CP concentration), the velocity of the reaction dropped to 0%. This indicates that above a certain concentration of the MtaF CP there was no activation of BpsA at all, indicating Sfp had a powerful preference for MtaF over BpsA. Overall Sfp gave relatively low $IC_{50}$s with most CP domains,

indicating an ability to activate all of the competing CP domains, as would be expected for this PPTase which has been widely described as having broad substrate specificity (Quadri et al., 1998). Gsp and MtaA also displayed relatively low $IC_{50}$ values with most competing CP domains indicating that they also had broad specificities, on a comparable level with Sfp. EntD had higher $IC_{50}$ values indicating it was less effective at activating this range of CP domain substrates. For EntD, the assay was also hindered by the fact that the activity of EntD with BpsA was so low (the catalytic efficiency differed from the other enzymes measured here by 2-3 orders of magnitude as measured in section 5.3.3), which is the reason for the large confidence intervals seen for the EntD $IC_{50}$ values in Table 5.8. Interestingly EntD had a larger $IC_{50}$ for the free BpsA CP domain compared to the other three PPTases, indicating that it may have a preference for activating CP domains (or at least the BpsA CP domain) when it is part of a complete NRPS enzyme, compared to the CP domain alone.

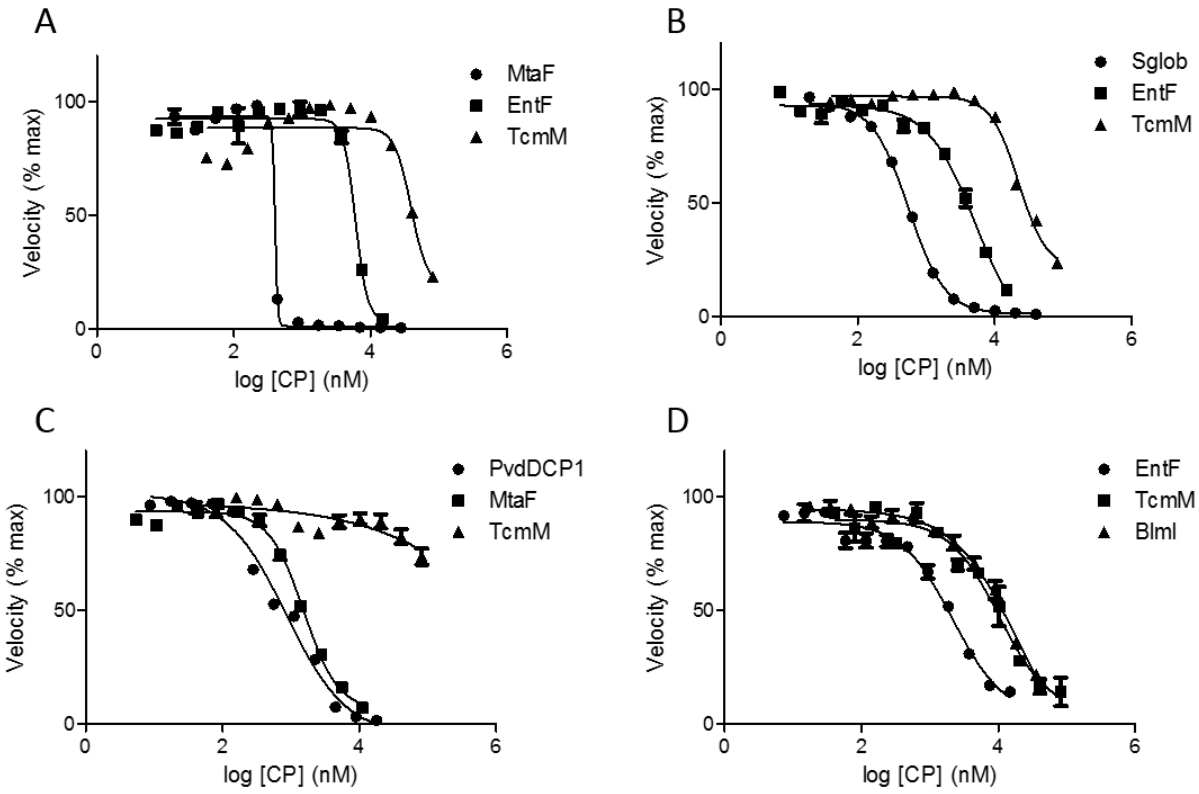



**Figure 5.11 Representative $IC_{50}$ curves for each PPTase with a range of competing CP domains.** Plots shown indicate indigoidine synthesis velocity values at different concentrations of a competing CP domain, presented as a percentage of the maximum velocity value for a given PPTase/CP pair, for the PPTases A) Sfp, B) Gsp, C) MtaA, and D) EntD. Data points were generated from three repeats and error bars are +/- one SEM. Four-parameter dose-response curves were generated using Graphpad Prism® and used to derive $IC_{50}$ values.

## 5.4 Discussion

In this chapter the eDNA PPTases discovered in the metagenomics bioprospecting study described in chapter 4 were characterised further based on their activity and specificity in a series of BpsA based assays. All of these eDNA PPTases were found to have a broad ranging activity with a number of different CP domain substrates. Only a subset of these PPTases were characterised *in vitro* due to issues with soluble protein expression, and the majority of those that were able to be purified had low activity in *in vitro* tests. Due to the inability to characterise all of the eDNA PPTases, the scope of the chapter was broadened and the same BpsA based assays were also used to characterise the activity of several other previously identified PPTases. The simplicity of these assays allowed for a greater examination of the activity and specificity of some of these enzymes than had previously been described. Some of these selected PPTases had been characterised before *in vitro*, to varying extents (ranging from the well-studied Sfp to the PPTase PPT1448a, which had not before been characterised as a PPTase). This research provided confirmation that PPT1448a displays PPTase activity *in vivo* and *in vitro* (though activity with BpsA was too low for full kinetic or competition assays to be performed).

The eDNA PPTases identified in Chapter 4 were tested here and found to be a lot less specific for certain CP domain substrates than they had first appeared to be, with expression of the individual cloned genes resulting in increased indigoidine synthesis in all three of the BpsA variant strains (EcoBlue1, EcoBlueP, and EcoBlueE) as compared to the expression of whole eDNA inserts. A contributing factor was likely that some of the PPTase genes may not have expressed well in their heterologous *E. coli* host when they were present in the middle of a larger eDNA insert, a factor in any functional metagenomic screening method (Gabor et al., 2004). The fact that activity was seen in at least one strain for each of the eDNA insert-containing plasmids indicated there was at least a minimal amount of PPTase expression from the eDNA plasmids. It was likely that this level of expression was the same across each of EcoBlue1, EcoBlueP, and EcoBlueE. The low expression levels of some of these eDNA inserts may have over-emphasised the extent of CP domain substrate specificity possessed by any individual PPTase. For example, in the present chapter it was shown that NZ 9 had a relatively low amount of activity with the EntF CP domain, both *in vivo* and *in vitro*, however this PPTase originally appeared to have no activity as an eDNA insert in the BpsA reporter strain EcoBlueE.

An important part of the metagenomic screening strategy presented in the previous chapter was having multiple CP-substituted BpsA variants, which led to the discovery of a wider range of

PPTases and thereby associated natural product biosynthetic genes. Here however it has been shown that that the inability to recover all of the eDNA PPTases using WT BpsA alone was not because some of these PPTases were entirely incapable of activating WT BpsA. In fact, all of the discovered eDNA PPTases were able to activate WT BpsA quite effectively when expressed at higher levels in *E. coli*. From the work described in the present chapter it now seems that some of these PPTases had only appeared unable to activate WT BpsA because their eDNA expression levels were too low (yet still sufficient for them to activate alternative CP-substituted variants, enabling their discovery in either EcoBlueP or EcoBlueE instead). This suggests that another approach for improving the BpsA metagenome screening method, rather than attempting to create and then screen eDNA libraries with multiple CP-substituted versions of BpsA, would be to increase the expression levels of the eDNA inserts, for example by using alternative hosts instead of *E. coli* (Gabor et al., 2004; Troeschel et al., 2010).

During the course of cloning the eDNA PPTases, three PPTases were cloned which had two different possible start codons, these were Sw 4 (construct 4a being 38 residues longer than 4b), Sw 7 (7a being 28 residues longer than 7b) and NZ 3 (3a being 20 residues longer than N3b). The fact that the longer genes all gave greater levels of activity (as tested *in vivo* for Sw 4a and 4b, Sw 7a and 7b, and NZ 3a and 3b (Fig 5.2), as well as *in vitro* for Sw 4a and 4b (the only pair able to be purified; Table 5.4)) suggested that the shorter versions were in fact truncated versions of the full-length genes. In two previous studies looking at N-terminal truncations of PPTases, the truncated proteins have been completely inactive (Huang et al., 2006; Jiang et al., 2013), suggesting this region is important for protein function. This was consistent with the results seen for NZ 3, where the truncated protein was almost completely inactive (displaying only a small amount of activity when expressed in EcoBlueP and no activity in the other two strains). It was not however consistent with the other two truncations, which had much longer sections of the N-terminus removed yet still retained some activity (including almost equal *in vitro* activity in the case of Sw 4a and 4b). This indicates that the N-terminus may not be strictly essential for PPTase function. It would be interesting to see if a compromise could be achieved whereby a portion of the N-terminus is removed to generate a loss of activity with one substrate but not another, thereby increasing the specificity of a target PPTase.

Overall the eDNA PPTases were generally difficult enzymes to purify and work with, though based on previous literature this is not entirely unexpected for PPTases. While a little over half were able to be purified, most of these had very low activity which made them difficult to characterise via the BpsA assays. One unusual result was the fact that three of the purified PPTases (NZ 6, 7 and

8) displayed no activity at all *in vitro*, even though all displayed a moderate level of activity with WT BpsA when expressed in EcoBlue1. For NZ 8 it may have been the case that not enough protein was purified, as only small amounts were detected. However for NZ 6 and 7 relatively large amount of protein were purified and yet absolutely no activation of BpsA was seen, even when assays were incubated for an extended time. This could indicate that these enzymes were highly unstable under the *in vitro* conditions that were employed.

Of the eDNA PPTases that were characterised *in vitro*, NZ 9 was substantially more active than all of the other eDNA PPTases (although PPTases NZ 1 and 2 did have reasonable activity levels as well). NZ 9 was also shown in the competition assay to be active with all CPs tested, with similar $IC_{50}$ values to the previously known broad specificity PPTases Sfp, Gsp and MtaA. This suggests that while none of the eDNA PPTases found may be particularly useful in applications that require highly specific PPTases, NZ 9 may be a good candidate for applications that require a PPTase with a broad range of activity, such as in the heterologous expression of NRPS or PKS genes.

Of the additional non-eDNA PPTases that were examined, the activities of most were consistent with their previously described abilities to activate a wide range of substrates. Sfp, the standard broad range PPTase had activity with every CP tested in the competition assay. This was not unique however and Gsp and MtaA also showed the ability to activate a wide range of other CP domains. Each of the PPTases showed a different pattern for the CP domains they were preferentially able to activate. For example, Gsp had the greatest preference for the CP domain SgcC2, compared to its ability to activate BpsA and the other CP domains it was tested with, whereas Sfp had the greatest preference for MtaF compared to the other CP domains it was tested with. EntD exhibited some unusual results, for instance having a relatively high $IC_{50}$ for the BpsA CP domain. A limitation of the results given for EntD is the low activity that EntD displayed with BpsA means these reactions were extremely slow, and as a result have a lot higher margin of error than all the other PPTases tested. Unfortunately with the data presented in this chapter it is not possible to compare between the PPTase enzymes directly, as they all have different baseline levels of activity with BpsA. To measure $k_{cat}/K_m$ for each PPTase/CP pair we would need to calculate the kinetic parameters for the PPTases with respect to the concentration of BpsA, not CoA as was performed here. This work will be performed in the future.

An unusual exception for the additional PPTases tested in the *in vitro* assays was Svp. This has previously been shown to be a reasonably active PPTase (when tested kinetically alongside Sfp, a similar or better amount of activity was seen with several CP substrates, notably far greater activity with the CP TcmM as a substrate (Sanchez et al., 2001)). Particularly relevant here, Svp

was the PPTase to be used in the first heterologous expression with BpsA, and was shown to have the ability to activate BpsA *in vitro* (Takahashi et al., 2007). Yet in the present study Svp did not show enough activity with BpsA to even be tested in the competition assay with a wider range of CP domain substrates. It would have been interesting to see if Svp had a low $IC_{50}$ value for the CP TcmM if we had been able to measure Svp in the competition assay, as all of the other PPTases tested with TcmM had a low degree of activity with it, compared to BpsA as measured in the competition assay. Svp did appear to be very active (with all CP-substituted variants) of BpsA when tested *in vivo*, which suggests that the purified protein may have been less stable under the conditions that were employed than the other proteins tested here.

Overall every purified PPTase tested in this chapter could activate every CP domain substrate tested to at least some degree, indicating it is difficult to find PPTases that are specific towards a particular CP domain (which is a property desired in many applications). The research described in next chapter therefore adopted a directed evolution approach in order to alter the CP substrate specificity of the PPTase, EntD.

# Chapter 6: Engineering the PPTase EntD for altered substrate specificity

## 6.1 Introduction

PPTases vary in the degree to which they are able to recognise and activate different CP domain substrates. Some have a narrow range of specificity for their native CP substrate, while others (for example the broad range PPTase Sfp) are active with a large number of different CP substrates. Generally type II PPTases have a relatively broad substrate range, often showing substantial activity with non-cognate CP domains (Beld et al., 2013), which was also consistent with the results presented in Chapter 5. There are certain applications for PPTases, such as site-specific protein labelling, where it would be advantageous to have PPTases that exhibit a narrower substrate specificity. Earlier in this work it was thought that such PPTases may have been found in Chapter 4, by virtue of the metagenomic screening method used (i.e. certain PPTases were only recovered when screening with BpsA variants containing a specific CP domain substitution). However, when these PPTases were examined in more detail in Chapter 5, all of them were found to be substantially less specific for individual CP substrates than preliminary results had suggested.

An alternative approach to obtaining more specific PPTases is to engineer an existing PPTase enzyme to modify its specificity. This chapter describes the development of a screen to alter PPTase substrate specificity, retaining activity with one CP substrate and losing activity with another. The *E. coli* type II PPTase EntD was chosen for this proof-of-principle study. One advantage of using EntD was that the WT enzyme expressed well and was easily purified (as opposed to many other PPTases, including many of the eDNA PPTases examined in Chapter 5). Another advantage was the availability of a *ΔentD E. coli* strain which was created during the course of this thesis (Section 4.2.1). Here, this was used to screen for variants that were able to complement the *ΔentD* deletion. Evolved EntD variants that retained activity with the endogenous *E. coli* enterobactin synthetase NRPS enzymes were selected under iron restricted conditions, on the basis that only host cells able to produce enterobactin siderophore would be able to grow. These variants were simultaneously counter-selected for a loss of activity with the BpsA CP domain, measured by a loss of pigment production at a colony level. It was hypothesised that this dual positive and negative selection would enable identification of PPTase variants that had evolved greater specificity for the native enterobactin NRPS CP domains.

### 6.1.1 PPTases for site specific protein labelling

Site-specific protein labelling is an approach that has a range of applications including the study of protein structure, dynamics and interactions (Foley and Burkart, 2007; Rashidian et al., 2013). One method for the site-specific modification of proteins is to use PPTases to add CoA-derived labels to CP-derived tags. In this method, a target protein is tagged with a CP domain or a smaller peptide tag that is also able to serve as a PPTase substrate, enabling the PPTase to attach a label derived from a modified CoA analogue. The types of CoA analogues that can potentially be used for this are varied, with many small molecules (e.g. fluorophores, biotin, quantum dots) able to be conjugated onto the thiol group of CoA (La Clair et al., 2004; Meier et al., 2006; Sunbul et al., 2008). The tethered small molecules are then transferred onto the tag by the PPTase. Many of the PPTases previously studied in this regard (including Sfp and EntD) have been found to be relatively flexible in the CoA analogues that they can accept (Chen et al., 2009; La Clair et al., 2004). In structural studies of Sfp bound to CoA, the 3'-phospho-5'-ADP moiety was shown to be important in CoA binding, while the Ppant moiety did not play a significant role, an observation thought to explain this promiscuity towards CoA analogues with Ppant modifications (Reuter et al., 1999). The CP domains themselves are relatively small (80-90 amino acids), and so their presence may be less likely to disturb protein function than many other common labelling proteins which are much bigger (for example GFP is 238 amino acids long) (Rashidian et al., 2013; Yin et al., 2004). More recent developments evolving the CP domain substrate have led to even smaller, less obtrusive peptide tags, just 10-12 amino acids in length (Yin et al. 2005; Zhou et al. 2007; Zhou et al. 2008). Another recent development of this technology has been the ability to reverse the labelling process, i.e. to remove labels by treatment with acyl carrier protein hydrolase (AcpH) (Kosa et al., 2012).

Currently the site specific labelling of proteins by PPTases is limited to at most a dual labelling system, employing two PPTases and two CP-derived tags (Figure 6.1). In the original version of this technology the more specific PPTase AcpS is used first to label an ACP substrate, followed by Sfp which has a broader substrate range and labels any remaining CP domains present in the reaction mix (primarily its own PCP targets but possibly also any ACP domains left unlabelled by AcpS) (Vivero-Pol et al., 2005). This technology has since been improved via development of smaller peptide tags, these being evolved so that their activating PPTase would have greater specificity with them, which resulted in the Sfp labelled S6 tag and the AcpS labelled A1 tag (Zhou et al. 2007; Zhou et al. 2008). Despite an increase in specificity there are still limitations with this technique, and some background non-specific labelling is still seen (Marchetti et al., 2014). The current

system is also limited to just two PPTase/tag pairs, only allowing the labelling of two protein substrates.

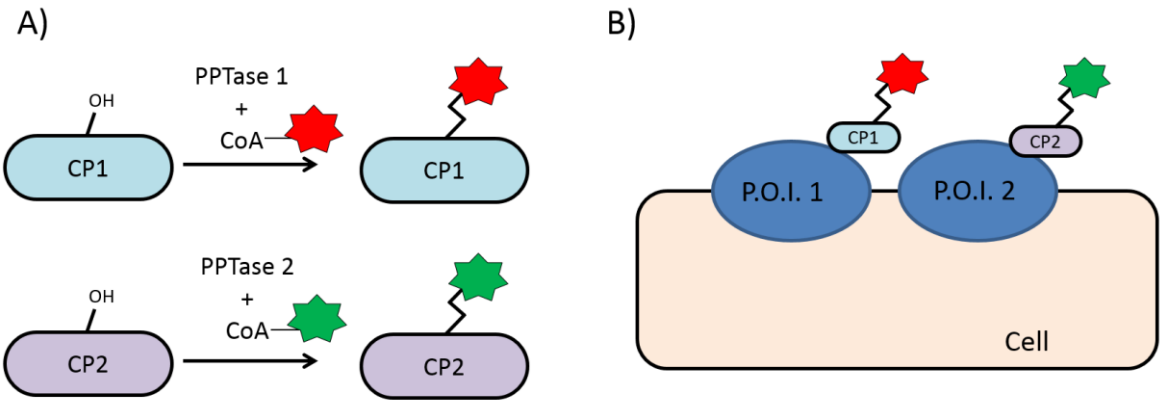



**Figure 6.1 Use of PPTases in site-specific protein labelling.** A) Two potential PPTases with non-overlapping CP-substrate specificities can label different CP tags with different fluorescent CoA analogues. B) Given two specific PPTase CP pairs, two proteins of interest (P.O.I) can be differentially labelled. Adapted from (Vivero-Pol et al., 2005).

