

CHARACTERISATION, MANIPULATION AND DIRECTED EVOLUTION OF
NON-RIBOSOMAL PEPTIDE SYNTHETASE ENZYMES

By

Jeremy George Owen

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

2010

Abstract

Non-ribosomal peptide synthetases (NRPS) are large, modular enzymes that synthesise biologically active secondary metabolites from amino acid precursors without the need for a nucleic acid template. NRPS play an integral role in microbial physiology and also have potential applications in the synthesis of novel peptide molecules. Both of these aspects are examined in this thesis.

Under conditions of iron starvation *Pseudomonas syringae* synthesises siderophores for active uptake of iron. The primary siderophore of *P. syringae* is pyoverdine, a fluorescent molecule that is assembled from amino acid (aa) precursors by NRPS. Five putative pyoverdine NRPS genes in *P. syringae* pv. *phaseolicola* 1448a (Ps1448a) were identified and characterised *in silico* and their role in pyoverdine biosynthesis was confirmed by gene knockout. Creation of pyoverdine null Ps1448a enabled identification of a previously uncharacterised temperature-regulated secondary siderophore, achromobactin, which is NRPS independent and has lower affinity for iron. Pyoverdine and achromobactin null mutants were characterised in regard to iron uptake, virulence and growth in iron-limited conditions. Determination of the substrate specificity for the seven adenylation (A) domains of the Ps1448a pyoverdine sidechain NRPS was also attempted. Although ultimately unsuccessful, these attempts provided a rigorous assessment of methods for the expression, purification and biochemical characterisation of A-domains.

The Ps1448a NRPS were subsequently employed in domain swapping experiments to test condensation (C) domain specificity for aa substrates during peptide formation *in vivo*. Experiments in which the terminal C- and/or A-domain of the *Pseudomonas aeruginosa* (PAO1) pyoverdine NRPS system were replaced with alternative domains from Ps1448a and PAO1 were consistent with previous *in vitro* observations that C-domains exhibit strong sidechain and stereo-selectivity at the downstream aa position, but only stereo-selectivity at the upstream aa position.

These results prompted investigation into the role of inter-domain communication in NRPS function, to test the hypothesis that the thiolation (T) domain enters into specific interactions with other domains, which might provide an alternative explanation for the diminished activity of recombinant NRPS enzymes. A recently characterised single-module NRPS, *bpsA*, was chosen as a reporter gene for these experiments based on its ability to generate blue pigment in *Escherichia coli*. Substitution of the native *bpsA* T-domain consistently impaired function, consistent with the hypothesis. It was shown that directed evolution could be applied to restore function in substituted T-domains. Mutations that restored function were mapped *in silico*, and a structural model for interaction between the thioester (TE) and T-domain of BpsA was derived.

The utility of *bpsA* for discovery and characterisation of phosphopantetheinyl transferase (PPTase) enzymes was also investigated. *In vivo* and *in vitro* assays for determination of PPTase activity were developed and a high-throughput screen for discovery of new PPTases in environmental DNA libraries was successfully implemented.

Acknowledgements

Dr David Ackerley, my supervisor. Dave, I have always been able to count on your support. Not once in the four years I have known have you let me down. You have inspired me to be creative and taught me to be scientific. Furthermore you are cool. Yes, for those of you who didn't know Dave is a super guy.

The rest of the Ackerley gang, thank you all for brightening up the lab with your excellent selves. Janine, for sharing your lab wisdom with an inept newbie, thanks to you I can now use a pipette and various other nifty lab doo-dackies. Mark, thanks for doing the experiments in chapter 5 in spite of my efforts to thwart your progress with unidentified restriction sites (we can all have a good laugh about it now). My buddies, Sandi, Sha, Chris, Ian and Gareth – thanks for being fun to hang out with in and out of the lab. Thanks Ronan and the O'Toole-labites for sharing the microbiology lab with the ever expanding Ackerley lab.