It would be useful to have additional PPTases that exhibit a narrow substrate specificity, i.e. that also exhibit very little cross reactivity with other CP substrates. Past attempts to improve the specificity of the PPTase/tag pair have always evolved the CP substrate, with no attempt to evolve the PPTases themselves. The aim of this chapter was to develop a screen to evolve a target PPTase to have highly specific activity for a given CP substrate.

### 6.1.2 Engineering PPTases for altered substrate specificity

There has been one previous published study to engineer PPTase substrate specificity, evolving Sfp to alter its substrate specificity with CoA (Sunbul et al., 2009). Though Sfp had been shown to be promiscuous with many CoA analogues, 3'-dephospho CoA (dpCoA) was shown to be a poor substrate for Sfp (Quadri et al., 1998). In this evolutionary scheme Sfp was co-displayed on a phage surface alongside the small peptide tag ybbr. The addition of biotin-conjugated dpCoA meant mutant enzymes that were able to catalyse the reaction with dpCoA could be subsequently recovered via a streptavidin selection (with the gene encoding these enzyme variants also recovered within the phage particles). Sfp variants were created via site directed mutagenesis, with four sites rationally selected (residues 28, 31, 44 and 90) based on the crystal structure of Sfp (Reuter et al., 1999). The top mutant recovered (R4-4) exhibited 329 fold higher activity with dpCoA over CoA. However, this selectivity was only observed when labelling the ybbr tag, and no difference in selectivity was seen when labelling a full CP substrate. By chance during the course

of the experiment additional mutations (due to PCR errors) were recovered at two sites where the sequence had not been randomised, one of which proved to be beneficial. Beyond this, there have been no published attempts to engineer the CP substrate specificity of a PPTase.

#### 6.1.2.1 Proof-of-principle engineering of the EntD PPTase

For the development of a dual selection/screen that would serve as a proof-of-principle for being able to evolve PPTases to have altered CP substrate activity, the type II PPTase EntD from *E. coli* was chosen. It was reasoned that the *ΔentD E. coli* strain (created in Chapter 4) would allow for a powerful selection, resulting in the ability to interrogate larger variant libraries. The WT EntD enzyme is also able to be expressed as a soluble protein and purified, characteristics important for verifying activity via *in vitro* tests, as well as for enabling downstream site specific labelling applications.

The native role for EntD in *E. coli* is activation of the NRPS enzymes of the enterobactin synthesis pathway (specifically the EntB and EntF enzymes). Enterobactin is a siderophore that is involved with iron uptake, making it essential for *E. coli* growth under conditions of iron starvation, and non-essential under normal growth conditions. EntD is already regarded as a narrowly specific PPTase, having much greater activity for its own substrates rather than foreign CPs. For example, *entD* was not able to complement an *acpS* mutant in *E. coli* (Flugel et al., 2000). Also, the heterologous expression of many NRPS and PKS enzymes within *E. coli* usually requires the co-expression of a more broadly-acting PPTase (for example Sfp) for the production of an active NRPS/PKS enzyme (Kealey et al., 1998; Ku et al., 1997).

The aim of the research described in this chapter was to decrease or abolish the activity of EntD with the CP domain of BpsA, while retaining activity with its native *E. coli* CP domains, those of EntB and EntF. The WT EntD enzyme already exhibited a preference for the latter substrates; for example, it was shown earlier that EntD has a greater activity with its own native substrate EntF than that of WT BpsA (Chapter 5). The *ΔentD* strain of *E. coli* that was created in Chapter 4 was used to select EntD variants able to complement this mutation under iron restricted selection, by activating the native *E. coli* CP domains involved in enterobactin synthesis. Only variants that were still functional as PPTases able to activate the CP domains of the enterobactin synthesis pathway were able to grow, eliminating non-functional variants. An immediate selection step that eliminates non-viable PPTase enzymes is a huge advantage and means a larger number of variants can be screened more easily. This allowed the creation of a totally random variant library via error prone PCR. As there is no structural information available for EntD, a more targeted approach

would have been difficult. At the same time, to screen for mutants with improved selectivity the *entD* variant library was expressed in the EcoBlue1 strain and those that were able to grow under iron restricted selection were counter-screened for a loss of activity with the CP of BpsA, by selecting for reduction (or ideally, elimination) of indigoidine production. The directed evolution scheme is illustrated in Figure 6.2.

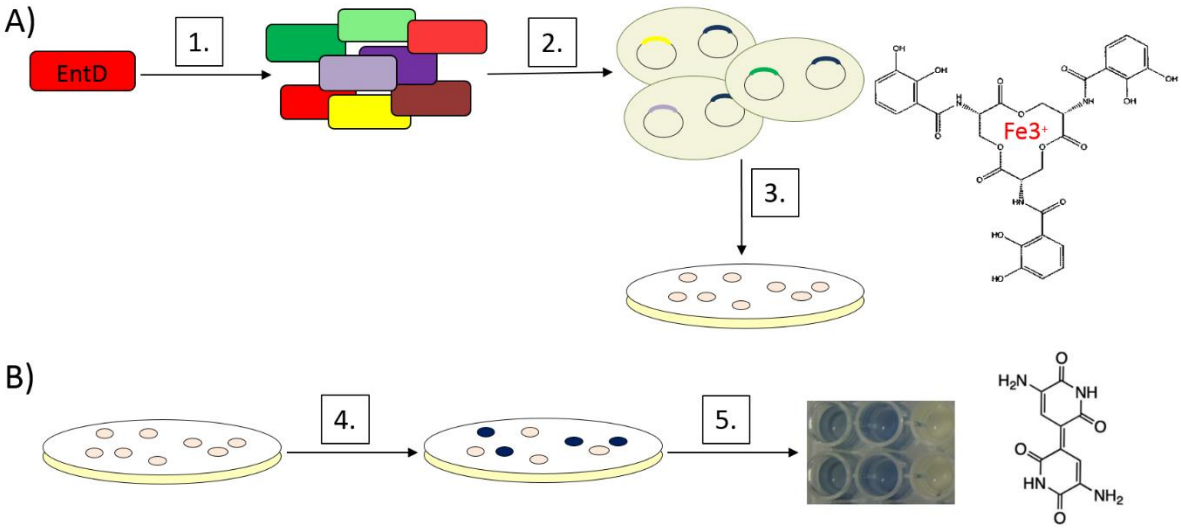



**Figure 6.2 Directed evolution scheme for EntD evolution.** A) Selection of *entD* variants that are able to activate the native CP domains of the enterobactin synthesis NRPS. (1) *entD* is amplified by error prone PCR to create a library of random variants. (2) *entD* variant library is used to transform cells harbouring BpsA. (3) Transformed cells are plated on iron restricted media, where only host cells expressing EntD variants that can still activate the CP of the enterobactin NRPS are viable. B) Selection for loss of EntD activity for BpsA. (4) Induction of BpsA via addition of IPTG. (5) Screening for white colonies (those that have lost the ability to activate BpsA and therefore cannot produce indigoidine).

### 6.1.3 Aims

- Develop a screen to select for EntD variants that retain activity with the EntF CP domain and lose activity with the BpsA CP domain.
- Verify that the EntD variants recovered have altered activity levels with these CP substrates.

# 6.2 Results

## 6.2.1 Optimisation of EntD and BpsA expression under iron restricted conditions

### 6.2.1.1 Development of a co-expression system for EntD and BpsA

Previously in this research, the expression of *bpsA* and the activating PPTase genes was simultaneous, from different plasmids that were each under IPTG inducible control (i.e., the pET plasmid for the PPTase genes and the pCDFDuet plasmid for *bpsA*). Both genes were expressed using a delayed induction method to address the limitation that if activated BpsA is expressed and indigoidine is produced before colonies can form, then the mildly toxic effects of indigoidine prevents colonies from forming after transformation (Owen, 2010). For this screen however, constitutive expression of a functional EntD variant was required for growth. In order for a BL21 *ΔentD* transformant to grow on iron restricted media, the enterobactin biosynthesis genes would need to be activated by an EntD variant capable of recognising its cognate substrates, i.e. the CP domains of EntF and EntB, to enable production of enterobactin and thus survival under iron restricted selection. At the same time, delayed induction of BpsA was still required to avoid the indigoidine toxicity problem, so plasmids with separately inducible controls for *entD* and *bpsA* genes were required. As delayed expression systems for BpsA (either autoinduction or late addition of IPTG under scooped agar plates (section 2.10.1)) had already been optimised, *entD* was cloned into a different expression system. The *entD* gene was also the smaller of the two and easier to clone. Due to availability and its ability to be maintained together with pCDFDuet in the same host cell, the pBAD plasmid (which induces genes via an arabinose promoter) was chosen for EntD expression.

Expression of pBAD:*entD* in both EcoBlue1 and EcoBlueE was tested at different concentrations of arabinose (0.001%, 0.01% and 0.5%, w/v) in LB plates with BpsA induced via IPTG addition using the scoop method as described in section 2.10.1. At the lower concentration (0.001%) there was low indigoidine production overall, but with much more being produced in the EcoBlueE strain (compared to EcoBlue1). At the two higher concentrations (0.01% and 0.5%) similar levels of indigoidine production were seen in both the EcoBlue1 and EcoBlueE strains, with a faster reaction time seen at the higher concentration. At 0.001% (w/v) arabinose EntD was clearly better able to activate the EntF CP-substituted BpsA, exaggerating the preference that EntD already had for the

EntF CP-substituted BpsA compared to the WT BpsA. Based on these results, the higher concentration of 0.01% was chosen for final screening plates to more easily see changes in activity in the EcoBlue1 background.

#### 6.2.1.2 Optimisation of solid media for iron restricted growth and indigoidine production

Iron restricted plates consist of minimal media to which has been added the iron chelator 2,2'-dipyridyl (Roche and Walsh, 2003). For this work the usual carbon source for minimal media plates, glucose, was replaced with glycerol, as glucose represses pBAD expression. The optimal concentration of 2,2'-dipyridyl for distinguishing functional EntD variants was then tested. The *E. coli* strains BL21, BL21 *ΔentD*, and BL21 *ΔentD* pBAD:*entD* were each plated out at a density of approximately 100 colonies/plate at three concentrations of 2,2'-dipyridyl (100 µM, 50 µM and 10 µM) on minimal media agar plates containing 0.01% (w/v) arabinose, and growth seen at 48 h was recorded (Table 6.1). There was no growth at all on any of the plates after just 24 h. After 48 h, a concentration of 10 µM 2,2'-dipyridyl was not enough to see any growth difference between the three strains. A difference was seen at 50 µM, with normal growth for BL21 WT and reduced growth for both the *ΔentD* strains, but the presence of pBAD:*entD* did not make a difference. At 100 µM 2,2'-dipyridyl there was no growth at all in the knock out strain, and this was rescued by overexpressing *entD* in pBAD, at similar levels as the WT strain. It was subsequently decided to incubate plates for 72 h, at which point larger colonies were seen on the BL21 and BL21 *ΔentD* pBAD:*entD* plates, while the *ΔentD* strain still had no growth. This made it easier to see any difference in indigoidine levels between the larger colonies.

**Table 6.1 Growth of BL21 strains at three different concentrations of 2,2'-dipyridyl in minimal media agar plates containing 0.01% (w/v) arabinose after 48 h**

| ***E. coli* strain** | **2,2'-Dipyridyl concentration (µM)** | | |
|---|---|---|---|
| | **10** | **50** | **100** |
| BL21 | Normal | Normal | slow |
| BL21 *ΔentD* | Normal | slow | none |
| BL21 *ΔentD* pBAD:*entD* | Normal | slow | slow |

Next the expression and detection of indigoidine on iron restricted minimal media plates was optimised, using the strain EcoBlue1 pBAD:*entD*. L-glutamine at the previously optimised concentration of 100 mM was sufficient in these plates to enable indigoidine expression and caused no problems with growth. Some pigmentation was visible before induction, which may have been due to leaky expression of BpsA from the pCDFDuet plasmid. This may have been

caused by the absence of glucose from the media, which would normally counteract IPTG induction. This leaky expression was presumed to be minimal, given that the cells were still able to grow, and that induction was still required to get substantial indigoidine production. An issue that arose with induction was that due to the 72 h incubation of the plates, scooping the agar out to add the ITPG became very difficult due to the degree to which the agar had dried out. To solve this problem plates were made with a higher percentage of agar (1.7% compared to the usual 1.5%) and each plate was made from 25 ml of media rather than the usual 20 ml. Plates were also sealed with parafilm during the 37 °C incubation. Another problem encountered was that after induction, the degree of indigoidine production was non-uniform across the plate (e.g., Figure 6.3). All colonies are identical and should theoretically be producing the same level of indigoidine, yet the representative colonies in Figure 6.3 appear to be much darker in the middle of the plate and lighter around the sides. It was reasoned that this could be due to non-equal amounts of induction resulting from uneven spreading of the IPTG under the agar. These uneven pigment levels had been seen before in other screening methods (for example the eDNA screening in Chapter 4). This was not considered a huge problem at the time as in that screen any level of blue pigmentation was considered a hit. In the current screen however the level of pigmentation is important, as the desired phenotype is that of impaired pigment production. It was necessary to ensure any white or light blue hits found were a result of the EntD variant losing activity for the BpsA CP domain, and not because of the location of the colony being in an area of poor induction on the plate. I was unable to resolve the uneven pigment production levels, so each potential hit had to be evaluated relative to the pigmentation of other immediately neighbouring colonies as opposed to the entire plate.

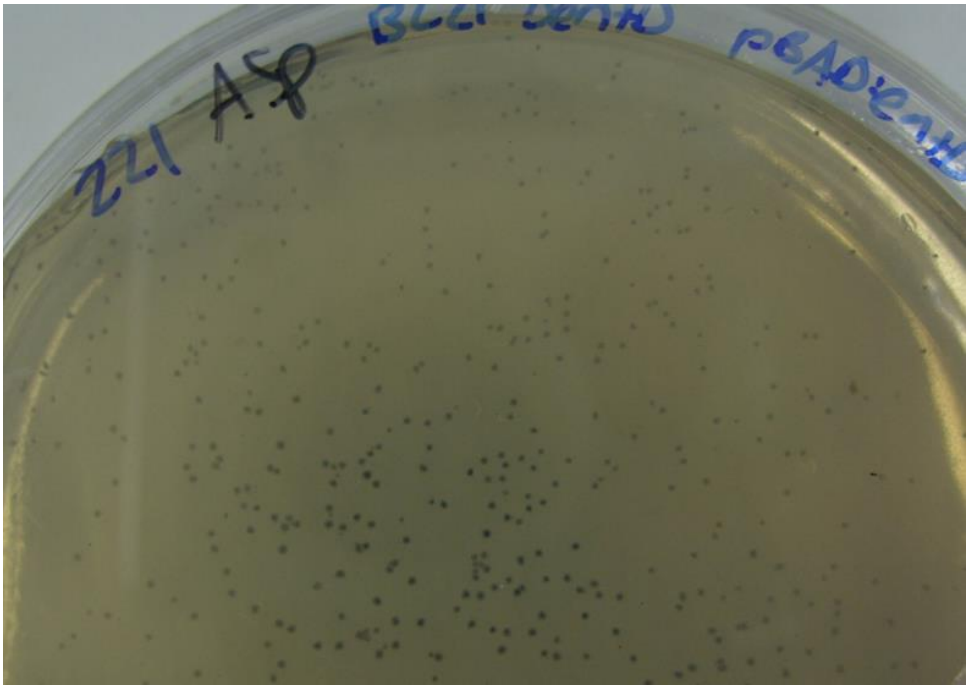


**Figure 6.3. Assessment of pigment production on low-iron agar plates.** EcoBlue1 pBAD:*entD* was plated on low-iron agar plates containing 100 mM L-glutamine. After 72 h incubation at 37 °C the plate was induced with 100 µl 2.5% (w/v) IPTG spread under the agar, and incubated at room temperature for a further 24 h before the photo was taken.

### 6.2.2 Creation of EntD variant libraries for screening

An *entD* variant library was created via error prone PCR, using the GeneMorph® Mutazyme® II polymerase and the WT *entD* as a template, with a high error rate applied. The gene library was cloned into the pBAD plasmid vector via compatible restriction sites (XhoI and HindIII), and used to transform electro-competent EcoBlue1. In order to estimate cloning efficiency, the WT *entD* gene was cloned into the same vector and also used to simultaneously transform electro-competent EcoBlue1. This transformation was plated onto LB plates containing 100 mM L-glutamine and 0.01% (w/v) arabinose, and induced with IPTG after initial growth as described in section 2.10.1. The percentage of blue colonies observed here was 87%, this was taken as an appropriate estimate for the cloning efficiency of the *entD* variant library. To check sequence diversity ten clones were sequenced. A high error rate had been chosen with the Mutazyme® II polymerase, as experience in our lab has shown that in practice the error rate is less than stated by the manufacturer. This seemed to be the case here as well, with two clones out of ten still being WT *entD* and five clones having just one mutation in the 630 bp *entD* gene. The remaining three clones had either two or three mutations.

To initially assess the percentage of variants encoding functional PPTases within the library, transformants were plated out onto both LB and low-iron agar pigment production plates. LB plates were incubated at 37 °C for 16 h prior to induction with IPTG, whereas low-iron plates were incubated at 37 °C for 72 h prior to induction. A survival percentage (corresponding to the percentage of functional PPTases in the library) was calculated by comparing the number of colonies growing on LB plates with those that grew on low-iron plates. The survival rate was estimated at 80% (taking into account the previously estimated insert ratio). This estimation was confounded by the fact that the plates appeared to have a lot of satellite colonies. This can be seen in Figure 6.4, which illustrates a number of larger colonies (which have blue centres indicative of leaky *bpsA* expression resulting in a small amount indigoidine production before induction) surrounded by a few smaller satellite colonies. Satellite colonies grow by scavenging enterobactin from the surrounding larger colonies as opposed to having a functional PPTase of their own. This problem has been seen before in iron-limited screens (Charlop-Powers et al., 2013).

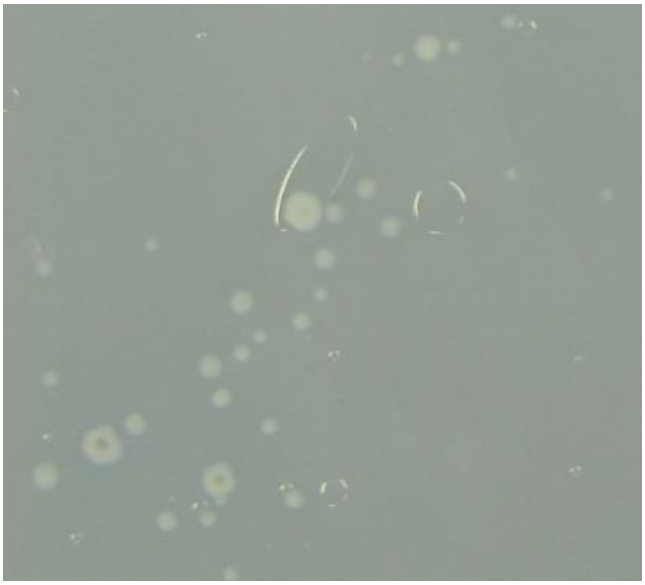


**Figure 6.4 Satellite colonies on low-iron agar plates.** EcoBlue1 expressing the *entD* library was plated on a low-iron plate and incubated for 72 h at 37 °C. Photo was taken prior to IPTG induction.

Another way to estimate the number of functional PPTases was to look at the number of blue colonies on LB plates for the EntD variant library *vs*. WT EntD. When estimated in this way, it seemed that closer to 50-60% of PPTases in the EntD variant library were still able to activate BpsA. The difference in these numbers is likely due to the survival rate being an over-estimate owing to the satellite colonies; alternatively, it may be that it is easier for the EntD variants to lose activity with the BpsA CP domain than with the enterobactin CP domains. The former explanation

being more likely to explain the discrepancy in estimates, it was assumed the number of functional EntD variants was closer to 50-60%.

#### 6.2.2.1 Library screening and further optimisation

Initial library cloning efforts were relatively inefficient for reasons that are unknown; in total approximately $5 \times 10^3$ clones were screened on low-iron pigment (indigoidine) production plates. This was however sufficient to identify several promising hits. Hits were selected as being those colonies that were white or had reduced pigment production in comparison to the surrounding colonies in the immediate vicinity. To cater for false positives resulting from the uneven pigmentation distribution or the presence of the satellite colonies on the low-iron plates, additional downstream screening steps were introduced. After the initial low-iron plate screening, colonies that looked promising (i.e., white or light blue, and ideally large and far away from other colonies) were streaked onto LB pigment production plates. This second step of streaking hits onto an LB plate allowed a better qualitative differentiation of pigment production levels. The whitest of these variants were cultured and plated out individually onto both LB and low-iron pigment production plates at a density of ~100 colonies/plate. This confirmed their ability to grow under iron limiting conditions (implying they were still able to activate their native CP substrates), as well as examining whether they were truly unable to activate WT BpsA. In total approximately 250 hits were chosen from iron restricted screening plates for streaking on LB. Twenty five of these hits were then plated on both LB and low-iron agar, and from this eight promising hits which had low to no indigoidine production as well as the ability to grow under low-iron selection were identified. These eight hits were subjected to further testing in the work described in section 6.2.3.

#### 6.2.2.2 Investigation of an alternative indigoidine-based screening strategy using EcoBlueE

Streaking out potential hits onto LB pigment production plates was a quick and easy step to overcome the problem of uneven pigment production leading to uncertainty about which EntD variants had lost the ability to activate BpsA. The extra step of growing colonies on low-iron media to overcome the problem of satellite colonies was a much more laborious strategy, with individual plates required to re-streak each potential hit. An alternative approach for confirming that the EntD variants still had activity with the EntF CP domain was explored. Instead of looking at the ability of EntD variants to complement the *ΔentD* strain under iron limitation, their ability to activate an EntF-substituted BpsA was tested (using the EcoBlueE strain). From the 250 initial hits recovered in the screening in section 6.2.2.1, 80 of the whitest hits on secondary screening LB

pigment production plates were selected (i.e., those hits that had been confirmed for a loss of activity with BpsA). Cultures for these hits were pooled and plasmids were isolated via miniprep. This new enriched library of EntD variants that had lost activity with the BpsA CP was then transformed into EcoBlueE and plated on LB pigment production plates. Cells were induced in the usual manner after overnight growth and this time colonies with the strongest pigment production were chosen as hits (i.e. those that retained activity with the EntF CP domain). This resulted in seven more hits being recovered and subjected to further verification.

### 6.2.3 Verification of potential hits on solid media

The fifteen hits found via the screening methods described in sections 6.2.2.1 and 6.2.2.2 were subjected to additional verification to confirm they had the desired activities with the appropriate substrates. First plasmids carrying each PPTase variant were separated from the BpsA-containing plasmid by transforming the mixed plasmids into DH5α, and selecting for the pBAD plasmid using ampicillin. The pBAD plasmid for each EntD variant hit was then used to transform both EcoBlue1 and EcoBlueE, and each was tested for pigment production on LB agar containing L-glutamine and arabinose. After initial colony growth, BpsA was induced with IPTG as described in section 2.10.1. Eight out of the 15 hits tested appeared to have gained the specificity that was selected for, i.e. appeared to have lost the ability to activate the WT BpsA (loss of indigoidine production when expressed in EcoBlue1), while retaining activity with the EntF CP domain (indigoidine was produced when expressed in EcoBlueE). Of these eight hits, four each were recovered by the two different screening methods described in sections 6.2.2.1 and 6.2.2.2. As an example, the results for Clone 1 are displayed in Figure 6.5, compared to WT EntD. The time taken for indigoidine to be produced was similar for most hits, except for Clone 4 which was much slower. This indicated that this clone may have also lost some activity with the EntF CP domain as a substrate.

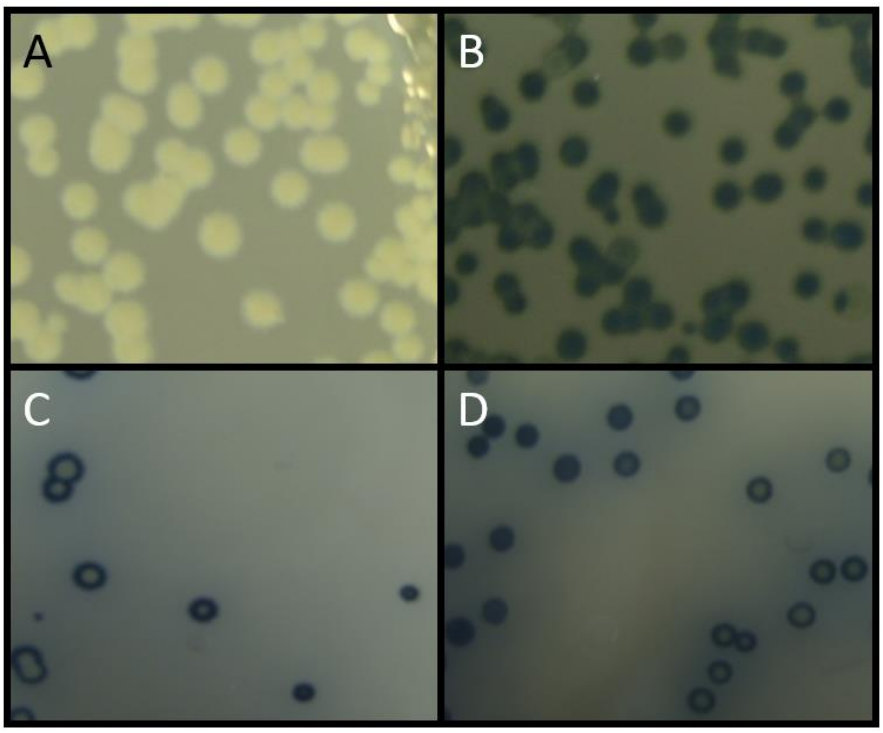



**Figure 6.5. Verification of CP specificity for Clone 1.** Clone 1 was expressed in both A) EcoBlue1 and B) EcoBlueE. WT EntD is shown for comparison in C) EcoBlue1 and D) EcoBlueE. Each strain was plated on LB agar with 0.01 % (w/v) arabinose and 100 mM L-glutamine and induced with 100 µl 2.5 % (w/v) IPTG before being incubated at room temperature. Photos were taken 48 h post induction.

The other seven of the 15 hits tested proved to be false positives. Most of these still had at least limited activity with both substrates, resulting in indigoidine production in both EcoBlue1 and EcoBlueE. This was surprising given each of the potential hits had already been double checked for loss of ability to activate BpsA. It may be relevant that here that each of the EntD variants was re-transformed into EcoBlue1 - possibly individual *bpsA* genes in the previous screening strains had acquired deleterious mutations, with the mild toxicity of indigoidine providing some selective pressure for these mutations. There was also one false positive for which there was no indigoidine production in either EcoBlue1 or EcoBlueE. This may have been a satellite colony, a completely non-functional PPTase that appeared white so was recovered as a hit yet was unable to produce its own enterobactin and was just scavenging some from a nearby colony. Another possibility is that the EntD variant could still activate its native substrate EntF CP in *E. coli* (so it could produce enterobactin and grow on iron restricted plates), but it could not activate the EntF CP-substituted BpsA.

### 6.2.4 Sequence analysis of verified hits

Hits that were verified in section 6.2.3 were sequenced, and a list of all clones and their mutations is given in Table 6.2. The mutation rate among the hits was relatively high, with 2-3 amino acid mutations per gene (the size of *entD* is 630 bp). Though only a small library was screened, the hits recovered were diverse, with seven out of eight hits sequenced being unique (Clone 7 was recovered twice).

**Table 6.2. Mutations present in improved *entD* variants**

| Clone | Mutations present |
|---|---|
| 1 | A45T, Y88N |
| 2 | L33S, R47H, Y88N |
| 3 | P25Q, F151L |
| 4 | P12T, Y88N, E112G |
| 5 | W35L, A45T, Y88N |
| 6 | L42Q, Y88N |
| 7 | G76D, P81S, Y88N |

All but one of the hits contained a shared mutation, Y88N. That this mutation was recovered together with so many alternative mutations is strong evidence that it had arisen several times during the epPCR reaction and was actively selected for. However, it was not found on its own so it was unclear what the effects of any of the other mutations that were found alongside it might be. Clone 3 was the only variant that did not contain this mutation, thought it did contain the mutation F151L. This Phe residue is highly conserved within PPTases, with all known PPTases having either a Phe or a Trp residue at this location.

### 6.2.5 Quantification of EntD variant activity with different CP substrates

#### 6.2.5.1 Optimisation of arabinose concentration for liquid culture assay

In order to obtain quantitative data on the activity of the EntD variants with different CP substrates, a liquid culture assay was performed to compare activity in EcoBlue1 versus EcoBlueE. It was first necessary to optimise the arabinose concentration, as previously this assay had been run when both the genes were under IPTG control. The previously optimised conditions of 0.5 mM IPTG, 100 mM L-glutamine at 18 °C were kept constant and the concentration of arabinose was varied from 2% to 0.00002% (w/v) in six 10-fold serial dilution steps. The pBAD:*entD* plasmid was

expressed in both EcoBlue1 and EcoBlueE. The levels of pigment production observed at both 24 and 48 h are displayed in Figure 6.6. At the higher concentrations of arabinose (2% and 0.2%), both strains produced similar levels of indigoidine at 24 h, with no noticeable difference between EcoBue1 and EcoBlueE (Figure 6.6 A). After 48 h a greater differentiation in pigment production between the two strains was seen (Figure 6.6 B). At the lower concentrations of arabinose, EcoBlueE pBAD:*entD* had much higher levels of indigoidine production than EcoBlue1 pBAD:*entD*, with the two lowest arabinose concentrations showing no visible pigment production in EcoBlue1 (Fig 6.6 B). These results indicate that when there is a lower amount of EntD present (i.e., at lower arabinose concentrations), then this small amount of enzyme is sufficient to activate the EntF CP-substituted BpsA, but a larger concentration of EntD enzyme is required to activate WT BpsA. So that any potential changes in the substrate specificity of the evolved EntD variants were not exaggerated due to the lower concentration of EntD enzyme, the second highest arabinose concentration (0.2%) was chosen for use in the final assay.

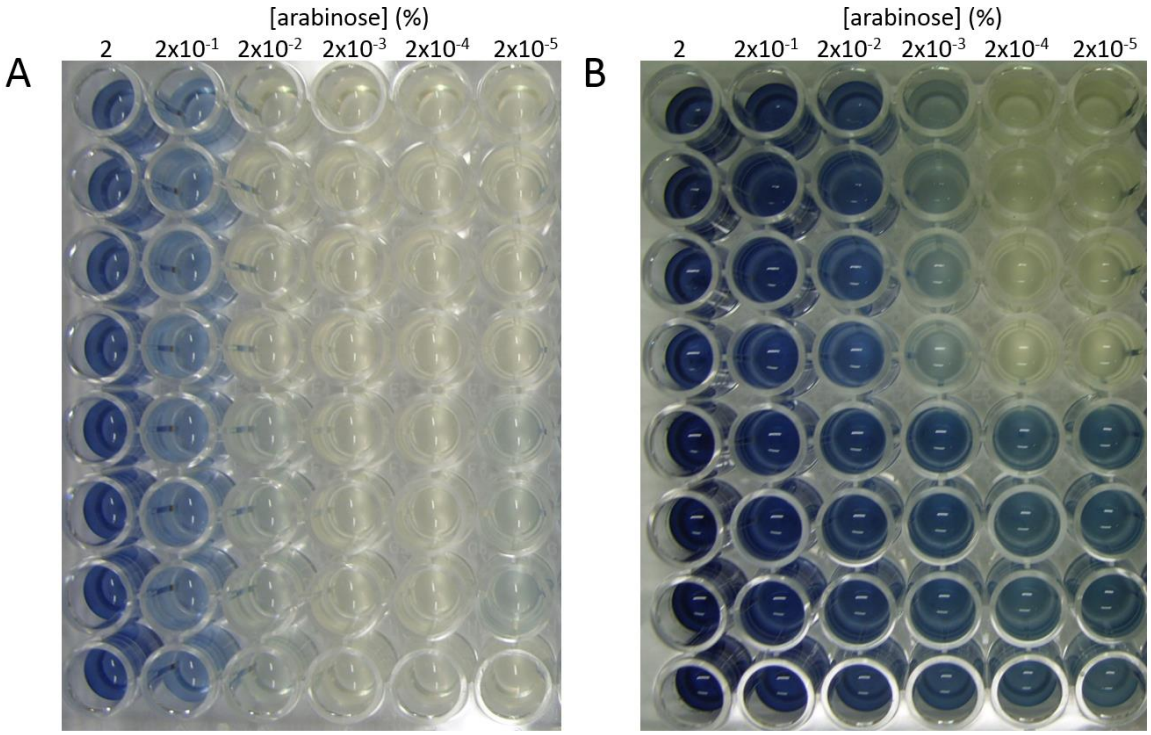



**Figure 6.6. Evaluation of optimal arabinose concentration for indigoidine synthesis.** Columns contain a 10-fold serial dilution series from 2% to 0.00002% of arabinose, The top four rows contain EcoBlue1 pBAD:*entD* and the bottom four rows contain EcoBlueE pBAD:*entD*. Each well contains a final volume of 150 µl with 0.5 mM IPTG and 100 mM L-glutamine, as well as 100 µg/ml of ampicillin and spectinomycin. The plate was incubated at 18 °C with shaking at 200 rpm and photographed at A) 24 h and B) 48 h.

#### 6.2.5.2 Quantitative assessment of hits

Each of the seven sequence verified hits was quantitatively assessed by measurement of indigoidine production in liquid culture, which had previously proven to be a more sensitive and quantitative measure of PPTase activity (Chapter 3). Each EntD variant was evaluated both in EcoBlue1 and EcoBlueE, and Figure 6.7 displays the pigment production for each variant at 24 and 48 h. After 24 h incubation, most of the variants tested showed very little activity in EcoBlue1, and were all much less active than WT EntD in this strain. Clones 4 and 7 seemed to have lost activity with both substrates but all others had retained their activity with the EntF CP domain, still able to activate the EntF-substituted BpsA variant in EcoBlueE. More activity was seen with all clones after 48 h, and the quantification of this indigoidine production at both time points is displayed in Figure 6.8. This shows that at 24 h, all EntD variants generated substantially less indigoidine production than the WT EntD, with Clones 4, 6, and 7 resulting in no visible indigoidine production at this stage. After 48 h, increased indigoidine production was seen across all the strains (except for Clone 4 in EcoBlue1, which still appeared inactive with WT BpsA). For every variant tested, there was a greater level of indigoidine production in EcoBlueE than EcoBlue1. Most variants did however have lower activity in EcoBlueE than the WT EntD enzyme, indicating that in reducing activity with the WT BpsA, overall activity with the EntF CP seemed to have decreased as well. The exception to this is Clone 1 which, compared to WT EntD, both lost activity in EcoBlue1 and gained activity in EcoBlueE.

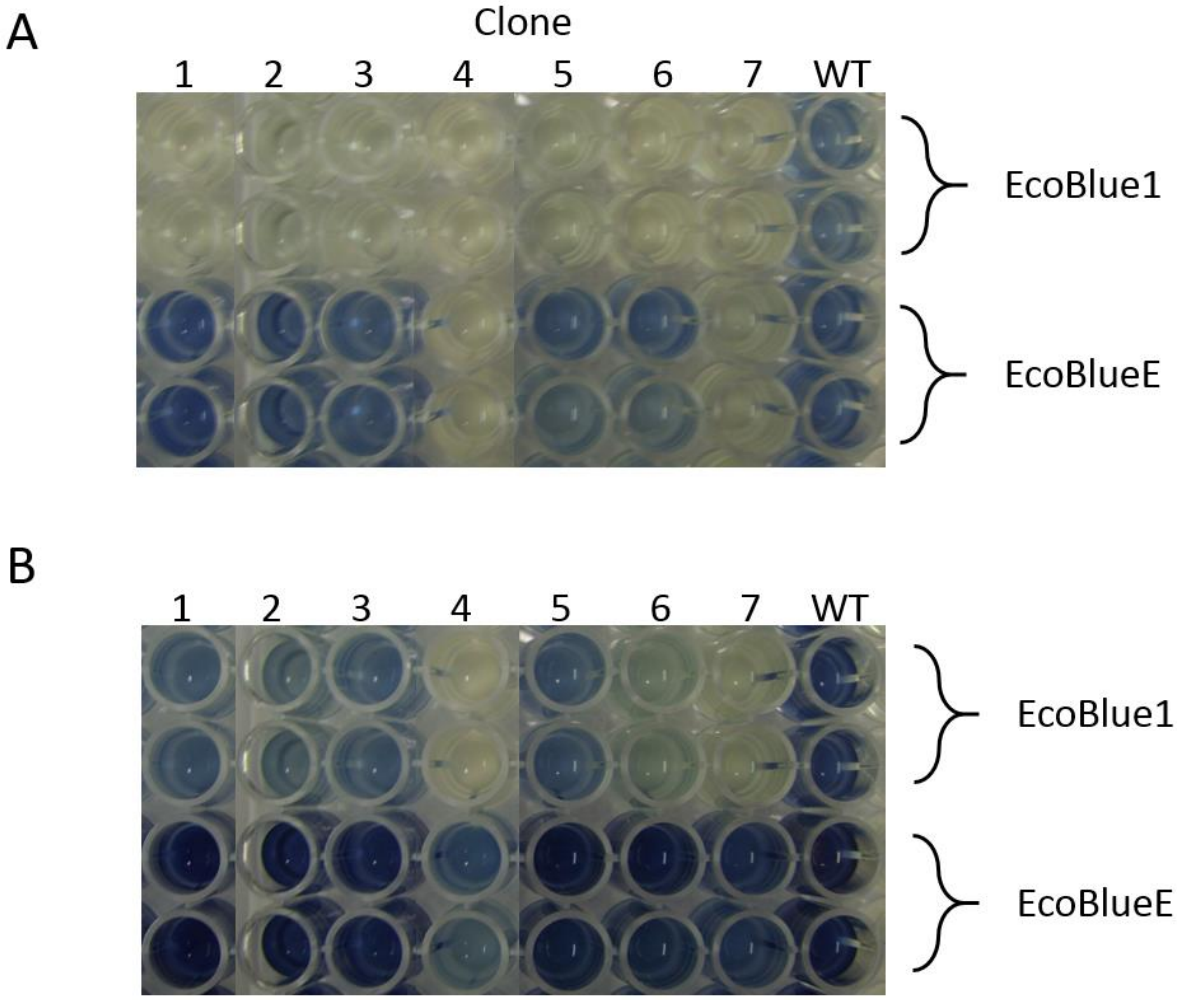



**Figure 6.7 Indigoidine production of the seven EntD variants in liquid media.** Each of the seven clones was measured for indigoidine production in EcoBlue1 and EcoBlueE. Wells each contained 130 µl of pigment production media (LB with 0.2% (w/v) arabinose, 100 mM L-glutamine, 0.5 mM IPTG, 100 µg/ml ampicillin and spectinomycin) inoculated with 20 µl of overnight culture. Plates were incubated at 18 °C with shaking at 200 rpm and photographs were taken after A) 24 h and B) 48 h.