Thanks to all the friendly and approachable staff members and students who I have pestered for advice and equipment loans. Cam, Craig and Neville – thanks for un-breaking everything, the department would fall apart without you. Thank you Paul T-S for doing such a great job of heading up our fine biology department. To all the people on the fifth floor of Kirk building who have helped me organise my often chaotic life, thanks for telling me things that I should already know, and being so nice about it.

Anna, my green-eyed tigress. Tolerance and grace. I have been creeping toward the finish line in ever decreasing increments for quite some time now. You have been amazing. Also, without your editorial skills I would asses a great many things according to the manufactures directions.

To my family. Mum, Dad, Simon and Olivia. I have been at university getting an education for a fair amount of time now. All the important things, I have learned from you.

Table of Contents

Chapter 1 – Introduction.....	1
1.1 – Non-Ribosomal Peptide Synthetases	1
1.2 – An overview of core NRPS domains	3
1.3 – Adenylation domain structure and catalytic mechanism	5
1.3.1 – A-domain specificity and the non-ribosomal code	9
1.4 – The Thiolation domain.....	10
1.5 – The Condensation domain	14
1.5.1 – C-domain specificity	17
1.5.2 – C-domain evolution and specialisation	17
1.5.3 – C-domain structure.....	18
1.6 – The Thioesterase Domain	19
1.6.1 – Mechanism of product release by TE-domains	20
1.6.2 – TE-domain structure	21
1.6.3 – Type II TE-domains	22
1.7 – Phosphopantetheinyl transferases	22
1.8 – NRPS tailoring domains	23
1.9 – Manipulation of NRPS.....	25
1.9.1 – Module swaps and site directed mutagenesis	26
1.9.2 – Combinatorial biosynthesis using the daptomycin synthetase system	28
1.9.3 Development of an NRPS tool box.....	31
1.10 – Aims of this study	32
Chapter 2 – Material and methods.....	34
2.1 – General reagents and materials	34
2.2 – Enzymes	34
2.3 – Bacterial strains and plasmids.....	34
2.3.1 – Bacterial strains.....	34
2.3.2 – Plasmids	35
2.4 – Oligonucleotide primers.....	36
2.5 – Media	39
2.5.1 – Media supplements	40
2.6 – Growth and maintenance of bacteria	41
2.7 – Routine molecular biology.....	41

2.7.1 – PCR protocols	41
2.7.1 – Isolation purification and manipulation of DNA	44
2.7.2 – Preparation and transformation of competent cells	44
2.7.3 – In frame deletion of genes from the chromosome of Ps1448a	48
2.8 – Siderophore characterisation.....	49
2.8.1 – CAS agar assays for iron uptake.....	49
2.8.2 – CAS media assays for iron uptake	50
2.8.2 – EDDHA Inhibitory Concentration (IC ₅₀) assays	50
2.8.3 – Ps1448a pathogenicity tests in <i>Phaseolus vulgaris</i>	50
2.8.4. – Achromobactin purification.....	51
2.8.5 – Pyoverdine purification.....	52
2.8.6 – Assessment of pyoverdine production in wild type and recombinant <i>P. aeruginosa</i> strains.....	52
2.8.7 – Determination of intracellular pyoverdine levels by flow cytometry	53
2.9 – Protein expression and purification	53
2.9.1 – Protein expression.....	53
2.9.2 – Protein purification by Ni-NTA affinity chromatography.....	55
2.9.3 – SDS-polyacrylamide gel electrophoresis.....	57
2.10 – A-domain specificity assays	58
2.10.1 – ATP/ ³² PPi exchange assays	58
2.10.2 – PPi release assay RA1.....	59
2.10.3 – PPi release assay RA2.....	60
2.11 – Directed evolution Protocols.....	60
2.11.1 – Library generation.....	61
2.11.2 – First tier screening.....	64
2.11.3 – Second tier screening	64
2.12 – Indigoidine quantification	65
2.13 – Enzyme kinetics	65
2.13.1 – Activation of BpsA by 4'-PP attachment.....	65
2.13.2 – Determination of kinetic parameters for BpsA.....	65
2.13.3 – Determination of kinetic parameters for PPTases	66
2.14 – Screening of soil derived eDNA library for PPTase genes.....	66
2.15 –In silico manipulation using Vector NTI and DeepView	67
2.15.1 – Design of scheme three for seamless T-domain substitution into BpsA	67
2.15.2 – Structural modelling.....	68
Chapter 3 – Siderophores in <i>Pseudomonas syringae</i>: Synthesis and physiological significance.....	69
3.1 – Summary	69
3.2 – Introduction.....	70
3.2.1 – Pyoverdine synthesis in fluorescent <i>pseudomonads</i>	70