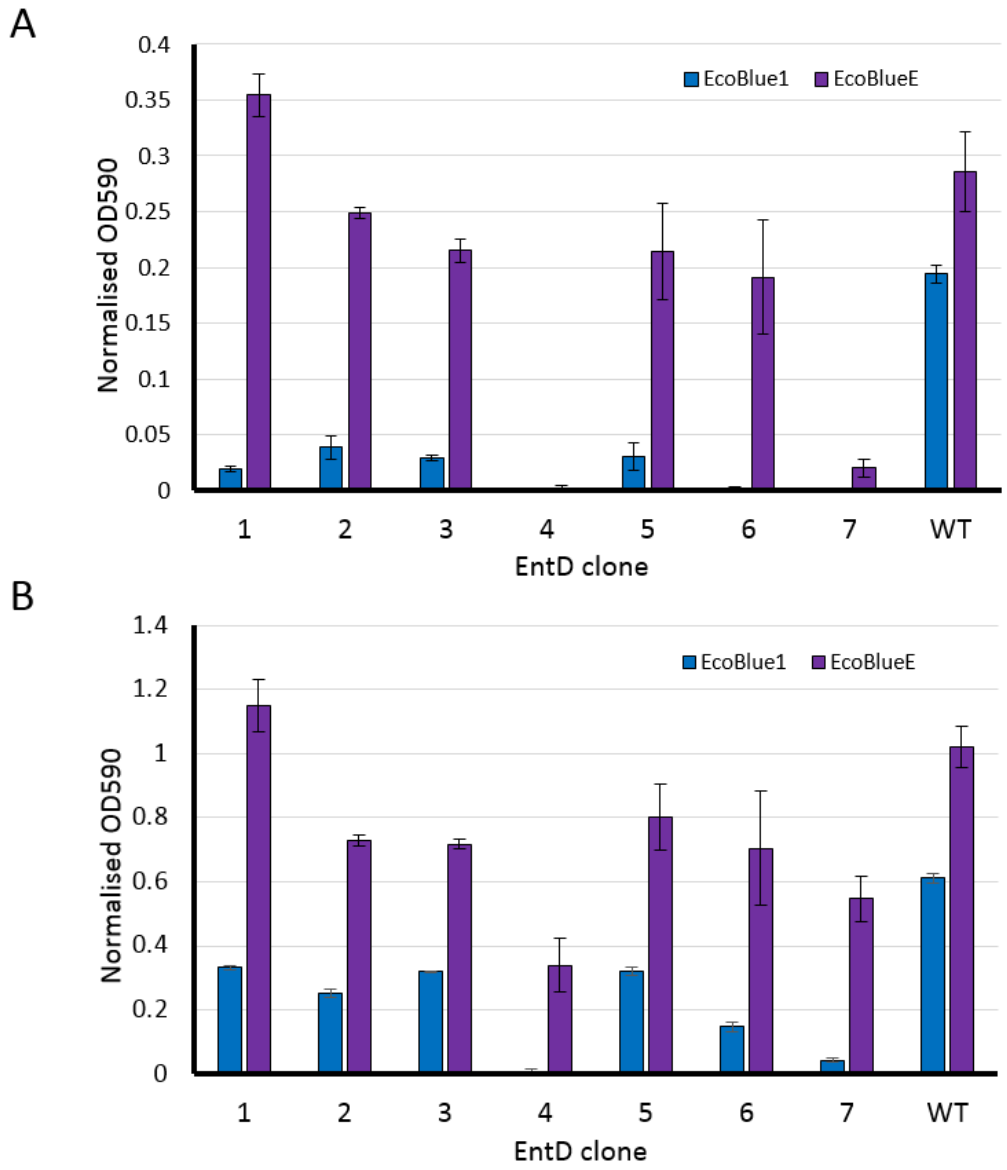



**Figure 6.8 Quantitative assessment of indigoidine production resulting from the seven EntD variants being co-expressed with two BpsA variants that contain different CP domains.** Quantification of indigoidine production resulting from the seven EntD variants (compared with WT EntD) being expressed in EcoBlue1 or EcoBlueE. Each was set up in quadruplicate in a 96-well microplate in pigment production media and incubated at 18 °C. After A) 24 h and B) 48 h absorbance at 590 nm and 800 nm was measured using a microplate reader. Absorbance at 590 nm was normalised for cell density as measured at 800 nm to give a relative measure of indigoidine, as described in section 3.2.1.2. Error bars are +/- one standard deviation.

The results of this liquid assay were somewhat inconsistent with the agar plate assay in section 6.2.3. Previously (in Chapter 3) it was observed that the liquid assay is more sensitive and hits that are slow to turn blue on agar (with no pigment produced within the first 48 h), will not produce any indigoidine in the liquid assay. However here Clones 2, 3 and 5 all had no activity at all on the agar plate assay in section 6.2.3 when expressed in EcoBlue1, yet were able to produce indigoidine in the liquid culture assay here. This could be due to the fact that here the expression of the EntD

variants were under arabinose control, and it may be that the concentration of arabinose in the agar was not high enough for substantial levels of induction.

### 6.2.6 Kinetic analysis of selected EntD variants

Several EntD variants spanning a range of activity levels as measured in section 6.2.5.2 were selected to be purified and characterised kinetically, for activity with WT BpsA and the EntF-substituted BpsA. The EntD variants were already in the pBAD plasmid, which adds a His6 tag to enable nickel affinity purification, but sufficient levels of soluble protein could not be obtained following arabinose induction. As expression of EntD via IPTG induction had already been optimised for expression in pET, it was decided to clone four of the *entD* variants (1, 3, 4 and 6) into pET. The encoded EntD enzyme variants were then purified following the protocol for the purification of PPTases described in section 2.8.5.2, with expression carried out in LB at 18 °C as for WT EntD. Initial tests for activity were performed to establish whether each variant possessed enough activity for full enzyme kinetic assays to be performed. Overall, all the variants tested had very low activity when compared to WT EntD, and because of this kinetics for most EntD variants were unable to be determined. An exception was Clone 1, for which kinetic analysis was performed (using purified EntF CP substituted BpsA as substrate) in two independent biological replicates (two independently expressed and purified protein preparations, with an assay performed in triplicate for each protein preparation). The Michaelis-Menten curve for varying CoA concentration for each of the replicates is shown in Figure 6.9. The two replicates were consistent, and the kinetic parameters derived from each were averaged and given in Table 6.3, along with the kinetic parameters for WT EntD. Note that Clone 1 did not exhibit sufficient activity with WT BpsA to measure any activity in this assay.

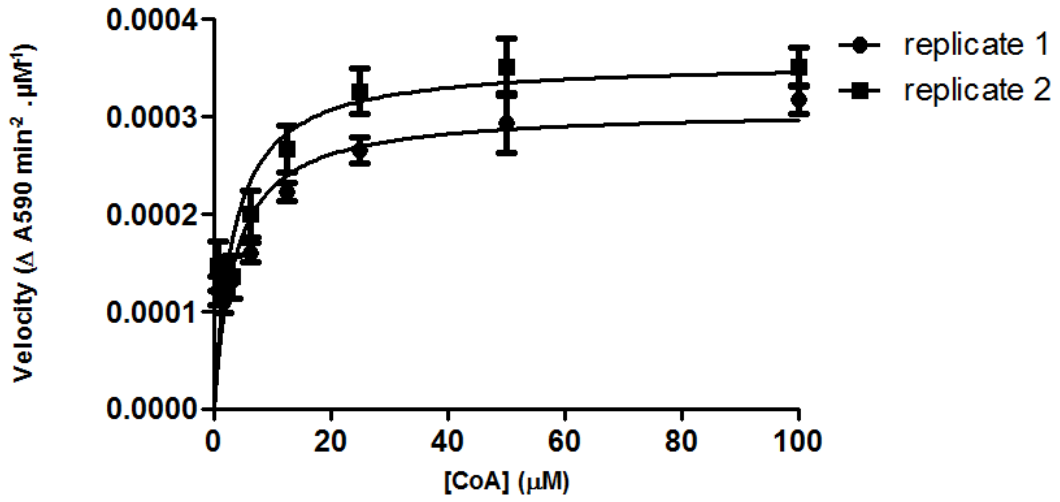



**Figure 6.9 Michaelis-Menten curves for two biological replicates of the EntD variant "Clone 1" with the EntF CP-substituted BpsA as substrate.** Plots indicate the maximum velocity of indigoidine production at different concentrations of CoA. Curves were fitted using Graphpad Prism® and used to derive kinetic parameters for each replicate of Clone 1. Data points were generated from three repeats and error bars are +/- SEM.

**Table 6.3 Kinetic parameters of EntD Clone 1 and EntD WT**

| EntD variant | BpsA substrate | $V_{max}$ ($\Delta$A590.min$^{-2}$.mM$^{-1}$) | $K_m$ (μM) | $V_{max}/K_m$ ($\Delta$A590.min$^{-2}$.μM$^{-1}$.mM$^{-1}$) |
|---|---|---|---|---|
| Clone 1 | BpsA (EntF CP) | 0.33 ± 0.014 | 3.39 ± 0.59 | 0.098 ± 0.018 |
| Clone 1 | BpsA | n.d | n.d | n.d |
| WT EntD | BpsA (EntF CP) | 18.8 ± 0.91 | 0.96 ± 0.23 | 19.6 ± 4.8 |
| WT EntD | BpsA | 5.3 ± 0.26 | 1.24 ± 0.28 | 4.28 ± 0.99 |

n.d. = not detected

While the Clone 1 variant had lost a lot of activity (around a 200-fold drop in catalytic efficiency (i.e., $V_{max}/K_m$) when compared to WT EntD) the specificity of the enzyme was greatly improved, with this variant having no detectable activity at all with the WT BpsA enzyme. The three other variants that were purified (Clones 3, 4, and 6) had also lost a lot of activity, to such an extent that it was not possible to measure their kinetic activity. However, it was possible to see that they had also gained the desired specificity and were able to activate the EntF-substituted BpsA preferentially over WT BpsA (data not shown).

## 6.3 Discussion

The aim of this chapter was to conduct proof-of-principle directed evolution studies to modify the PPTase EntD in order to narrow its CP domain substrate specificity. EntD was evolved to be more selective for a native CP substrate (the CP of EntF), while losing activity with a second CP substrate, the native BpsA CP domain. Despite the small size of the EntD variant library screened, I was

successful in recovering several variant EntD enzymes which displayed this altered substrate specificity. While there was a loss in overall catalytic efficiency, the top variant recovered had retained enough activity with EntF in order to derive kinetics with the EntF CP-substituted BpsA, while displaying no measurable activity with WT BpsA in the same assays. The screen developed here will be able to be adapted for evolution of other PPTases with other CP substrates.

In total seven clones were identified that had low activity with the WT BpsA CP domain, and higher relative levels of activity with the EntF CP domain. The top performing EntD variant recovered was Clone 1. An *in vivo* liquid assay was performed to quantify the activity of the EntD variants with the CP domains of BpsA and EntF, as measured by indigoidine production in both EcoBlue1 and EcoBlueE. Clone 1 appeared to be substantially more selective than WT EntD, with a greater difference in the activity between EcoBlue1 and EcoBlueE (Figure 6.8). Clone 1 appeared *in vivo* to have even more activity with the EntF CP-substituted BpsA than WT EntD did (all other variants were less active in EcoBlueE as compared to WT EntD) when tested in the liquid culture assay. However when purified enzyme for Clone 1 was tested kinetically with BpsA (section 6.2.6), it was found to be much less active than WT EntD with the EntF CP substituted-BpsA, with an approximate 200-fold drop in catalytic efficiency under the *in vitro* conditions tested. Even though there was this large loss of activity with EntF as the CP substrate, there was no detectable level of activity with WT BpsA, showing that the screen had been successful in recovering an EntD variant with narrowed CP substrate specificity. A possible reason for the difference between the *in vivo* and *in vitro* activity levels seen is that the purified enzyme may be a lot less stable then the purified WT EntD, due to the mutations it contained. Future work could examine this, for instance running a thermostability assay to compare the stability of the EntD variants against the WT enzyme (Niesen et al., 2007).

The remaining EntD variants were all less active than Clone 1, as measured via indigoidine production in the *in vivo* liquid culture assay. This also appeared to be the case at the purified protein level as there was not enough activity seen with any of the other purified enzyme variants to run kinetic assays. Despite this low overall activity, all of these clones did appear to display improved selectivity, having a greater level of activity with the EntF CP domain over the BpsA CP domain, in both solid and liquid media assays. However, for the reasons noted above it was difficult to accurately judge the extent of improvement, as the timing of measurement influenced relative activity levels, and it is unclear whether fold-improvement in activities is the most appropriate measure for comparing different enzymes. While it would be preferable to lose activity with the BpsA CP substrate without any concomitant loss in catalytic activity with the EntF

CP substrate, if the activity with WT BpsA was completely eliminated, this may yet prove useful in some applications. Such an enzyme would exhibit no background activation with WT BpsA, which might be very useful as long as sufficient activity was retained with the EntF CP domain. Clones 4 and 7 may be good examples of this, as it looks as if activity with WT BpsA has been almost completely removed.

Due to the fact all seven EntD variants contained multiple (two or three) mutations, without further experimental data it is hard to draw firm conclusions about what the effect of individual mutations may have been. Clearly Y88N looks to be an important mutation, being present in six out of seven of the variants, including the top hit, Clone1. It would be of interest to examine the activity of an EntD clone with the single point mutation Y88N. As all of the other variants containing Y88N have less activity than Clone 1, it also seems likely that the other mutations present in these variants (Clones 2, 4, 5, 6 and 7) may be detrimental, or at best neutral. The use of site directed mutagenesis to generate a Y88N single mutant and then – one at a time – add in the additional mutations that were recovered, would enable a detailed comparison of the effects of each mutation on overall catalytic activity and selectivity. Unfortunately however, time constraints precluded these experiments from being performed. Instead, preliminary hypotheses were generated by analysing the location and nature of the different substitutions that appeared, and comparing these to the *in vivo* assay data. For example, Clone 5 contains the same two mutations as Clone 1 (Y88N and A45T) as well as the mutation W35L; Clone 5 has less activity (with both CP substrates) than Clone 1, so this implies W35L must be detrimental to overall enzyme activity. Clones 1 and 6 are also very similar, both containing the Y88N point mutation as well as one additional other mutation (A45T for Clone 1and L42Q for Clone 6). These additional mutations are in a similar location of the protein. While there is no solved structure available for EntD, alignments with type II PPTases that have had their structure solved (Bunkoczi et al., 2007; Jung et al., 2014; Reuter et al., 1999; Vickery et al., 2014) suggest that these mutations are likely to occur in the second alpha helix of the N-terminal of the enzyme, which may be involved in CoA binding (Reuter et al., 1999). A45T is a relatively conservative amino acid change whereas L42Q is not, and this more dramatic change may be responsible for the greater loss of total activity observed for Clone 6.

The worst performing EntD variant (with both CP substrates) was Clone 4, which includes the mutation E112G. This position is equivalent to E109 in Sfp, which is important in $Mg^{2+}$ binding (Reuter et al., 1999) - in a mutagenesis study the mutation E109D was found to completely ablate activity (Mofid et al., 2004). Recently this position was identified as the second residue in a triad

of residues (Asp-Glu-Glu) that are important in coordinating the $Mg^{2+}$ ion (Wang et al., 2014). The second residue was found to be non-essential in this role, and an alternate residue to Glu is found in a number of different species. Gly was not one of the residues that was ever found at this position.

Overall most of the mutations in recovered clones were in the N terminal domain of the protein. This is in contrast to other studies that have stated that the specificity for carrier protein recognition comes from the C terminal domain (Jiang et al., 2013; Tufar et al., 2014). However, it was suggested by Jiang et al. (2013) that while the C terminus was important for specificity, the N terminus was important in overall enzyme activity. This is consistent with the fact all of these variants (with N terminal mutations) did have a lot lower activity overall than the WT enzyme.

The fact that this screen was a survival based screen on agar plates (i.e., only functional PPTases with the desired substrate specificity for the EntF CP domain could survive under iron limiting selection) meant that a large number of clones should have been able to be screened relatively easily. One of the limitations encountered in this work was that the library of EntD variants produced via error prone PCR was very small, and so only around 5000 clones were screened. Despite trying to improve this and determining that there was high vector quality with a good insert ratio and highly competent cells the library size was not able to be increased in the time available. Nonetheless, that there were still a good number of hits found indicates that the substrate specificities of PPTases may be relatively easily manipulated.

Another limitation encountered was the large number of false positive hits that were detected. Out of the fifteen hits that were selected for a final verification step in section 6.2.3, eight were confirmed as having the sought-after selectivity while seven were false positive hits. As there was a problem with satellite colonies being found on screening plates, it was thought that this might be one reason for false positive hits. A completely non-functional PPTase could be recovered because it had scavenged enterobactin, appearing active with EntF, while also appearing to have lost activity with WT BpsA. However, this was only the case for one of the seven false positive hits, which was unable to activate either BpsA variant upon secondary verification (i.e., giving no indigoidine production in EcoBlue1 or EcoBlueE). This indicates that the additional steps taken to double-check activity with the EntF CP domain (either low-iron plate testing as in section 6.2.2.1 or transforming a mixed plasmid population en masse into EcoBlueE as in section 6.2.2.2) were largely successful in eliminating any non-functional PPTases. The remaining six false positive hits had not lost activity with WT BpsA upon verification, and displayed activity levels in EcoBlue1 and EcoBlueE similar to that of WT EntD (indeed, although not mentioned previously, sequencing

revealed that three of these clones carried no mutations at all). Selection of these unimproved variants may have stemmed from the fact that there were inconsistencies in indigoidine production on the screening plates. However, as all potential hits were subsequently streaked onto pigment producing LB to confirm loss of indigoidine production, it seems more likely that these variants may have by chance accrued mutations in their *bpsA* reporter genes or plasmids that led to a diminished indigoidine output. Owing to the toxicity of indigoidine, instability in the BpsA plasmid has been seen before (Owen, 2010). During the verification step pBAD plasmids containing the EntD variants were retransformed into EcoBlue1 for verification where it was seen that they were indeed able to activate indigoidine synthesis.

While this screen was specific for the EntD PPTase and relied on the fact that we had a *ΔentD E. coli* strain available, it could also be applied to the engineering of other PPTases. For example, if another PPTase knock out strain of a different siderophore-producing bacterial species was available, then the same principles could be applied. Based on the pilot results described in this chapter, further work in our laboratory is looking at performing an equivalent screen using the *pcpS* PPTase in *P. aeruginosa.* A conditional knock out of the essential gene *pcpS* has previously been created along with a chromosomally inserted *E. coli acpS* gene (Barekzi et al., 2004). The *acpS* gene allows activation of the pathways of primary metabolism necessary for growth, while not activating the siderophore pathways of secondary metabolism. Ongoing work in our lab is being performed by Masters student Jack Sissons (under my co-supervision) with the aim of evolving both PcpS and EntD so that they have complementary (i.e., non-overlapping) CP specificities. For EntD, the same screen will be performed as in this chapter but using the PvdD CP-substituted BpsA variant in place of WT BpsA, in order to obtain an EntD variant that retains activity with the native EntF CP domain and loses activity with the PvdD CP domain. The complementary study with PcpS will then be performed, in the *ΔpcpS P. aeruginosa* strain, to select for PcpS variants that lose activity with the EntF CP-substituted BpsA variant.

As part of the screening method developed in this chapter, a second tier assay for recovering hits was investigated. Instead of relying on growth on low-iron media to select for variants which had activity with the EntF CP domain, a subset of variants already found to have lost activity with WT BpsA were transformed into the EntF CP-substituted BpsA variant strain, EcoBlueE, and selected based on indigoidine production in this strain. This was to overcome the problem of having to test potential hits to ensure they were not satellite colonies, which needed to be done individually and was time consuming. This approach was successful in uncovering four out of eight of the original hits during the course of this work, with two of these hits having identical sequences. There is the

potential to solely use this method in order to engineer PPTase selectivity. This would not be reliant on having a particular PPTase knock out strain, so would be a more flexible system, though would rely on having the appropriate CP-substituted BpsA variants available for screening. It would also have the disadvantage of not being a survival based system, with no easy way to select against non-functional PPTase variants.

It remains to be seen whether the specificity of the EntD variants recovered here will translate to a specificity for the CP as stand-alone proteins or in a different context, i.e. with the CP or a derivative thereof attached to another protein in a labelling application. The fact that the EntF CP has essentially been tested both in the context of its native EntF NRPS as well as the BpsA NRPS is a good indication that the activity seen here will not just be specific to one situation. There is also the question of whether the CP domains here can be shortened to smaller peptide tags and while retaining the same properties, and further work by Jack Sissons in our lab is investigating this.

# Chapter 7: Summary, Conclusions and future directions

## 7.1 Research motivation

The focus of this research was a class of natural product biosynthetic enzymes, the NRPSs, and the enzymes responsible for activating them, the PPTases. NRPSs are of interest due to the wide range of bioactive products they synthesise, with two key areas of research being the manipulation of existing NRPS enzymes, or the discovery of new enzymes, in order to create or find new bioactive natural products. PPTases are also important enzymes that have potential for use in a variety of applications, for which an understanding of the reaction between the CP domains of NRPSs and the PPTases is essential. This reaction was a key theme of this thesis, and the work presented here was motivated by the potential to use a unique NRPS, BpsA, to study this reaction further.

The initial aim of this research was to use the model NRPS BpsA, expanding upon a series of versatile assays that were first developed as a part of the doctoral work of Dr Jeremy Owen. As part of this preceding work, a number of different CP domains were substituted into BpsA in order to investigate the interactions between the CP domain and the TE domain. As a result of this, several functional CP-substituted BpsA variants were created. The first aim of this thesis was to further progress the CP-substitution work, focusing on specific interactions between these domains, in light of new information which contradicted some of Dr Owen's previous conclusions. The second aim of this thesis was to utilise the engineered CP-substituted BpsA variants and apply them to a metagenomic screening method, to expand upon a small pilot metagenomics study that Dr Owen had performed using WT BpsA. During the course of this screening a number of novel PPTase genes were discovered, and these were investigated further, this time using BpsA to characterise the PPTases in terms of their activity and specificity. The motivation behind this large scale characterisation was to better understand the properties of these enzymes in order to discover promising PPTases for use in a variety of biotechnological applications, in particular PPTases that might exhibit a strong preference for specific CP domain sequences. Upon noticing a lack of any clear specificity of the PPTases tested for their CP substrates, the final aim of this work was to utilise the enzyme BpsA in a completely novel screen which could be applied in directed evolution experiments to narrow the specificity of a target PPTase.

## 7.2 Key findings

### 7.2.1 Carrier protein substitutions in BpsA

A greater understanding of the interactions between the different domains and/or modules of NRPS enzymes may increase our ability to manipulate these enzymes to create new and novel products. Currently such enzyme manipulation experiments generally result in enzymes with decreased function (Calcott and Ackerley, 2014). Previously, the NRPS BpsA has been used in CP substitution experiments, with CP domains being engineered within the BpsA NRPS in order to increase enzyme function, identified by an increase in indigoidine production levels. The first aim of this thesis was to examine features of CP domain substitutions that were likely to result in a functional CP-substituted BpsA enzyme. During the course of Dr Owen's previous substitution experiments he made two observations, both of which appeared to be important factors in whether or not substitutions would be functional. The characteristics of the CP domains that resulted in functional CP-substitutions were: 1) these CPs were located upstream from TE domains in their native context (consistent with the CP domain of BpsA being situated immediately upstream from the BpsA TE domain); and 2) these CPs exclusively had hydrophobic residues at positions +4 and +24 (relative to the invariant Ser). Subsequently, four CP domains were found which appeared to contradict these observations (these four CP domains being located upstream from TE domains in their native enzymes, but having the polar residue Thr at the +4 position). The goal of the work in presented in Chapter 3 was to use these four CP domains to test the previous observations, and to answer the question of whether the native location of the CP domain or the identity of the +4 residue was the more important factor in predicting successful CP domain swaps in BpsA. These four domains were all substituted into BpsA, and all were functional to some extent. This indicated that the original location of the CP domain was the more important factor in predicting the success of swaps. However, although functional, all substitutions showed substantial decreases in activity compared to the WT BpsA enzyme.

The importance of the +4 residue, as well as the +24 residue, was also investigated using the CP domain PvdDCP1. Previous substitution into BpsA had shown that while the initial substitution was non-functional (PvdDCP1 was upstream from a C domain in its endogenous enzyme), functionality could be restored via directed evolution. Here, the PvdDCP1 domain was substituted into BpsA and site directed mutagenesis was used to create variants which contained all possible amino acids at the +4 and +24 positions, both individually and in combination. Results confirmed the importance of a hydrophobic residue at the +4 position, but this position appeared to be of

minor importance compared to the +24 position, with residue optimisation yielding smaller increases in indigoidine production. The earlier work by Dr Owen had led to the hypothesis that the +24 position played a key role in the hydrophobic interface between the CP and the TE domain, with hydrophobic residues at this site leading to improved enzyme function. The work presented in this thesis demonstrated that a wider range of residues appeared to be tolerated at this position (although many of the most active mutants did contain a hydrophobic residue). The results suggested that not just the hydrophobicity but the size of the residue was also important, with smaller less bulky residues resulting in more active CP-substituted enzymes.

The key findings from Chapter 3 were that both the location of a CP domain in its native NRPS enzyme (i.e., the identity of the domain that it is immediately upstream of) and the identity of residues at key positions are important in predicting functionality of substitutions. If either condition is met, there is a high likelihood of the resulting substitution being functional. A combination of these factors can also work together synergistically, with the +4 and +24 positions having a much greater effect when both were optimised, compared to either alone. The work in this thesis looked at interactions specifically in BpsA; this will potentially allow further BpsA variants to be created that could potentially be used in other applications (such as in the metagenomic screen performed in Chapter 4 of this thesis). The residues confirmed as being important here have also been identified as being important in interdomain communications in other NRPS enzymes, so these rules are likely to be more generally applicable as opposed to pertinent to just BpsA.

### 7.2.2 Discovery of PPTases and natural product genes

The mining of bacterial metagenomes from uncultivable microorganisms is an important tool in the search for new natural products or enzymes which may have biotechnological applications. The work presented in Chapter 4 describes a screen using BpsA to identify PPTases from a metagenomic sample. In this screen, eDNA was co-expressed in the presence of *bpsA*; if the eDNA insert contained a PPTase that was both expressed, and capable of recognising and activating BpsA, then the resulting indigoidine production led to the recovery of PPTase-containing clones. Two small insert eDNA libraries (the Sw library and the NZ library) were screened, using both WT BpsA and two CP-substituted variants. During the course of this research, 21 hits were found, each containing a novel PPTase. Eight of these hits (approximately 40% of the total) also contained evidence that they were part of a biosynthetic gene cluster. The fact that these were such small insert libraries means that some of the remaining PPTases may also have been present within biosynthetic gene clusters but the recovered fragment was not identifiable as such due to the lack

of surrounding DNA. For instance in two of the remaining inserts, the PPTase gene was the only ORF present, and in several other small inserts the PPTase gene was present at the edge of the insert, and may have in fact been adjacent to biosynthetic genes immediately to the right or left, that were not represented in the recovered insert. This suggests that, barring a statistical anomaly, 40% might be expected to be the minimum number of PPTase genes that occur alongside natural product biosynthetic genes in nature, and therefore that using this screen is an excellent way to enrich for such genes. In fact a later screen, which used a derivative method to recover PPTase genes from a metagenomic library, found that up to 70% of the clones they recovered also had NRPS or PKS genes present (Charlop-Powers et al., 2013). Another key result that came out of Chapter 4 was that use of different CP-substituted BpsA variants successfully increased the number of hits that were recovered, presumably due to the fact that different PPTases were able to activate different CP domains. For example, three out of the seven hits from the Sw soil library and five out of fourteen hits from the NZ soil library were not recovered when the libraries were screened with WT BpsA.

As well as providing a means for discovering new natural product biosynthetic genes, this is also a novel method for the discovery of PPTase genes. The use of a direct screen for the discovery of new PPTase genes has never previously been reported. As a family, PPTases do not produce a product that would be identifiable in any phenotypic screening. They also have low sequence homology, so any large scale discovery based on sequences has been difficult (Lambalot et al., 1996). PPTases themselves have utility in a wide variety of biotechnological applications. Many of the PPTases currently used in these applications (e.g. Sfp, AcpS), are likely employed on a historical basis, i.e. these were among the first PPTases discovered and there has not been a comprehensive analysis or effort to find superior alternatives. Due to the fact that we had found such a large number of new PPTase enzymes in the course of this screen, plus the fact we could use BpsA in order to easily measure the activity and specificity of these enzymes, the characterisation of these PPTases became the subject of the next research chapter.

### 7.2.3 Characterisation of PPTases

In the various applications for which PPTases are employed, PPTases with either a particularly broad or particularly narrow substrate specificity for different CP domains are potentially advantageous in different applications. For instance a PPTase with a broad range of specificity would be generally useful for heterologous expression of NRPS/PKS genes. This may be especially useful in the future for engineered NRPS/PKS enzymes which may combine modules and therefore CP domains which have not existed in the same species before and may require an exceptionally

broad range PPTase; or for metagenome derived NRPSs or PKSs which again may require a unique PPTase to maximise their functionality. PPTases with narrower substrate specificity might also prove advantageous, for example in site specific protein labelling applications, either improving existing dual systems of labelling, or else enabling the specific labelling of three or more target sites.

As one of the key results of the previous chapter had been recovery of a large number of novel PPTase enzymes, the next aim of this research was to characterise these eDNA PPTases. It was originally hypothesised that the screening method, whereby different PPTases had been recovered based on their ability to recognise different CP-substituted BpsA variants, would have selected for a subset of PPTases that had a narrow substrate specificity. To test whether this was in fact the case, PPTase genes were cloned individually into pET28a(+) expression vectors, and analysed in a series of *in vitro* and *in vivo* assays to monitor their activity with both WT and CP-substituted BpsA variants. Unfortunately it became apparent that none of the eDNA PPTases exhibited a profoundly narrow CP domain substrate specificity, with each of the PPTases able to activate every CP-substituted variant of BpsA they were tested with. The enzymes also proved to be extremely difficult to purify, with only half of these eDNA PPTases able to be purified and characterised *in vitro*. Disappointingly, the majority of these exhibited very low activity in these *in vitro* tests. However, one of the eDNA PPTases, NZ 9, did appear to have a relatively high activity level, orders of magnitude higher than the other eDNA PPTases tested. NZ 9 was also evaluated in a competition assay, which assessed its ability to activate other CP domains relative to BpsA. It was found that NZ9 could activate all the CP domains it was tested against including both PCP and ACP domains. This, along with the fact it was so active *in vitro* (which implies the enzyme was relatively stable), may indicate that this could be a potential PPTase to use in applications that require a broad activity PPTase.

A selection of other PPTases (Sfp, Gsp, MtaA and EntD), which had been characterised to some extent previously, were also tested via the competition assay, and all showed at least some degree of activity with all six CP domains they were tested with. This is consistent with published literature, which describes the majority of type II PPTases as having a broad substrate specificity. As none of the discovered eDNA PPTases nor any other PPTases tested appeared to possess an especially narrow substrate specificity, the final aim of this thesis employed directed evolution techniques in an attempt to narrow the specificity of a PPTase.

### 7.2.4 Engineering of PPTase substrate specificity

In order to engineer a PPTase with narrow substrate specificity, a directed evolution strategy was developed. Some of the specific PPTase/CP pairs that have previously been developed have low efficiencies of the labelling reactions, or exhibit a level of background activation whereby PPTases are still able to label substrates they are meant to have lower activity with; so there is room to improve upon these systems. Other studies have focused in the past on evolving the CP substrates to create more specific PPTase/CP pairs, but prior to the present study no work had been performed in terms of engineering the PPTases themselves. As part of the work characterising the eDNA PPTases in Chapter 5, three of the PPTases had been cloned as truncated versions. It was noted that these different versions of the same PPTase displayed different levels of specificity with some of the CP-substituted variants tested. This provided a first indication that it should be possible to engineer the sequence of a PPTase in order to change its substrate specificity.

A novel screen for the evolution of PPTase substrate specificity was presented in Chapter 6. A dual selection was developed wherein evolved variants of the PPTase *entD* were selected for on iron limiting media (selecting variants that retained the ability to activate the CP domains of the enterobactin siderophore synthesis pathway, one of which is the CP of EntF). These *entD* variants were co-expressed with the *bpsA* gene, and also selected for the loss of ability to activate the wild type BpsA CP, by selecting variants that exhibited decreased levels of indigoidine production. The result was the discovery of several variants with narrowed substrate specificity, i.e. those that had retained activity with their native CP substrates while losing activity with BpsA. The top variant, Clone 1, appeared to have an even greater activity with EntF than WT *entD* in an *in vivo* in liquid culture assay. However, when this variant was purified and tested in an *in vitro* kinetic assay the overall catalytic efficiency of this variant in fact appeared to be substantially decreased in comparison to WT. The reasons for this anomaly are unknown, but may reflect diminished stability of the evolved variant *in vitro*. Irrespective, it was concluded that the screen that was developed here is likely to prove effective when applied to a larger library size, and moreover has the ability to be adapted to evolve other PPTases as well, which will help in the goal to find multiple specific PPTase/CP pairs.

## 7.3 Critical evaluation of BpsA assays used in this study

### 7.3.1 Comparison of results seen in *in vivo* vs *in vitro* BpsA assays

There were several different approaches taken for measuring either BpsA or PPTase activity via indigoidine production. Indigoidine production was either measured *in vivo*, for the co-expression of *bpsA* and PPTase genes in *E. coli*, on solid media or in liquid media; or else measured *in vitro* for purified enzymes. Each approach has advantages but also limitations, and an understanding of these limitations is required in order to fully appreciate the appropriate situations to use each method, and also to understand any discrepancies in results when measuring the same enzymes across multiple assays. Measuring indigoidine production on agar plates is the most high throughput method, with millions of colonies able to be screened relatively easily. The time taken for indigoidine production to become apparent on agar plates can be taken as a rough estimate of PPTase activity, but this is the least quantitative method, particularly when using the scoop method for administration of IPTG. Due to the high throughput nature of this method it is best for situations where screening of large numbers of colonies is required, and so was used to find initial hits in both directed evolution and metagenomic screening (Chapters 3, 4 and 6). For a more quantitative approach, *bpsA* and PPTase containing strains were grown in liquid media and indigoidine production was quantitatively assessed through measurement of absorbance at 590 nm. This is more of a medium throughput approach and is good when some level of activity is already known. One of the limitations of this method is that the assay has a relatively short linear range, and hence is not an effective means to measure the activity of PPTase or BpsA variants that vary greatly from one another. In practice this means that if using this method when the activities of the tested variants are too low, comparison with WT enzymes cannot be accurately made. This is largely due to the fact that indigoidine is unstable in aqueous solution, and either precipitates out of solution or converts to a colourless form. After a peak of indigoidine concentration is seen, the levels then appear to drop (e.g., Figure 5.5). Thus, when comparing two variants that have substantially different activities, at a given time point one variant may still be in a stage of increasing indigoidine production while the other may have already reached a peak and be decreasing. This problem of indigoidine instability is also seen in *in vitro* assays, but is not as much of an issue here as each reaction is run over a much shorter period of time (15-90 min) and measured at much smaller intervals. This is the most quantitative method, giving kinetic parameters for enzymes measured. However, it is also the most low throughput, as purified enzymes are required.

These measures of indigoidine production were often inconsistent when results for the same enzyme were compared in different assays. For example, in Chapter 3 some CP-substituted BpsA variants appeared slow acting on solid media and fast acting on liquid media, though these were both *in vivo* assays. The reasons for this were unclear but the liquid assay was likely more accurate (as they are more quantitative). Another difference seen for the solid *vs.* liquid media *in vivo* assays was dependent on the method of induction for the PPTase. For example, in Chapter 3, where CP-substituted BpsA variants were tested for their ability to produce indigoidine, the solid media assay appeared to be more sensitive, with the strains containing the lowest activity variants able to produce indigoidine on solid media despite not producing measurable levels of indigoidine in liquid media. However, the opposite result was observed in Chapter 6, measuring the activity of the engineered *entD* variants. Here the liquid assay was more sensitive for the lowest activity variants and indigoidine production was seen in the liquid assay whereas none was seen on solid media. This is likely due to the different methods of induction employed, with the PPTases in Chapter 6 induced via the pBAD promoter using arabinose. Compared to IPTG induction, this is a tighter method of induction and levels of PPTase expression on solid media in this instance were likely too low for activity to be seen. The level of induction is therefore likely something that needs to be taken into account, especially with the solid agar assays being used more for larger screening efforts. If the goal of a screen was to recover PPTases with higher activity, then decreasing the levels of induction in agar plates may be the appropriate course of action. Conversely if the goal were to recover clones with even small amounts of activity, as might be the case with metagenomic screening, then higher amounts of induction would be required.

The other main issue was the difference seen between the *in vitro* method and the aforementioned *in vivo* screens. Often the activity levels seen *in vitro* were much lower than they had appeared to be *in vivo*. For example, in Chapter 5, Svp displayed no activity *in vitro* despite showing a high amount of activity *in vivo*. It was also seen in Chapter 6 where all of the evolved *entD* variants had extremely low activity compared to WT *entD*, which did not reflect in the results of the *in vivo* assays. One explanation for the inconsistencies seen may be due to problems with the stability of the purified protein. For many *in vitro* downstream applications a purified protein would be required, so this will require further investigation in future work.

### 7.3.2 Limitations of the BpsA assays used to judge the specificity of PPTases

As all of the assays used in this thesis to assess the activity of PPTases were based on the enzyme BpsA, this means they are only describing the ability of a given PPTase to activate BpsA, which is

not the native substrate of any of the PPTases tested here. In order to test the ability of PPTases to activate other CP domains, two methods were used. First the CP-substituted BpsA enzymes were employed, offering an alternative CP substrate for a given PPTase to activate while still producing indigoidine as the coloured product to report activity. This approach is limited however in terms of the range of CP domains that could be tested, as creating functional CP-swaps is a very laborious process. The other way of testing the activity of a PPTase with alternative CP domains was to use the competition assay, where BpsA and any alternative CP domain compete for a pool of limited CoA, which allows calculation of the preference a given PPTase shows for the competing CP domain as compared to WT BpsA. It is easier to test a wider variety of CP domains using this assay as no enzyme evolution is required, and therefore all of the CPs have their WT sequence. These cannot be compared directly due to the fact that each PPTase has a different level of activity with BpsA to begin with but an approximation can be calculated based on the kinetic parameters of the PPTases with respect to BpsA (Owen et al., 2011). This work will be done in the future for the PPTases investigated here.

## 7.4 Future directions

### 7.4.1 BpsA in metagenomic screening

The work described in this thesis validated the use of BpsA as a tool for the discovery of both PPTase genes and their associated natural product biosynthetic gene clusters. As both of the libraries screened in the current study were small insert eDNA libraries, it was not possible to recover entire biosynthetic gene clusters. Future work will focus on the screening of libraries containing much larger eDNA fragments (using cosmids or bacterial artificial chromosomes) in order to discover entire biosynthetic gene clusters. An advantage of finding complete gene clusters for natural product genes is that the gene clusters may be able to be functionally expressed in *E. coli,* or other suitable heterologous expression strains, and ultimately used to produce a novel natural product, as an alternative to chemical synthesis.