3.2.2 – Regulation of pyoverdine synthesis	71
3.2.3 – NRPS independent siderophores.....	72
3.3 – Identification and in silico characterisation of the pyoverdine NRPS genes.....	73
3.4 – Confirmation of pyoverdine NRPS identity by gene knockout.....	77
3.4.1 – Construction of pyoverdine NRPS mutant strains	77
3.4.2 – Initial phenotypic characterisation of NRPS knockout strains	79
3.4.3 – Complementation of a pyoverdine NRPS deletion <i>in trans</i>	80
3.5 – Identification and genetic characterisation of achromobactin as a secondary siderophore of Ps1448a.....	81
3.5.1 – <i>In silico</i> characterisation of the Ps1448a achromobactin biosynthetic locus	82
3.5.2 – Generation and characterisation of achromobactin deficient Ps1448a strains	84
3.6 – Assessment of relative fitness of mutant stains under iron starvation conditions	86
3.7 – Assessment of pathogenicity of mutant strains in <i>Phaseolus vulgaris</i>	87
3.8 – Mass spectrometric analysis of pyoverdine and achromobactin purified from Ps1448a	88
3.9 – Discussion	92
Chapter 4 – Techniques for purification and characterisation of NRPS A-domains.....	96
4.1 Introduction.....	96
4.2 General strategy for determination of substrate specificity	97
4.2.1 Design and construction of AT bi-domain expression vectors.....	99
4.2.2 – Preliminary expression testing.....	100
4.2.3 – Optimisation of expression conditions to achieve soluble recombinant proteins	100
4.2.4 – Assessment of substrate specificity by ATP/ ³² PPi exchange assay.....	103
4.3 – Exploring alternative protein expression strategies	110
4.3.1- <i>Pseudomonas</i> species as an expression host.....	110
4.3.2 – Purification and assessment of proteins produced in <i>E. coli</i> co-expressing molecular chaperone proteins	118
4.3.3- Refolding of a denatured AT bi-domain isolated from inclusion bodies	120
4.3.4 – Exploration of alternative domain boundaries and constructs for substrate specificity assessment	122
4.4 – Discussion	126
Chapter 5 – Probing C-domain specificity using the pyoverdine biosynthetic NRPS of <i>Pseudomonas aeruginosa</i> as a model system.....	131
5.1 – Introduction.....	131
5.1.1 – Pyoverdine synthesis as a model system for module swapping experiments	131
5.1.2 – A description of previous module swapping experiments conducted using the <i>P. aeruginosa</i> pyoverdine biosynthetic system.....	132

5.1.3 – Aims and overview	133
5.2 – Design and construction of recombinant <i>pvdD</i> genes	134
5.2.1 – Nomenclature for describing modified genes and strains created in this study	136
5.3 – Assessment of function of recombinant genes.....	137
5.3.1 – Pyoverdine production on solid media	137
5.3.2 – Pyoverdine production in liquid media.....	139
5.3.3 – Assessment of pyoverdine function for strain harbouring modified <i>pvdD</i> genes	141
5.3.4 – Analysis of intracellular pyoverdine levels in modified strains using flow cytometry	141
5.4 – Discussion	143
Chapter 6 – Probing T-domain interactions using BpsA as a reporter system	148
6.1 – Introduction.....	148
6.1.1 – T-domain structure and function.....	151
6.1.2 – Aims	152
6.2 – Examination of T-domain interactions in silico.....	152
6.2.1 – Alignment based analysis of T-domain interactions.....	152
6.2.2 – Structural insights into T-TE interaction	153
6.3 – Cloning of BpsA and determination of compatibility with <i>Pseudomonas</i> PPTases..	156
6.3.1 – Construction of plasmids for expression of BpsA in <i>E. coli</i> and <i>Pseudomonas</i> species	156
6.3.2 – Assessment of BpsA activity in <i>Pseudomonas</i> species	157
6.4 – Development of plasmid based platforms for the co-expression of BpsA and an activating PPTase in <i>E. coli</i>	158
6.5 – Development and testing of T-domain swapping platforms.....	159
6.5.1 – Substitution scheme one	159
6.5.2 – Substitution scheme two	160
6.5.3 – Substitution scheme three	160
6.5.4 – System for description of modified BpsA variants.....	161
6.6 – Assessment of modified BpsA variants created using swapping schemes 1-3.....	163
6.7 – Development of a high throughput screening process for directed evolution of recombinant <i>bpsA</i> genes	165
6.7.1 – Development of a first tier agar plate based screening system.....	165
6.7.2 – Optimisation of library generation for directed evolution	166
6.8 – Adaptation of a T-domain from DhbF to function in BpsA using directed evolution	167
6.9 – Adaptation of a T-domain from PvdD to function in BpsA	169
6.9.2 – Recovery and assessment of hits.....	169
6.9.3 – Second round evolution	171
6.10 – Adaptation of the T-domain of EntF to function in BpsA	172

6.11 – Summary of sequence data from directed evolution experiments	174
6.12 – Discussion	176
6.12.1 – Interpretation of results from T-domain swapping experiments.....	176
6.12.2 –Structural analysis of type I mutations from first round evolution experiments	177
6.12.3 – Interpretation of type II mutations from first round evolution experiments.....	180
6.12.4 –Interpretation of results from second round evolution experiment.....	181
6.12.5 – A proposed model for T-domain interaction with the TE-domain of BpsA.....	182
6.12.6 – Concluding comments on T-domain function and interactions.....	185
Chapter 7 – BpsA as a tool for the characterisation and discovery of phosphopantetheinyl transferase enzymes	186
7.1 – Introduction.....	186
7.1.2 – Aims and overview	186
7.2 – In vivo characterisation of PPTase enzymes using BpsA.....	187
7.3 – In vitro activity of BpsA	188
7.3.1 – Purification of his-tagged BpsA and PPTases	188
7.3.2 – Initial <i>in vitro</i> characterisation of BpsA.....	189
7.3.3 – Kinetic studies.....	191
7.4 – Determination of kinetic parameters for PPTases using a BpsA reporter system	194
7.5 – BpsA as a tool for discovery of PPTase genes in environmental DNA libraries	197
7.5.1 – Screening of an eDNA library and recovery of hits	198
7.5.2 – Sequence analysis of hits from eDNA library screen	199
7.6 – Discussion	200
7.6.1 – Interpretation of BpsA kinetic data.....	200
7.6.2 – Interpretation of PPTase kinetic data and assay validation	201
7.6.2 – Utility of BpsA for recovery of novel PPTases and secondary metabolite clusters.....	201
Chapter 8 – Conclusions.....	203
8.1- Research motivation	203
8.2 – Key findings.....	204
8.2.1 – Biosynthesis and physiological significance of siderophores in <i>P. syringae</i>	204
8.2.3 – Manipulation of pyoverdine biosynthetic NRPS in <i>P. aeruginosa</i>	205
8.2.4 – Substitution and directed evolution of T-domains in BpsA.....	206
8.2.5 – Discovery and characterisation of PPTases	207
8.3 – Research applications.....	208
8.4 – Future directions	210
8.4.1 – Siderophore characterisation.....	210
8.4.2 – T-domain function and interaction.....	211
8.4.3 – Discovery and characterisation of PPTases	212