In the screen described in this thesis, each of the CP-substituted BpsA variants were plasmid-based, and libraries were screened repeatedly in different strains containing each plasmid. An option for improving upon this system for future experiments is to use genomic incorporation to integrate multiple *bpsA* variants into the chromosome of a single host strain. This would cut down on the amount of screening that needs to be done. Another way to improve upon this screen in the future would be the creation of even more CP-substituted BpsA variants, using even more diverse CP domains (e.g., including ACPs and ArCPs). The work described in Chapter 3 could be

used to guide the creation of such variants, i.e. by selecting CP domains that are found immediately upstream from TE domains in their native context. Low functioning substitutions could then be improved in a targeted fashion via site directed mutagenesis at the +4 and +24 sites, a more straight-forward approach than a completely random mutagenesis, and easier to implement. The combination of choosing CP domains (upstream from TE domains) that are more likely to be functional, and the knowledge of the exact residues to target if improvements in function are required, means that CP domains will overall be changed very little from their original sequence. This is an advantage as the whole point of using these CP-substituted variants is to present alternative sites of recognition for the PPTase enzymes. Helix II is thought to be the part of the CP domains that the PPTase recognises, and it is only the +4 residue that is in this helix, while the +24 residue is in helix III, indicating that evolving CP-substituted BpsA variants in this manner should not dramatically alter the site of PPTase recognition.

An important factor to take into account however, is that the work in Chapter 5 indicated that all of the PPTases discovered in Chapter 4 could in fact recognise all of the CP-substituted BpsA variants, but some were just not adequately expressed within the eDNA inserts. Another future direction then would be to investigate ways of increasing the gene expression of the eDNA inserts. Increasing these expression levels may be of even more of a concern when looking at larger insert vectors like BACs (with a given PPTase likely to be even further away from any strong promoters the plasmid may possess). Two approaches to increase expression may be to use alternative host strains, e.g. *Streptomyces* species that are better suited to heterologous expression of many natural product biosynthetic genes; or else using transposon mediated random insertion of promoters into the DNA fragments (Troeschel et al., 2010).There has also been work showing that expression of heterologous sigma factors in *E. coli* can result in these strains being able to recognise a wider number of heterologous promoters (Gaida et al., 2015).

### 7.4.2 Engineering of PPTases to have altered CP substrate specificity

The work in Chapter 6 showed that it was possible to engineer altered substrate specificity for the PPTase EntD, and the strategy developed here should be applicable to other PPTase enzymes. If a PPTase knock out strain is available, then variants can be screened in the same manner, selecting for siderophore production through growth on iron limiting plates. As part of ongoing work in our lab by Masters student Jack Sissons, EntD is being engineered to lose activity with a different CP substrate (the PvdDCP1 domain, screening for decreased indigoidine production in the reporter strain EcoBlueP). In parallel, the PPTase PcpS from *P. aeruginosa* is also being evolved to retain

activity with PvdDCP1 (through growth on iron limiting media, necessitating production of the NRPS pyoverdine) and to lose activity with the EntF CP-substituted variant of BpsA (variant libraries of *pcpS* being screened for loss of indigoidine production in the EcoBlueE reporter strain).The aim of this work is to create a complementary pair, an EntD variant which has activity with EntF and no activity with PvdD, and a PcpS variant which has activity with PvdD and none with EntF. These could then be used for the application of site specific protein labelling, and perhaps combined with existing specific PPTase/CP pairs and used to label a number of different proteins with no cross reactivity.

The proof-of-principle EntD evolution study described here focused on the ability of EntD to recognise (or not recognise) the entire CP domain. The ability of the evolved variants to specifically activate a CP as a distinct protein, as opposed as a domain within BpsA, needs to be investigated further. Additional to this, ongoing work in our lab is also investigating the potential to use small protein tags in the place of an entire CP domain. Small tags based on the CP domain sequences are being designed and tested to see whether the PPTase is capable of activating them and still showing specificity. Another future direction is the use of BpsA to screen for such small tag sequences, by substituting just the helix II region of BpsA. This region has been substituted before by (Mofid et al., 2002), who showed that altering this region alone is enough to influence which PPTase activates a target CP domain. Another option for a future experiment would be to use a directed evolution approach to alter just this region, helix II of the BpsA, and thereby evolve a tag for greater specificity with a PPTase.

The enzyme engineering that has been performed to alter the CP domain substrate specificity of EntD, and any possible future work to alter the specificity of other PPTases, could also be applied to better understand the interaction between a PPTase and CP domain. Though this interaction is important in the physiological function of PPTases, as well as for various PPTase applications, little is known about how PPTases recognise specific CP domains. Only one type II PPTase has had its structure solved together with a PCP partner, showing the regions that are important for the general interaction. Without the comparison of other CP domains, it is difficult to infer the extent to which this interaction may differ for different substrates, and what drives the specificity a PPTase has for one CP over another. A mutational analysis is another approach to answer these questions. A pattern may emerge if several PPTases are evolved (starting with *entD* and *pcpS*) for specificity with particular substrates.

An important aspect of future work would be the need to tease apart individual mutations, as with so many multi-site mutants generated via epPCR it can be difficult to assess the role of specific

mutations. Mutations that are found repeatedly, such as Y88N in the evolved EntD variants (Chapter 6), need to be tested as single mutations. This also may help with the fact that all of the evolved *entD* variants appeared to have lost so much activity when tested *in vitro*. This may have been caused by protein instability which may in turn have been a result of introducing too many mutations into the WT enzyme. If we could elucidate only the key amino acids to change in order to alter substrate specificity, this may lead to more active evolved variants.

## 7.5 Concluding remarks

This thesis describes the development and implementation of several assays that use BpsA as a tool to give insights into both PPTases and CP domains. Firstly, further insights were gained about specific features of CP domains that promote formation of functional CP-substituted BpsA variants. Functional CP-substituted variants as well as WT BpsA were then used in a series of screens and assays to both discover and characterise novel PPTase enzymes. Upon finding all PPTases to have a broad specificity across the tested CP domains, a novel screen was developed in order to engineer PPTase substrate specificity. Also based on the activity of BpsA, this screen was used to engineer the PPTase EntD in a proof of principle experiment and was successful in creating a PPTase with more narrow substrate specificity. This will form the basis of future work for developing PPTase/CP pairs which have non-overlapping activities.

# Appendix 1: BpsA as a tool to quantify L-glutamine

Work was undertaken contributing to a project not directly relating to this thesis, the development of BpsA as a biosensor in order to quantify L-glutamine. My role in this work was to perform initial experiments testing the ability of BpsA to quantify L-glutamine in complex cell culture media (LB and DMEM), this resulted in a patent of which I was one of the co-authors (Ackerley et al., 2015). Following is a fact sheet produced by Viclink, Victoria University of Wellington's commercialisation office.

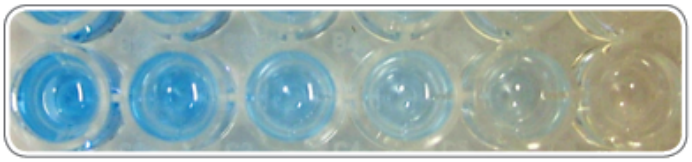
# *RAPID AND SENSITIVE L-GLUTAMINE MEASUREMENT

## *Glutamine Quantification Kit*

**VICTORIA UNIVERSITY OF WELLINGTON (VUW) RESEARCHERS HAVE DEVELOPED A NOVEL COLOURMETRIC ASSAY FOR DIRECT AND SPECIFIC MEASUREMENT OF GLUTAMINE.**

**KEY FEATURES:**

- **Rapid**
  Under half the time when compared to other tests.
- **Robust**
  The assay works in a range of conditions without the need for prior sample processing.
- **Sensitive**
  The assay has a wide linear detection range, accurate between 50µM and 1000µM.
- **Simple**
  Simple and amenable to high-throughput.
- **Direct**
  Unlike other diagnostic kits, glutamine is converted directly to a detectable product. This reduces complexity when working with samples that also contain glutamate.



**THIS NOVEL COLOURMETRIC ASSAY PROVIDES A SIMPLE SOLUTION FOR GLUTAMINE MEASUREMENT IN COMPLEX BIOLOGICAL SOLUTIONS.**

**UNLIKE OTHER COMMERCIAL PRODUCTS, THE ASSAY DEVELOPED BY VUW'S APROF DAVID ACKERLEY OFFERS DIRECT, SPECIFIC MEASUREMENT OF GLUTAMINE WITHOUT RELYING ON A SECONDARY STEP TO REMOVE GLUTAMATE FROM THE READING.**

**USING A PROPRIETARY ENZYME TO CONVERT L-GLUTAMINE INTO A DETECTABLE BLUE PIGMENT THAT CAN BE QUANTIFIED USING STANDARD LABORATORY INSTRUMENTATION, THIS TECHNOLOGY OFFERS A RAPID AND SENSITIVE SOLUTION TO L-GLUTAMINE MEASUREMENT.**

### APPLICATION DEVELOPMENT OPTIONS

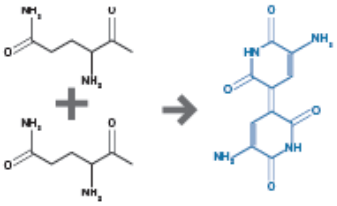
**Life Science** – Monitor tissue culture media content

**Diagnostics** – Detect metabolic disorders

**Manufacturing** – In line product testing and quality assurance

**Personal Care** – Measure glutamine levels during athletic activity to monitor for overtraining syndrome

### ENZYME-BASED CONVERSION



Synthesis of blue pigment from two molecules of glutamine, catalysed by our enzyme.

| | VUW Solution | Competitor Solution |
|---|---|---|
| **Assay Time** | <Less than 45mins (including incubation time of 30 mins) | Approximately 2 hours (including 100 min incubation) |
| **Sample size** | Minimum of 10µl of sample size. | Minimum of 250µL sample size. |
| **Detection Range** | 50µM to 1000µM | 50µM to 1000µM |
| **Substrate Measurement** | Simple direct measurement of glutamine. | Requires glutamine to be converted to glutamate. |
| **Complexity of kit** | 5 reagents | 9 reagents. Buffers in the kit are also required to be pH verified. |
| **Sample format** | Water, dilute organic solvents, blood plasma, urine, cell culture media, bacterial growth media. | Secondary processing required for complex biological samples. |
| **Shelf life** | >12 months at 4°C | >12 months at 4°C |

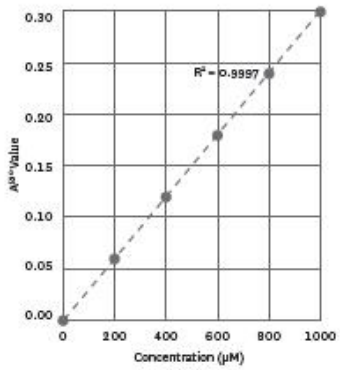
viclink

CONTINUED OVERLEAF...

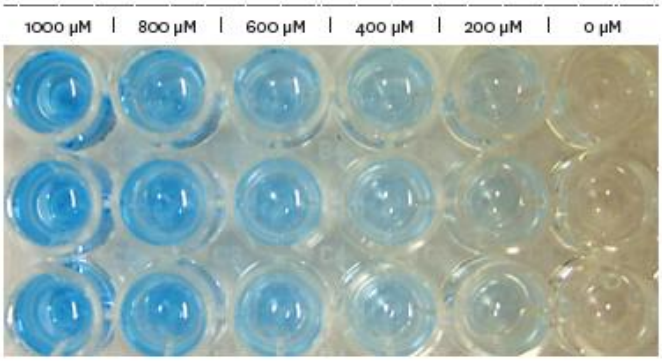
## *Glutamine Quantification Protocol*

**Step 1:** Combine Buffer 1, Enzyme, and ATP to produce Activation Buffer.
**Step 2:** Add 30 µl of Activation Buffer to each well in a 96 well plate.
**Step 3:** Add 10 µL of Standard or Sample and mix thoroughly.
**Step 4:** Incubate at room temperature for 30 minutes.
**Step 5:** Heat Stop Buffer to 60°C.
**Step 6:** Add Stop Buffer 190 µl to each well and mix thoroughly.
**Step 7:** Read plate at 590 nm to generate standard curve and sample values.

**L-GLUTAMINE STANDARD CURVE**




**L-GLUTAMINE CONCENTRATION**




**KIT COMPONENTS**

Buffer 1 (Tris-Cl, pH 7.8)
Enzyme
ATP
Glutamine standards
Solubilisation buffer

VICTORIA LINK LTD (VICLINK) HELPS STAFF AND STUDENTS TURN THE WORLD-CLASS RESEARCH AT VICTORIA UNIVERSITY INTO COMMERCIAL OPPORTUNITIES.

VICLINK IS A TECHNOLOGY TRANSFER COMPANY WHOLLY OWNED BY VICTORIA UNIVERSITY. THE COMPANY DRAWS ON YEARS OF EXPERIENCE TO IDENTIFY UNIVERSITY IP WITH COMMERCIAL POTENTIAL AND CONNECT IT WITH INDUSTRY THROUGH LICENSING OR NEW VENTURE FORMATION.

VICLINK OFFERS BUSINESSES, INVESTORS AND ACADEMICS SMART CONNECTIONS.

FOR FURTHER INFORMATION REGARDING LICENSING OPPORTUNITIES FOR THIS TECHNOLOGY, PLEASE CONTACT:

**DR ANNE BARNETT**
**SENIOR COMMERCIALISATION MANAGER**
**VICLINK**
**PH: +64 4 463 9456**
**MOB: +64 22 1088 110**
**ANNE.BARNETT@VICLINK.CO.NZ**

CONTACT US
**INFO@VICLINK.CO.NZ**
**OR +64 4 463 5135**

POSTAL ADDRESS
**VICTORIA LINK LTD, PO BOX 1762**
**WELLINGTON 6140, NEW ZEALAND**

STREET ADDRESS
**EA120, EASTERFIELD BUILDING**
**KELBURN PARADE, KELBURN**
**WELLINGTON 6012, NEW ZEALAND**

For more information about Viclink, visit
***www.viclink.co.nz***

viclink

# Appendix 2: Sequences of synthesised genes

Sequences in italics denote sequence that was added (e.g. containing restriction sites) to the gene

**Sfp**

*CCCCCAT*ATGAAAATCTACGGTATCTACATGGACCGTCCGCTGTCTCAGGAAGAAAACGAACGTTTCATG
ACCTTCATCTCTCCGGAAAAACGTGAAAAATGCCGTCGTTTCTACCACAAAGAAGACGCGCACCGTACCCT
GCTGGGTGACGTTCTGGTTCGTTCTGTTATCTCTCGTCAGTACCAGCTGGACAAATCTGACATCCGTTTCT
CTACCCAGGAATACGGTAAACCGTGCATCCCGGACCTGCCGGACGCGCACTTCAACATCTCTCACTCTGG
TCGTTGGGTTATCGGTGCGTTCGACTCTCAGCCGATCGGTATCGACATCGAAAAAACCAAACCGATCTCTC
TGGAAATCGCGAAACGTTTCTTCTCTAAAACCGAATACTCTGACCTGCTGGCGAAAGACAAAGACGAACA
GACCGACTACTTCTACCACCTGTGGTCTATGAAAGAATCTTTCATCAAACAGGAAGGTAAAGGTCTGTCTC
TGCCGCTGGACTCTTTCTCTGTTCGTCTGCACCAGGACGGTCAGGTTTCTATCGAACTGCCGGACTCTCAC
TCTCCGTGCTACATCAAAACCTACGAAGTTGACCCGGGTTACAAAATGGCGGTTTGCGCGGCGCACCCGG
ACTTCCCGGAAGACATCACCATGGTTTCTTACGAAGAACTGCTGTAA*GTCGACCCCC*

**Svp**

*CCCCCATATG*ATCGCTGCTCTGCTGCCGAGTTGGGCTGTAACCGAACACGCCTTCACCGATGCGCCGGAC
GACCCGGTGAGCCTGCTTTTCCCGGAAGAGGCGGCACACGTGGCCCGTGCGGTTCCGAAACGCCTGCAT
GAATTTGCCACCGTCCGTGTTTGTGCTCGTGCGGCGTTGGGTCGCCTGGGCCTGCCTCCGGGCCCGCTGT
TACCGGGCCGTCGCGGCGCACCGTCCTGGCCGGACGGTGTAGTGGGTTCGATGACTCACTGTCAGGGTT
TCCGCGGTGCGGCTGTGGCACGAGCGGCTGACGCCGCTTCCCTGGGCATCGACGCTGAACCGAACGGTC
CGCTGCCGGATGGTGTTCTGGCAATGGTTAGTCTGCCGTCCGAACGTGAATGGCTGGCAGGTCTGGCCG
CTCGCCGCCCGGACGTCCATTGGGACCGTCTGCTGTTCAGCGCGAAAGAATCTGTGTTTAAAGCCTGGTA
TCCGCTGACCGGTCTGGAACTGGATTTCGATGAAGCGGAACTGGCGGTTGATCCGGACGCGGGTACCTT
CACTGCTCGCCTCCTGGTTCCGGGTCCGGTGGTAGGTGGCCGTCGTCTGGACGGCTTTGAAGGTCGTTGG
GCGGCAGGTGAAGGTCTGGTTGTTACCGCCATTGCGGTTGCAGCTCCAGCAGGCACCGCAGAAGAATCT
GCCGAGGGCGCCGGCAAAGAGGCTACTGCAGATGACCGCACTGCGGTTCCGTAA*GTCGACCCCC*

**Gsp**

*CCCCCAT*ATGATTGAAATGCTGTTCGTAAAAGTCCCGAACGAAATCGACCGTCACGTTTTCAATTTTCTGA
GCTCCAACGTTTCCAAAGAAAAACAACAGGCTTTCGTTCGCTACGTCAACGTTAAGGATGCTTATCGTTCC
CTGCTGGGCGAGCTGTTGATCCGCAAATACCTGATCCAGGTCTTAAACATCCCTAATGAAAACATCCTGTT
CCGTAAAAACGAATACGGCAAACCGTTCGTAGATTTCGACATTCACTTTAACATCTCTCACTCCGATGAAT
GGGTTGTTTGCGCTATTTCTAACCATCCGGTGGGTATCGACATTGAGCGTATCTCCGAGATCGACATCAA
GATCGCAGAACAGTTCTTCCACGAAAACGAATACATCTGGCTGCAGTCGAAAGCCCAAAACTCTCAGGTG
TCTTCTTTTTTTGAGCTGTGGACCATTAAAGAGAGCTATATTAAGGCCATTGGTAAAGGCATGTATATTCC
TATCAACTCTTTTTGGATCGATAAGAACCAGACTCAGACGGTGATTTATAAACAAAACAAAAAAGAACCG
GTTACCATCTATGAACCGGAACTGTTCGAAGGTTACAAGTGCAGTTGCTGCAGCCTTTTCAGCTCTGTTAC
GAACCTCAGCATCACCAAACTCCAGGTTCAGGAACTGTGCAACCTGTTCCTGGATTCTACCTTTAGCGAAA
ACAACAACTTCTAG*GTCGACCCCC*