8.5 – Concluding remarks	213
Appendix One – Sequence alignments	214
A1.1 – Phylogenetic analysis of T-domains	214
A1.1.1 – Examination of a module duplication event in PvdD	218
A1.2 – Complete alignments for all improved clones	219
Appendix 2 – Unit conversion for PPTase kinetic parameters.....	222
Bibliography	224

List of Tables

Table 2.1 – E.coli strains used in this study.....	34
Table 2.2 – <i>P. aeruginosa</i> strains used in this study.....	35
Table 2.3 – <i>P. syringae</i> strains used in this study.....	35
Table 2.4- Plasmids used in this study.....	35
Table 2.5 – Primers used in this study.....	37
Table 2.6 – Antibiotics used in this study.....	41
Table 2.7 – Chaperone strains 1-4[106].....	54
Table 3.1 – Summary of PAO1 and Ps1448a pyoverdine gene alignment results	74
Table 3.2 – <i>In silico</i> prediction of A-domain specificity for Ps1448a pyoverdine.....	77
sidechain NRPS	77
Table 3.3 – Summary of <i>D. dadantii</i> and Ps1448a achromobactin gene alignment results	84
Table 3.4 – Diagnostic ions arising from MS/MS analysis of the m/z = 1141 pyoverdine species.....	92
Table 4.1 – Summary of purification procedures and assay results	126
Table 6.1 – system for describing modified BpsA genes/proteins	161
Table 6.2 – In vivo pigment synthesis capacity of recombinant BpsA genes.	164
Table 6.3 – Type I and II mutations recovered from first round directed evolution experiments	175
Table 6.4 – Type I and II mutations recovered from second round evolution of 3kF0.....	175
Table 7.1 – Kinetic parameters* for BpsA	193
Table 7.2 – Kinetic parameters for PcpS and PP1183.....	197
Table 7.3 – partial and complete genes found in eDNA fragments.....	199
Table A1.1 – Summary of alignment features shown in Figure A1.2.....	218

List of Figures

Figure 1.1 – Overview of NRPS function.....	2
Figure 1.2 –Function of core domains	4
Figure 1.3 – Adenylation domain structure	6
Figure 1.4 – The A-domain reaction cycle	8
Figure 1.5 – Post translational activation of T-domains	11
Figure 1.6 – Thiolation domain interactions during synthesis.....	12
Figure 1.7 – T-domain conformational switching	13
Figure 1.8 – C-domain function.....	14
Figure 1.9 – Two proposed mechanisms for peptide bond formation in NRPS:.....	16
Figure 1.10 – The C-domain of TycC6.....	19
Figure 1.11 – Product release by TE-domains.....	21
Figure 1.12 – N-methylation of amino acids by MT-domains in an NRPS module	24
Figure 1.13 – Formation of thiazoline and oxazoline groups by Cy domains.....	24
Figure 1.14 – Module swapping in the surfactin biosynthesis system	27
Figure 1.15 – Subunit exchange to produce novel daptomycin derivatives	29
Figure 1.16 – Module swapping to produce novel daptomycin derivatives in <i>Streptomyces roseosporus</i>	30
Figure 2.1 – Molecular weight markers used in this study	57
Figure 2.2 – Substitution scheme three cloning strategy	68
Figure 3.1 – Regulation of pyoverdine synthesis and transport in <i>P. aeruginosa</i>	72
Figure 3.2 – A Comparison of the pyoverdine loci of <i>P. aeruginosa</i> PAO1 and Ps1448a75	
Figure 3.3 – Strategy for in-frame deletion of siderophore biosynthetic genes.....	78
Figure 3.4 – Confirmation of <i>pspph1926</i> deletion by PCR and Southern blot.....	79
Figure 3.5 – Characterisation of Ps1448a pyoverdine NRPS knockouts.....	80
Figure 3.6 – Temperature-dependent production of a secondary siderophore by pyoverdine null Ps1448a.....	82
Figure 3.7 – Comparison of the achromobactin gene loci of <i>D. dadantii</i> and Ps1448a ...	83
Figure 3.8 – Liquid CAS assay	85

Figure 3.9 – Growth of WT and siderophore deficient strains during iron starvation.....	87
Figure 3.10 – Assessment of WT and mutant strains using standard pathogenicity test..	88
Figure 3.11 – MS analysis of pyoverdine purified from Ps1448a	89
Figure 3.12 – Ions arising from MS/MS analysis of pyoverdines isolated from Ps1448a	91
Figure 4.1 – pET28a+ map and features	98
Figure 4.2 – Construction of AT bi-domain expression vectors.....	99
Figure 4.4 – Ni-NTA purification of AT bi-domains	103
4.6 – Validation of exchange assay EA2 using EntF A-domain positive control.....	107
Figure 4.8 – Effect of ATP/Mg ²⁺ treatment on ATP/ ³² PPi exchange activity of 252.....	109
Figure 4.9 – Expression tests for pSX:251 in <i>P. syringae</i> , <i>P. putida</i> , <i>P. aeruginosa</i> and <i>E. coli</i>	112
Figure 4.11 – Validation of absorbance-based PPi release assay RA1 for determination of A- domain substrate specificity.....	115
Figure 4.12 – Monitoring continuous release of PPi by EntF A-domain	116
Figure 4.13 – Validation of A-domain specificity assay RA2.....	118
Figure 4.14 – Solubility trial using Chaperone strains 1-4	119
Figure 4.15 – Apparatus for on column refolding of denatured proteins	121
Figure 4.16 – SDS Page analysis of soluble 252 recovered from denatured inclusion bodies	122
Figure 4.17 – Assessment of L- lysine concentration on PPi release activity of x251...	125
Figure 4.18 – The relationship between energy and conformation of proteins in the cytoplasm	129
Figure 5.1 – Summary of previous domain swapping strategies using <i>pvdD</i>	133
Figure 5.2 – Outline of pSMC staging vector.....	135
Figure 5.3 – Recombinant genes <i>pvdD</i> created in this study	136
Figure 5.4 – Assessment of pyoverdine production strains on KB agar.....	138
Figure 5.5 – Quantitative assessment of pyoverdine production in KB broth.....	140
Figure 5.6 – Effect of arabinose concentration on pyoverdine production.....	140
Figure 5.7 Growth of mutant strains under iron starvation.....	141
Figure 5.8 – Determination of intracellular pyoverdine levels by flow cytometry	142
Figure 5.10 – Interaction of substituted CA domains with T-domains in PvdD	145

Figure 6.1 – BpsA	149
Figure 6.3 – Comparison of T-domain H and TE-states.....	154
Figure 6.4 – Hydrophobic interface of the T and TE-domains of EntF.....	155
Figure 6.5 – Expression of BpsA in <i>P. putida</i>	158
Figure 6.6 – Splice points for T-domain substitution schemes 1-3.	162
Figure 6.8 – Overview of directed evolution procedure	168
Figure 6.9 – Quantitative assessment of activity of improved variants of oDhbF	168
Figure 6.10A – Quantitative assessment of activity of improved variants of sIPvdD....	170
Figure 6.10B – Quantitative assessment of activity of improved variants of oPvdD.....	171
Figure 6.11 – Quantitative assessment of activity of improved variants of 3KFO.....	172
Figure 6.12 – Quantitative assessment of activity of improved variants of sEntf	173
Figure 6.15 – Locations of type II mutations mapped onto a structural model of the BpsA T- domain.....	181
Figure 6.17 – Schematic representation of the hydrophobic interface between T and TE- domains	184
Figure 7.2 – Preliminary analysis of indigoidine formation catalysed by BpsA.....	190
Figure 7.3 – The effect of pH on Indigoidine stability and BpsA reaction rate	191
Figure 7.4 – Determination of pigment synthesis velocity	192
Figure 7.5 – Derivation of kinetic parameters for BpsA	193
Figure 7.6 – Derivation of kinetic parameters for PPTase enzymes using the BpsA reporter assay.....	194
Figure 7.7 – Raw and processed data from PPTase comparison experiment	195
Figure 7.8 – Derivation of kinetic parameters for PcpS and PP1183	196
Figure 7.9 – Recovery of PPTases from an eDNA library	198
Figure A1.1 – Comparison of sequence diversity for T-domains and ACPs from type II FAS	215
Figure A1.2 – Interaction specific features of T-domain sequences	217
Figure A1.3 – T-domain sequence divergence following a putative module duplication event in PvdD	219