**MtaA**

*CCCCCAT*ATGCCGACTTCCAGCCCGGCTTTGCCGCTGCTGAAACTGCCGCCGGATGAAGTACACGTTTGG
ATTGTTGAACCGGAACGTATCACAGAGCCAGGTCTGCTCGAAAGCTACCGCGCGCTCCTAGATCCGGGC
GAACGCGACAAACAGCAGCGTTTCTACTTCGAACGTCATCGTCTGCAATACCTGGTTAGCCATGCGCTGG
TACGTCTGACCCTGAGTCGTTATGCCCCGGTTGCACCGGAAGCTTGGTCCTTTTCTGCGAACCAGTACGGT
CGTCCGGAAATCCGCGGTGAAGAAAAACCGTGGCTGCGTTTCAACCTGTCGCACACCGACGGTATGGCT
CTGTGCGCTGTCGCCCGTGATGTGGACGTAGGTGCAGATGTAGAGGACACCGAGCGCCGCGGTGAAACC
GTCGAAATCGCGGATTCTTTCTTCGCGCCGGCAGAAGTTGCGTCCCTGCGTGCTCTCCCGGTTAGCGGCC
AGCGTGAACGCTTTTTTGACTACTGGACCCTGAAAGAGGCGTACATTAAAGCTCGCGGGATGGGCCTGA
GCCTCCCGCTGGATCAGTTCGCGTTCGAAGTATCCCAGGGCCTGAGTACCCGCATTAGCTTCGACCCGCG
CCTGGTAGACGAACCGTCCCAGTGGCAGTTCGTCCGTTTTCGTCCGAGCCAGCGCCACGCAGCGGCGCTG
GCTGTCCGTCGTCCTTCCGAAGCGCCGCTGACCGTTCGTTTCCAACGTACCGTCCCGCTGCAGGACGATGC
ACCGGCTGAATATCTGAGCCGTGAGCGTATCCAGCCGCTGCGTCTCCGTATGCCGGGTGTTGGCGGCGG
TTGA*GTCGACCCCC*

**ACMSIII**

*AGCTCTGCAGAGCTCGTCGAGCGCCCC*GGCCGTGCACCGCGTACGCCGCAGGAACAGGTTGTCTGCGAA
CTGTTCGCGGAAGTGCTGGGTCGCCCGTTGGTTGGGGTCGATCAGGATTTCTTCGATCTGGGTGGTCACA
GCCTGCTGGCAACACGTCTTATCGCTCGTCTGCGTGCAGCGTTCGGTGTAGAACTGGGCTTACGTTCCCTG
TTCGAAGCACCTACGCCGGGTGGTATTGCAGCCCGTTTGGA*CCGCGAGGTCGCCCAGGAGTCTAGAAGC*
*T*

**GrsB**

*AGCTCTGCAGAGCTCGTCGAGCGCCCC*TACGTGGCGCCGCGTAACGAAATCGAAGCACAGCTGGTACTG
ATTTGGCAAGAAGTTCTGGGTATCGAACTGATCGGCATCACCGACAACTTTTTCGAACTCGGTGGTCACTC
TCTCAAGGCAACGCTGCTCGTCGCGAAAATCTACGAATATATGCAGATCGAAATGCCACTGAACGTGGTG
TTCAAACACAGCACCATTATGAAAATTGCAGAATATATTACT*CGCGAGGTCGCCCAGGAGTCTAGAAGCT*

**TycC**

*AGCTCTGCAGAGCTCGTCGAGCGCCCC*TACGTAGCACCGCGTAACGCTACCGAACAACAGCTGGCAGCTA
TCTGGCAGGAAGTGCTGGGTGTGGAACCGATTGGCATCACCGATCAGTTCTTCGAACTGGGCGGTCACA
GCCTTAAGGCTACCCTCCTGATCGCGAAAGTGTATGAGTATATGCAGATCGAACTCCCGCTGAACCTGAT
TTTCCAGTATCCGACTATCGAAAAAGTGGCGGATTTCATTACC*CGCGAGGTCGCCCAGGAGTCTAGAAGC*
*T*

**Pris3**

*AGCTCTGCAGAGCTCGTCGAGCGCCCC*GGTCGTCCGCCACGTGGTCCACGCGAAGAAATCCTGTGTGGCC
TGTTCGCTGAAGTTCTGGGTGCACCGCGTGTTGGCACTGACGACAACTTCTTCGAACTGGGCGGTCACTC
TATGCTGGCCACCCGTCTGGTCGGCCGTGTAAAAACCGTCCTGGGTGCAGACATCGGTGTGCGTACTCTG
TTCGAAGCACCGACGGTTGCTGCTCTGGCAGCACGTATTGAT*CGCGAGGTCGCCCAGGAGTCTAGAAGCT*

**BlmI**

*CCCCCAT*ATGAGTGCGCCACGTGGTGAACGTACCCGTCGTCGCGCTCTGGAACGTGACATCGCAGCGATT
TGGGCAGAAACTCTCGGCCGTGATAGCGTTGGGCCGCATGAAGATTTCGCCGCACTGGGCGGCAACTCG
ATCCATGCGATCAAAATTACCAACCGTGTGGAAGAACTCGTCGATGCTGAACTTTCTATTCGTGTTCTGCT
GGAAACCCGTACCGTGGCTGGTATGACCGACCACGTACATGCTACCCTGACGGGTGAGCGTGACCGTTG
A*GTCGACCCCC*

**SgcC2**

*CCCCCATATG*TCTACTGTGTCCGACACCGCCGCAGGCTCTTCTCTGGAAGAAAAAGTGACGCGTATCTGG
ACCGGTGTGCTGGGCACGTCAGGTGAAGAAGGTGCGACCTTTATTGAACTGGGTGGTCAATCCGTTTCC
GCGGTCCGCATCGCAACTCGTATCCAGGAGGAACTGGACATCTGGGTAGATATCGGCGTGCTGTTCGAT
GATCCGGACCTGCCGACGTTCATTGCGGCTGTTGTGCGTACTGCCGATGCGGCTGGCGGCGAAGGTAGC
GGTACCCAGTGA*GTCGACCCCC*

**MtaF**

*CCCCCAT*ATGTCTGCGTTCGCTAACGCTCTGAGCTCTGCCGAACCGAGCCAGCGTGGCCAGCTGCTGGGC
TCCTACTTATGCGAACAGGTTGCACACCTGCTGAAGATGCCGGTTGCGAAACTGGACCCGGAACAGCCTT
TAAACTCAATGGGTATGGACAGCCTGCTGAGCCTGGAACTGAAACATAAGATTCACGCGGAAACCGGCA
TGGATGTACCGCTGGACGAAGTTCTGCAGGGCGCCTCAATCGCTAACCTGTCGATGCGCCTGGCGGAAC
GTATGGGTAATTCCGGCCAGGCCGCG*TGAGTCGACCCCC*

**TcmM**

*CCCCCATATG*CCTCAGATCGGCCTGCCGCGTCTGGTGGAGATCATTCGTGAATGCGCCGGCGATCCGGAC
GAGCGTGATCTGGACGGTGATATCCTGGACGTTACCTATCAGGATCTGGGTTACGACTCTATCGCTCTGC
TGGAAATCTCTGCAAAACTGGAACAAGATTTGGGCGTTTCTATTCCGGGCGAAGAACTGAAGACGCCAC
GCCATACGCTGCACCTGGTTAACACTGAAACCGCTGGTGAGGTTGCGTGA*GTCGACCCCC*

Ackerley, D.F., Caradoc-Davies, T.T., and Lamont, I.L. (2003). Substrate specificity of the nonribosomal peptide synthetase PvdD from *Pseudomonas aeruginosa*. J. Bacteriol. *185*, 2848–2855.

Ackerley, D.F., Brown, A.S., and Robins, K.J. (2015). Methods of Detecting and Measuring Glutamine and Analogues Thereof, and Methods Related Thereto. PCT/NZ2014/050017

Allen, E.E., and Bartlett, D.H. (2002). Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9. Microbiology *148*, 1903–1913.

Asghar, A.H., Shastri, S., Dave, E., Wowk, I., Agnoli, K., Cook, A.M., and 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*, 349–361.

Baltz, R.H. (2002). Combinatorial glycosylation of glycopeptide antibiotics. Chem. Biol. *9*, 1268–1270.

Banik, J.J., and Brady, S.F. (2010). Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. Curr. Opin. Microbiol. *13*, 603–609.

Barekzi, N., Joshi, S., Irwin, S., Ontl, T., and Schweizer, H.P. (2004). Genetic characterization of *pcpS*, encoding the multifunctional phosphopantetheinyl transferase of *Pseudomonas aeruginosa*. Microbiology *150*, 795–803.

Beer, R., Herbst, K., Ignatiadis, N., Kats, I., Adlung, L., Meyer, H., Niopek, D., Christiansen, T., Georgi, F., Kurzawa, N., Meichsner, J., Rabe, S., Riedel, A., Sachs, J., Schessner, J., Schmidt, F., Walch, P., Niopek, K., Heinemann, T., Eils, R., et al. (2014). Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. Mol. Biosyst.

Beld, J., Sonnenschein, E.C., Vickery, C.R., Noel, J.P., and Burkart, M.D. (2013). The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. Nat. Prod. Rep. *31*, 61–108.

Berdy, J. (2005). Bioactive microbial metabolites - A personal view. J Antibiot *58*, 1–26.

Blackwell, J.R., and Horgan, R. (1991). A novel strategy for production of a highly expressed recombinant protein in an active form. FEBS Lett. *295*, 10–12.

Borchert, S., Stachelhaus, T., and Marahiel, M.A. (1994). Induction of surfactin production in *Bacillus subtilis* by *gsp*, a gene located upstream of the gramicidin S operon in *Bacillus brevis*. J. Bacteriol. *176*, 2458–2462.

Brachmann, A.O., Kirchner, F., Kegler, C., Kinski, S.C., Schmitt, I., and Bode, H.B. (2012). Triggering the production of the cryptic blue pigment indigoidine from *Photorhabdus luminescens*. J. Biotechnol. *157*, 96–99.

Bunkoczi, G., Pasta, S., Joshi, A., Wu, X., Kavanagh, K.L., Smith, S., and Oppermann, U. (2007). Mechanism and substrate recognition of human holo ACP Synthase. Chem. Biol. *14*, 1243–1253.

Cabail, M.Z., Li, S., Lemmon, E., Bowen, M.E., Hubbard, S.R., and Miller, W.T. (2015). The insulin and IGF1 receptor kinase domains are functional dimers in the activated state. Nat. Commun. *6*.

Caboche, S., Leclère, V., Pupin, M., Kucherov, G., and Jacques, P. (2010). Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. J. Bacteriol. *192*, 5143–5150.

Caffrey, P., Green, B., Packman, L.C., Rawlings, B.J., Staunton, J., and Leadlay, P.F. (1991). An acyl-carrier-protein – thioesterase domain from the 6-deoxyerythronolide B synthase of *Saccharopolyspora erythraea*. Eur. J. Biochem. *195*, 823–830.

Calcott, M.J. (2014). Genetic manipulation of pyoverdine non-ribosomal peptide synthetases to identify genetic constraints to effective domain recombination. Victoria University of Wellington, School of Biological Sciences.

Calcott, M.J., and Ackerley, D.F. (2014). Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. Biotechnol. Lett. *36*, 2407–2416.

Callegari, A., Luin, S., Marchetti, L., Duci, A., Cattaneo, A., and Beltram, F. (2012). Single particle tracking of acyl carrier protein (ACP)-tagged TrkA receptors in PC12*nnr5* cells. J. Neurosci. Methods *204*, 82–86.

Chalut, C., Botella, L., de Sousa-D'Auria, C., Houssin, C., and Guilhot, C. (2006). The nonredundant roles of two 4 '-phosphopantetheinyl transferases in vital processes of *Mycobacteria*. Proc. Natl. Acad. Sci. U. S. A. *103*, 8511–8516.

Charlop-Powers, Z., Banik, J.J., Owen, J.G., Craig, J.W., and Brady, S.F. (2013). Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis. ACS Chem. Biol. *8*, 138–143.

Chen, D., Wu, R., Bryan, T.L., and 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*, 511–513.

Chirgadze, N.Y., Briggs, S.L., McAllister, K.A., Fischl, A.S., and Zhao, G. (2000). Crystal structure of *Streptococcus pneumoniae* acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis. EMBO J. *19*, 5281–5287.

Copp, J.N., and Neilan, B.A. (2006). The phosphopantetheinyl transferase superfamily: Phylogenetic analysis and functional implications in cyanobacteria. Appl. Environ. Microbiol. *72*, 2298–2305.

Copp, J.N., Roberts, A.A., Marahiel, M.A., and Neilan, B.A. (2007). Characterization of PPTNs, a cyanobacterial phosphopantetheinyl transferase from *Nodularia spumigena* NSOR10. J. Bacteriol. *189*, 3133–3139.

Crosa, J.H., and Walsh, C.T. (2002). Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiol. Mol. Biol. Rev. *66*, 223–249.

Crosby, J., and Crump, M.P. (2012). The structural role of the carrier protein – active controller or passive carrier. Nat. Prod. Rep. *29*, 1111–1137.

Cude, W.N., Mooney, J., Tavanaei, A.A., Hadden, M.K., Frank, A.M., Gulvik, C.A., May, A.L., and Buchan, A. (2012). Production of the antimicrobial secondary metabolite indigoidine contributes to competitive surface colonization by the marine *Roseobacter Phaeobacter* sp. Strain Y4I. Appl. Environ. Microbiol. *78*, 4771–4780.

Daniel, R. (2004). The soil metagenome - a rich resource for the discovery of novel natural products. Curr. Opin. Biotechnol. *15*, 199–204.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. U.S.A. *97*, 6640–6645.

Davidovich, C., Goodrich, K.J., Gooding, A.R., and Cech, T.R. (2014). A dimeric state for PRC2. Nucleic Acids Res. *42*, 9236–9248.

von Döhren, H. (2009). A survey of nonribosomal peptide synthetase (NRPS) genes in *Aspergillus nidulans*. Fungal Genet. Biol. *46*, S45–S52.

Du, L., and Shen, B. (1999). Identification and characterization of a type II peptidyl carrier protein from the bleomycin producer *Streptomyces verticillus* ATCC 15003. Chem. Biol. *6*, 507–517.

Duckworth, B.P., and Aldrich, C.C. (2010). Development of a high-throughput fluorescence polarization assay for the discovery of phosphopantetheinyl transferase inhibitors. Anal. Biochem. *403*, 13–19.

Ebisu, K., Tateno, H., Kuroiwa, H., Kawakami, K., Ikeuchi, M., Hirabayashi, J., Sisido, M., and Taki, M. (2009). N-terminal specific point-immobilization of active proteins by the one-Pot NEXT-A Method. Chembiochem *10*, 2460–2464.

Elovson, J., and Vagelos, P.R. (1968). Acyl Carrier Protein X. Acyl carrier protein synthetase. J. Biol. Chem. *243*, 3603–3611.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol. Pharm. *5*, 191–211.

Fichtlscherer, F., Wellein, C., Mittag, M., and Schweizer, E. (2000). A novel function of yeast fatty acid synthase - Subunit alpha is capable of self-pantetheinylation. Eur. J. Biochem. *267*, 2666–2671.

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

Finking, R., Solsbacher, J., Konz, D., Schobert, M., Schafer, A., Jahn, D., and Marahiel, M.A. (2002). Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. J. Biol. Chem. *277*, 50293–50302.

Fischbach, M.A., and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. *106*, 3468–3496.

Flugel, R.S., Hwangbo, Y., Lambalot, R.H., Cronan, J.E., and Walsh, C.T. (2000). Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in *Escherichia coli*. J. Biol. Chem. *275*, 959–968.

Foley, T.L., and Burkart, M.D. (2007). Site-specific protein modification: advances and applications. Curr. Opin. Chem. Biol. *11*, 12–19.

Foley, T.L., and Burkart, M.D. (2009). A homogeneous resonance energy transfer assay for phosphopantetheinyl transferase. Anal. Biochem. *394*, 39–47.

Frank, B., Wenzel, S.C., Bode, H.B., Scharfe, M., Blöcker, H., and Müller, R. (2007). From genetic diversity to metabolic unity: studies on the biosynthesis of aurafurones and aurafuron-like structures in myxobacteria and streptomycetes. J. Mol. Biol. *374*, 24–38.

Gabor, E.M., Alkema, W.B.L., and Janssen, D.B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ. Microbiol. *6*, 879–886.

Gaida, S.M., Sandoval, N.R., Nicolaou, S.A., Chen, Y., Venkataramanan, K.P., and Papoutsakis, E.T. (2015). Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. Nat. Commun. *6*, 7045.

Gaitatzis, N., Hans, A., Müller, R., and Beyer, S. (2001). The mtaA Gene of the Myxothiazol Biosynthetic Gene Cluster from *Stigmatella aurantiaca* DW4/3-1 Encodes a Phosphopantetheinyl Transferase that Activates Polyketide Synthases and Polypeptide Synthetases. J. Biochem. *129*, 119–124.

Garcin, G., Paul, F., Staufenbiel, M., Bordat, Y., Van der Heyden, J., Wilmes, S., Cartron, G., Apparailly, F., De Koker, S., Piehler, J., Tavernier, J., and Uzé, G. (2014). High efficiency cell-specific targeting of cytokine activity. Nat. Commun. *5*.

Gehring, A.M., Lambalot, R.H., Vogel, K.W., Drueckhammer, D.G., and 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. Chem. Biol. *4*, 17–24.

George, N., Pick, H., Vogel, H., Johnsson, N., and Johnsson, K. (2004). Specific labeling of cell surface proteins with chemically diverse compounds. J. Am. Chem. Soc. *126*, 8896–8897.

Gokhale, R.S., Tsuji, S.Y., Cane, D.E., and Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. Science *284*, 482–485.

Goldsmith, M., Kiss, C., Bradbury, A.R.M., and Tawfik, D.S. (2007). Avoiding and controlling double transformation artifacts. Protein Eng. Des. Sel. *20*, 315–318.

Grünewald, J., Jones, D.H., Brock, A., Chiu, H.-P., Bursulaya, B., Ng, K., Vo, T., Patterson, P., Uno, T., Hunt, J., Spraggon, G., and Geierstanger, B.H. (2014). Site-specific dual labeling of proteins by using small orthogonal tags at neutral pH. ChemBioChem *15*, 1787–1791.

Haslinger, K., Redfield, C., and Cryle, M.J. (2015). Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. Proteins Struct. Funct. Bioinforma. *83*, 711–721.

Hermes, J.D., Blacklow, S.C., and Knowles, J.R. (1990). Searching sequence space by definably random mutagenesis: improving the catalytic potency of an enzyme. Proc. Natl. Acad. Sci. U. S. A. *87*, 696–700.

Huang, Y., Wendt-Pienkowski, E., and Shen, B. (2006). A dedicated phosphopantetheinyl transferase for the fredericamycin polyketide synthase from *Streptomyces griseus*. J. Biol. Chem. *281*, 29660–29668.

Humpert, M.L., Tzouros, M., Thelen, S., Bignon, A., Levoye, A., Arenzana-Seisdedos, F., Balabanian, K., Bachelerie, F., Langen, H., and Thelen, M. (2012). Complementary methods provide evidence for the expression of CXCR7 on human B cells. Proteomics *12*, 1938–1948.

Jacquier, V., Prummer, M., Segura, J.-M., Pick, H., and Vogel, H. (2006). Visualizing odorant receptor trafficking in living cells down to the single-molecule level. Proc. Natl. Acad. Sci. U. S. A. *103*, 14325–14330.

Jiang, H., Wang, Y.-Y., Ran, X.-X., Fan, W.-M., Jiang, X.-H., Guan, W.-J., and Li, Y.-Q. (2013). Improvement of natamycin production by engineering of phosphopantetheinyl transferases in *Streptomyces chattanoogensis* L10. Appl. Environ. Microbiol. *79*, 3346–3354.

Joshi, A.K., Zhang, L., Rangan, V.S., and Smith, S. (2003). Cloning, expression, and characterization of a human 4 ′-phosphopantetheinyl transferase with broad substrate specificity. J. Biol. Chem. *278*, 33142–33149.

Jung, J., Bashiri, G., Johnston, J.M., Brown, A.S., Ackerley, D.F., and Baker, E.N. (2014). Crystal structure of the essential *Mycobacterium tuberculosis* phosphopantetheinyl transferase PptT, solved as a fusion protein with maltose binding protein. J. Struct. Biol. *188*, 274–278.