Chapter 1 – Introduction

1.1 – Non-Ribosomal Peptide Synthetases

Non ribosomal peptide synthetases (NRPS) are enzymes found in many bacteria and fungi that catalyse the production of biologically active small peptides from aa precursors without the need for a nucleic acid template [1-3]. In non ribosomal peptide synthesis, the NRPS proteins themselves form the template that directs the number, order and identity of aa substrates found in the peptide product [4]. NRPS proteins consist of a series of discrete modules, each responsible for the recognition, activation and incorporation of a single aa substrate [5]. The first module in a NRPS system is known as the initiation module and the final module known as the termination module, intermediate to these are the elongation modules [1]. Each module contains a site for attachment of a 4'-phosphopantetheine (4'-PP) cofactor that serves to tether aa substrates and peptide intermediates via a thioester bond [6]. As synthesis proceeds, the growing peptide chain is passed from one 4'-PP cofactor to the next with the formation of a peptide bond displacing the thioester linkage. The termination module possesses thioesterase activity and cleaves the final product from its 4'-PP cofactor. In general the number of modules in a NRPS system corresponds to the number of aas in the molecule produced; furthermore the order in which these modules are encoded in the genome of the producer organism reflects the order in which aa substrates are incorporated into the peptide product. This phenomenon is known as the co-linearity rule [4, 7]. Modules within bacterial NRPS biosynthetic system are usually spread out over a number of interacting proteins; each protein typically contains 1-5 modules however up to 18 modules have been known to comprise a single NRPS protein [8]. Figure 1 outlines the modular structure and basic function of NRPS.