Kawamura, T., Stephens, B., Qin, L., Yin, X., Dores, M.R., Smith, T.H., Grimsey, N., Abagyan, R., Trejo, J., Kufareva, I., Fuster, M.M., Salanga, C.L., and Handel, T.M. (2014). A general method for site specific fluorescent labeling of recombinant chemokines. PLoS One *9*, e81454.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., and Barr, P.J. (1998). Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. Proc. Natl. Acad. Sci. U. S. A. *95*, 505–509.

Keller, M., and Zengler, K. (2004). Tapping into microbial diversity. Nat. Rev. Microbiol. *2*, 141–150.

Koglin, A., Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V.V., Blum, M.M., Mittag, T., Marahiel, M.A., Bernhard, F., and Dötsch, V. (2006). Conformational switches modulate protein interactions in peptide ant6ibiotic synthetases. Science *312*, 273–276.

Kosa, N.M., Haushalter, R.W., Smith, A.R., and Burkart, M.D. (2012). Reversible labeling of native and fusion-protein motifs. Nat. Methods *9*, 981–984.

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

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157*, 105–132.

La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. (2004). Manipulation of carrier proteins in antibiotic biosynthesis. Chem. Biol. *11*, 195–201.

Lai, J.R., Fischbach, M.A., Liu, D.R., and Walsh, C.T. (2006). A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. Proc. Natl. Acad. Sci. USA *103*, 5314–5319.

Lambalot, R.H., and Walsh, C.T. (1995). Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. J. Biol. Chem. *270*, 24658–24661.

Lambalot, R.H., and Walsh, C.T. (1997). Holo-[acyl-carrier-protein] synthase of *Escherichia coli*. Vitam. Coenzymes Pt I *279*, 254–262.

Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily - The phosphopantetheinyl transferases. Chem. Biol. *3*, 923–936.

Leblanc, C., Prudhomme, T., Tabouret, G., Ray, A., Burbaud, S., Cabantous, S., Mourey, L., Guilhot, C., and Chalut, C. (2012). 4'-phosphopantetheinyl transferase PptT, a new drug target required for *Mycobacterium tuberculosis* growth and persistence *in vivo*. PLoS Pathog. *8*.

Li, J.W.H., and Vederas, J.C. (2009). Drug discovery and natural products: end of an era or an endless frontier? Science *325*, 161–165.

Limmer, K., Pippig, D.A., Aschenbrenner, D., and Gaub, H.E. (2014). A Force-Based, Parallel Assay for the Quantification of Protein-DNA Interactions. PLoS ONE *9*, e89626.

Liu, Q., Ma, Y., Zhou, L.Y., and Zhang, Y.X. (2005). Gene cloning, expression and functional characterization of a phosphopantetheinyl transferase from *Vibrio anguillarum* serotypeO1. Arch. Microbiol. *183*, 37–44.

Lohman, J.R., Ma, M., Cuff, M.E., Bigelow, L., Bearden, J., Babnigg, G., Joachimiak, A., Phillips, G.N., and Shen, B. (2014). The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. Proteins Struct. Funct. Bioinforma. *82*, 1210–1218.

Maillard, R.A., Chistol, G., Sen, M., Righini, M., Tan, J., Kaiser, C.M., Hodges, C., Martin, A., and Bustamante, C. (2011). ClpX(P) generates mechanical force to unfold and translocate its protein substrates. Cell *145*, 459–469.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2673.

Marchetti, L., Callegari, A., Luin, S., Signore, G., Viegi, A., Beltram, F., and Cattaneo, A. (2013). Ligand signature in the membrane dynamics of single TrkA receptor molecules. J. Cell Sci. *126*, 4445–4456.

Marchetti, L., De Nadai, T., Bonsignore, F., Calvello, M., Signore, G., Viegi, A., Beltram, F., Luin, S., and Cattaneo, A. (2014). Site-specific labeling of neurotrophins and their receptors via short and versatile peptide tags. PLoS One *9*, e113708.

Martin, P., Marcq, I., Magistro, G., Penary, M., Garcie, C., Payros, D., Boury, M., Olier, M., Nougayrède, J.-P., Audebert, M., Chalut, C., Schubert, S., and Oswald, E. (2013). Interplay between siderophores and colibactin genotoxin biosynthetic pathways in *Escherichia coli*. PLoS Pathog. *9*, e1003437.

Martinez-Martin, N., Viejo-Borbolla, A., Martín, R., Blanco, S., Benovic, J.L., Thelen, M., and Alcamí, A. (2015). Herpes simplex virus enhances chemokine function through modulation of receptor trafficking and oligomerization. Nat. Commun. *6*.

Meier, J.L., and Burkart, M.D. (2009). The chemical biology of modular biosynthetic enzymes. Chem. Soc. Rev. *38*, 2012–2045.

Meier, J.L., Mercer, A.C., Rivera, H., and Burkart, M.D. (2006). Synthesis and evaluation of bioorthogonal pantetheine analogues for *in vivo* protein modification. J. Am. Chem. Soc. *128*, 12174–12184.

Meier, J.L., Niessen, S., Hoover, H.S., Foley, T.L., Cravatt, B.F., and Burkart, M.D. (2009). An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. ACS Chem. Biol. *4*, 948–957.

Mercer, A.C., and Burkart, M.D. (2007). The ubiquitous carrier protein - a window to metabolite biosynthesis. Nat. Prod. Rep. *24*, 750–773.

Mofid, M.R., Finking, R., and Marahiel, M.A. (2002). Recognition of hybrid peptidyl carrier proteins/Acyl carrier proteins in nonribosomal peptide synthetase modules by the 4 ’-phophopantetheinyl transferases AcpS and Sfp. J. Biol. Chem. *277*, 17023–17031.

Mofid, M.R., Finking, R., Essen, L.O., and Marahiel, M.A. (2004). Structure-based mutational analysis of the 4 ’-phosphopantetheinyl transferases Sfp from *Bacillus subtilis*: Carrier protein recognition and reaction mechanism. Biochemistry *43*, 4128–4136.

Mootz, H.D., Finking, R., and Marahiel, M.A. (2001). 4 ’-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. J. Biol. Chem. *276*, 37289–37298.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002a). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chembiochem *3*, 491–504.

Mootz, H.D., Schorgendorfer, K., and Marahiel, M.A. (2002b). Functional characterization of 4’-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of *Saccharomyces cerevisiae* lys5. FEMS Microbiol. Lett. *213*, 51–57.

Mosiewicz, K.A., Johnsson, K., and Lutolf, M.P. Phosphopantetheinyl transferase-catalyzed formation of bioactive hydrogels for tissue engineering. J Am Chem Soc *132*, 5972 – +.

Munro, J.B., Gorman, J., Ma, X., Zhou, Z., Arthos, J., Burton, D.R., Koff, W.C., Courter, J.R., Smith, A.B., Kwong, P.D., Blanchard, S.C., and Mothes, W. (2014a). Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. Science *346*, 759–763.

Munro, J.B., Nath, A., Färber, M., Datta, S.A.K., Rein, A., Rhoades, E., and Mothes, W. (2014b). A conformational transition observed in single HIV-1 Gag molecules during *in vitro* assembly of virus-like particles. J. Virol. *88*, 3577–3585.

Myers, J.A., Curtis, B.S., and Curtis, W.R. (2013). Improving accuracy of cell and chromophore concentration measurements using optical density. BMC Biophys. *6*, 4.

Nakano, M.M., Corbell, N., Besson, J., and Zuber, P. (1992). Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in *Bacillus subtilis*. Mol. Gen. Genet. *232*, 313–321.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat. Protoc. *2*, 2212–2221.

Oganesyan, N., Ankoudinova, I., Kim, S.-H., and Kim, R. (2007). Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. Protein Expr. Purif. *52*, 280–285.

Osbourn, A. (2010). Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. Trends Genet. *26*, 449–457.

Otten, M., Ott, W., Jobst, M.A., Milles, L.F., Verdorfer, T., Pippig, D.A., Nash, M.A., and Gaub, H.E. (2014). From genes to protein mechanics on a chip. Nat. Methods *11*, 1127–1130.

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

Owen, J.G., Copp, J.N., and Ackerley, D.F. (2011). Rapid and flexible biochemical assays for evaluating 4'-phosphopantetheinyl transferase activity. Biochem. J. *436*, 709–717.

Owen, J.G., Robins, K.J., Parachin, N.S., and Ackerley, D.F. (2012). A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. Environ. Microbiol. *14*, 1198–1209.

Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.-Y., Ofek, G., Stewart-Jones, G.B.E., Stuckey, J., Bailer, R.T., Joyce, M.G., Louder, M.K., Tumba, N., Yang, Y., Zhang, B., Cohen, M.S., Haynes, B.F., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature *514*, 455–461.

Parachin, N.S., and Gorwa-Grauslund, M.F. (2011). Isolation of xylose isomerases by sequence- and function-based screening from a soil metagenomic library. Biotechnol. Biofuels *4*, 9.

Parris, K.D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C.C., Seehra, J., and 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 Fold. Des. *8*, 883–895.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. Science *291*, 1790–1792.

Pfeifer, E., Pavela-Vrancic, M., von Doehren, H., and Kleinkauf, H. (1995). Characterization of tyrocidine synthetase 1 (TY1): Requirement of posttranslational modification for peptide biosynthesis. Biochemistry *34*, 7450–7459.

Puigbò, P., Guzmán, E., Romeu, A., and Garcia-Vallvé, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. *35*, W126–W131.

Puigbò, P., Romeu, A., and Garcia-Vallvé, S. (2008). HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. Nucleic Acids Res. *36*, D524–D527.

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

Rappé, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annu. Rev. Microbiol. *57*, 369–394.

Rashidian, M., Dozier, J.K., and Distefano, M.D. (2013). Enzymatic labeling of proteins: techniques and approaches. Bioconjug. Chem. *24*, 1277–1294.

Reuter, K., Mofid, M.R., Marahiel, M.A., and Ficner, R. (1999). Crystal structure of the surfactin synthetase-activating enzyme Sfp: a prototype of the 4 ’-phosphopantetheinyl transferase superfamily. Embo. J. *18*, 6823–6831.

Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. (2002). Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. J. Bacteriol. *184*, 654–665.

Roberts, G.A., Staunton, J., and Leadlay, P.F. (1993). Heterologous expression in *Escherichia coli* of an intact multienzyme component of the erythromycin-producing polyketide synthase. Eur. J. Biochem. *214*, 305–311.

Roche, E.D., and Walsh, C.T. (2003). Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. Biochemistry *42*, 1334–1344.

Rottier, K., Faille, A., Prudhomme, T., Leblanc, C., Chalut, C., Cabantous, S., Guilhot, C., Mourey, L., and Pedelacq, J.-D. (2013). Detection of soluble co-factor dependent protein expression in vivo: Application to the 4’-phosphopantetheinyl transferase PptT from Mycobacterium tuberculosis. J. Struct. Biol. *183*, 320–328.

Sanchez, C., Du, L.C., Edwards, D.J., Toney, M.D., and Shen, B. (2001). Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces verticillus* ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem. Biol. *8*, 725–738.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F., and Hutchinson, C.R. (2005). Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. Appl. Environ. Microbiol. *71*, 4840–4849.

Schoeler, C., Malinowska, K.H., Bernardi, R.C., Milles, L.F., Jobst, M.A., Durner, E., Ott, W., Fried, D.B., Bayer, E.A., Schulten, K., Gaub, H.E., and Nash, M.A. (2014). Ultrastable cellulosome-adhesion complex tightens under load. Nat. Commun. *5*.

Schwarzer, D., and Marahiel, M.A. (2001). Multimodular biocatalysts for natural product assembly. Naturwissenschaften *88*, 93–101.

Schwarzer, D., Finking, R., and Marahiel, M.A. (2003). Nonribosomal peptides: from genes to products. Nat. Prod. Rep. *20*, 275.

Shen, B., Summers, R.G., Gramajo, H., Bibb, M.J., and Hutchinson, C.R. (1992). Purification and characterization of the acyl carrier protein of the *Streptomyces glaucescens* tetracenomycin C polyketide synthase. J. Bacteriol. *174*, 3818–3821.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. Chem. Rev. *105*, 715–738.

Sieber, S.A., Walsh, C.T., and Marahiel, M.A. (2003). Loading peptidyl-coenzyme A onto peptidyl carrier proteins: A novel approach in characterizing macrocyclization by thioesterase domains. J. Am. Chem. Soc. *125*, 10862–10866.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Sunbul, M., and Yin, J. (2009). Site specific protein labeling by enzymatic posttranslational modification. Org. Biomol. Chem. *7*, 3361–3371.

Sunbul, M., Yen, M., Zou, Y.K., and Yin, J. (2008). Enzyme catalyzed site-specific protein labeling and cell imaging with quantum dots. Chem. Commun. 5927–5929.

Sunbul, M., Marshall, N.J., Zou, Y.K., Zhang, K.Y., and Yin, J. (2009). Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. J. Mol. Biol. *387*, 883–898.

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

Torsvik, V., Øvreås, L., and Thingstad, T.F. (2002). Prokaryotic diversity--magnitude, dynamics, and controlling factors. Science *296*, 1064–1066.

Troeschel, S.C., Drepper, T., Leggewie, C., Streit, W.R., and Jaeger, K.E. (2010). Novel tools for the functional expression of metagenomic DNA. Methods Mol. Biol. *668*, 117–139.

Tufar, P., Rahighi, S., Kraas, F.I., Kirchner, D.K., Löhr, F., Henrich, E., Köpke, J., Dikic, I., Güntert, P., Marahiel, M.A., and Dötsch, V. (2014). Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. Chem. Biol. *21*, 552–562.

Valley, C.C., Arndt-Jovin, D.J., Karedla, N., Steinkamp, M.P., Chizhik, A.I., Hlavacek, W.S., Wilson, B.S., Lidke, K.A., and Lidke, D.S. (2015). Enhanced dimerization drives ligand-independent activity of mutant EGFR in lung cancer. Mol. Biol. Cell mbc.E15–05 – 0269.

Van Lanen, S.G., Dorrestein, P.C., Christenson, S.D., Liu, W., Ju, J., Kelleher, N.L., and Shen, B. (2005). Biosynthesis of the beta-amino acid moiety of the enediyne antitumor antibiotic C-1027 featuring beta-amino acyl-S-carrier protein intermediates. J. Am. Chem. Soc. *127*, 11594–11595.

Vasquez, J.K., Chantranuvatana, K., Giardina, D.T., Coffman, M.D., and Knight, J.D. (2014). Lateral diffusion of proteins on supported lipid bilayers: additive friction of synaptotagmin 7 C2A–C2B tandem domains. Biochemistry *53*, 7904–7913.

Vickery, C.R., Kosa, N.M., Casavant, E.P., Duan, S., Noel, J.P., and Burkart, M.D. (2014). Structure, biochemistry, and inhibition of essential 4'-phosphopantetheinyl transferases from two species of *Mycobacteria*. ACS Chem. Biol. *9*, 1939–1944.

Vivero-Pol, L., George, N., Krumm, H., Johnsson, K., and Johnsson, N. (2005). Multicolor imaging of cell surface proteins. J. Am. Chem. Soc. *127*, 12770–12771.

Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E.N., and Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. Curr. Opin. Chem. Biol. *1*, 309–315.

Wang, Y.Y., Li, Y.D., Liu, J.B., Ran, X.X., Guo, Y.Y., Ren, N.N., Chen, X., Jiang, H., and Li, Y.-Q. (2014). Characterization and evolutionary implications of the triad Asp-Xxx-Glu in group II phosphopantetheinyl transferases. PLoS One *9*, e103031.

Weber, G., Schörgendorfer, K., Schneider-Scherzer, E., and Leitner, E. (1994). The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. Curr. Genet. *26*, 120–125.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure *8*, 407–418.

Wilmes, S., Beutel, O., Li, Z., Francois-Newton, V., Richter, C.P., Janning, D., Kroll, C., Hanhart, P., Hötte, K., You, C., Uzé, G., Pellegrini, S., and Piehler, J. (2015). Receptor dimerization dynamics as a regulatory valve for plasticity of type I interferon signaling. J. Cell Biol. *209*, 579–593.

Wong, L.S., Thirlway, J., and Micklefield, J. (2008). Direct site-selective covalent protein immobilization catalyzed by a phosphopantetheinyl transferase. J. Am. Chem. Soc. *130*, 12456–12464.

Yasgar, A., Foley, T.L., Jadhav, A., Inglese, J., Burkart, M.D., and Simeonov, A. (2010). A strategy to discover inhibitors of *Bacillus subtilis* surfactin-type phosphopantetheinyl transferase. Mol. Biosyst. *6*, 365–375.

Yin, J., Liu, F., Li, X.H., and Walsh, C.T. (2004). Labeling proteins with small molecules by site-specific posttranslational modification. J. Am. Chem. Soc. *126*, 7754–7755.

Yin, J., Straight, P.D., McLoughlin, S.M., Zhou, Z., Lin, A.J., Golan, D.E., Kelleher, N.L., Kolter, R., and Walsh, C.T. (2005). Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. Proc. Natl. Acad. Sci. U.S.A. *102*, 15815–15820.

Yin, J., Straight, P.D., Hrvatin, S., Dorrestein, P.C., Bumpus, S.B., Jao, C., Kelleher, N.L., Kolter, R., and Walsh, C.T. (2007). Genorne-wide high-throughput mining of natural-product biosynthetic gene clusters by phage display. Chem Biol *14*, 303–312.

Zhang, K., He, J., Yang, M., Yen, M., and Yin, J. (2009). Identifying natural product biosynthetic genes from a soil metagenome by using T7 phage selection. ChemBioChem *10*, 2599–2606.

Zhou, Z., Lai, J.R., and Walsh, C.T. (2006). Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. Chem. Biol. *13*, 869–879.

Zhou, Z., Cironi, P., Lin, A.J., Xu, Y.Q., Hrvatin, S., Golan, D.E., Silver, P.A., Walsh, C.T., and Yin, J. (2007a). Genetically encoded short peptide tags for orthogonal protein labeling by sfp and AcpS phosphopantetheinyl transferases. ACS Chem. Biol. *2*, 337–346.

Zhou, Z., Lai, J.R., and Walsh, C.T. (2007b). Directed evolution of aryl carrier proteins in the enterobactin synthetase. Proc. Natl. Acad. Sci. U. S. A. *104*, 11621–11626.

Zhou, Z., Koglin, A., Wang, Y., McMahon, A.P., and 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*. J. Am. Chem. Soc. *130*, 9925–9930.

Zhu, P., Zheng, Y., You, Y., Yan, X., and Shao, J. (2009). Sequencing and modular analysis of the hybrid non-ribosomal peptide synthase – polyketide synthase gene cluster from the marine sponge *Hymeniacidon perleve* -associated bacterium *Pseudoalteromonas* sp. strain NJ631. Can. J. Microbiol. *55*, 219–227.

Zimhony, O., Schwarz, A., Raitses-Gurevich, M., Peleg, Y., Dym, O., Albeck, S., Burstein, Y., and Shakked, Z. (2015). AcpM, the meromycolate extension acyl carrier protein of *Mycobacterium tuberculosis*, is activated by the 4'-phosphopantetheinyl transferase PptT, a potential target of the multistep mycolic acid biosynthesis. Biochemistry *54*, 2360–2371.

Ziomkiewicz, I., Loman, A., Klement, R., Fritsch, C., Klymchenko, A.S., Bunt, G., Jovin, T.M., and Arndt-Jovin, D.J. (2013). Dynamic conformational transitions of the EGF receptor in living mammalian cells determined by FRET and fluorescence lifetime imaging microscopy. Cytom. Part J. Int. Soc. Anal. Cytol. *83*, 794–805.