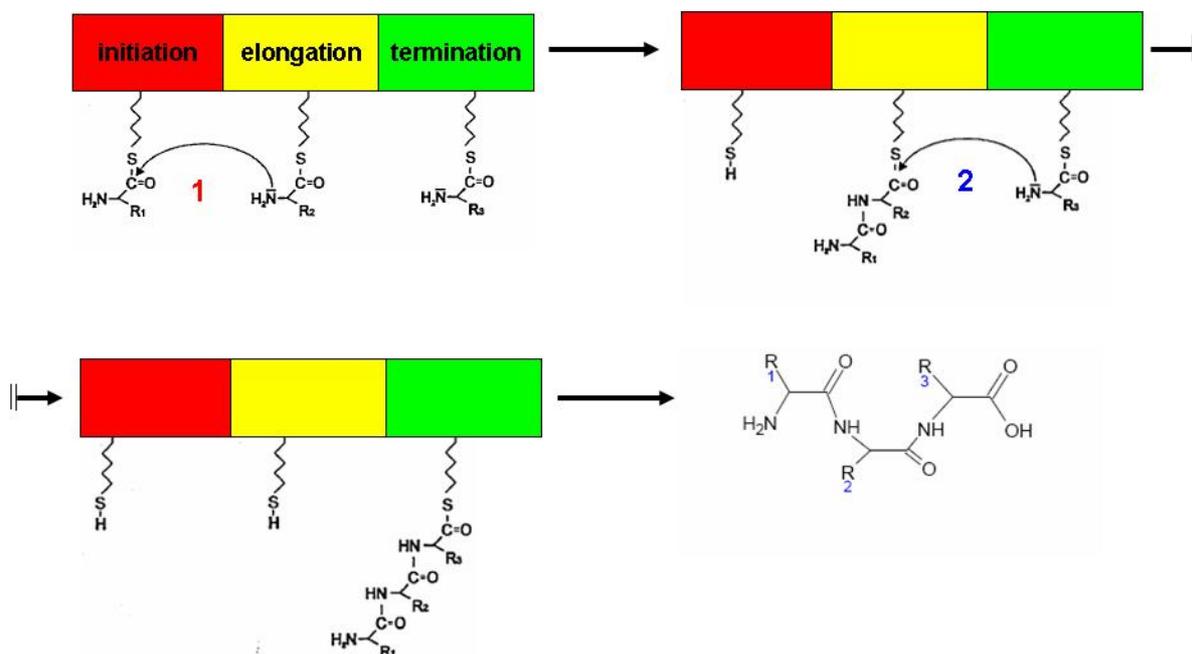


Figure 1.1 – Overview of NRPS function

A hypothetical tri-modular NRPS enzyme consisting of a single initiation, elongation and termination module. Each module is responsible for selecting a single aa substrate which is then tethered to the enzyme via a 4'-PP cofactor. Formation of peptide bond **1** results in a di-peptide linked to the elongation module, peptide bond **2** results in formation of a tri-peptide linked to the termination module. Thioesterase activity of the termination module cleaves the tri-peptide product from the enzyme.

NRPS enzymes are particularly interesting from a biotechnology perspective due to the immense diversity in structure and function of the molecules they synthesise. Unlike ribosomal peptide synthesis, which is restricted to simple linear extension using the 20 proteinogenic aas (21 if the rare alternative aa selenocysteine is considered), NRPS can utilise over 400 molecules as substrates [9] to produce biologically active linear, cyclic and branched cyclic molecules [3, 10-12]. NRPS derived natural products include molecules with antibiotic, cytostatic, bio-surfactant and immunosuppressive activity and the utility of NRPS for drug discovery is not limited to discovery and characterisation of naturally occurring enzyme systems. The modular nature of NRPS makes them potentially amenable to manipulation, either rationally or randomly, in order to produce novel bio-active compounds [1, 12-14]. Manipulation of NRPS, while by no means straight forward, is possible as a number of

successful experiments have shown [13, 15-20]; however these same experiments have also shown that before NRPS manipulation can be freely used as a tool for generation of novel biomolecules a deeper understanding of the way in which these enzymes function is required.

1.2 – An overview of core NRPS domains

NRPS modules can be further broken down into a series of discrete domains. These domains are considered autonomous and usually retain catalytic activity when purified and characterised individually [1, 21, 22]. NRPS domains can either be core domains (required for synthesis of a peptide backbone) or accessory domains (catalysing chemical modification of aa constituents). There are three domains mandatory for the function of a NRPS elongation module: these are the adenylation (A) domain, the thiolation (T) domain (also known as the peptidyl carrier protein, or PCP domain) and the condensation (C) domain [1, 23]. The A-domain serves three purposes: 1) aa substrate selection; 2) activation of that substrate in a reversible reaction that consumes ATP (Figure 1.2A reaction 1, 1.2B); 3) transfer of the activated amino adenylate to a 4'-PP cofactor on the T-domain.

The T-domain is the site of aa/peptide linkage to the NRPS module. This linkage occurs via the formation of a thioester bond that displaces the adenylyl group from the activated aa [1-3] (Figure 1.2A, reaction-2; and Figure 1.2C). The C-domain catalyses the condensation of substrates tethered to the 4'-PP cofactors of two adjacent modules [24](Figure 1.2 A, reaction-3). The termination module (Figure 1.1) of an NRPS system contains a thioesterase (TE) domain. The TE-domain is present only once in any given NRPS system and is responsible for release of the final product from the enzyme system by cleavage of the terminal thioester linkage (Figure 1.2A, reaction-4).

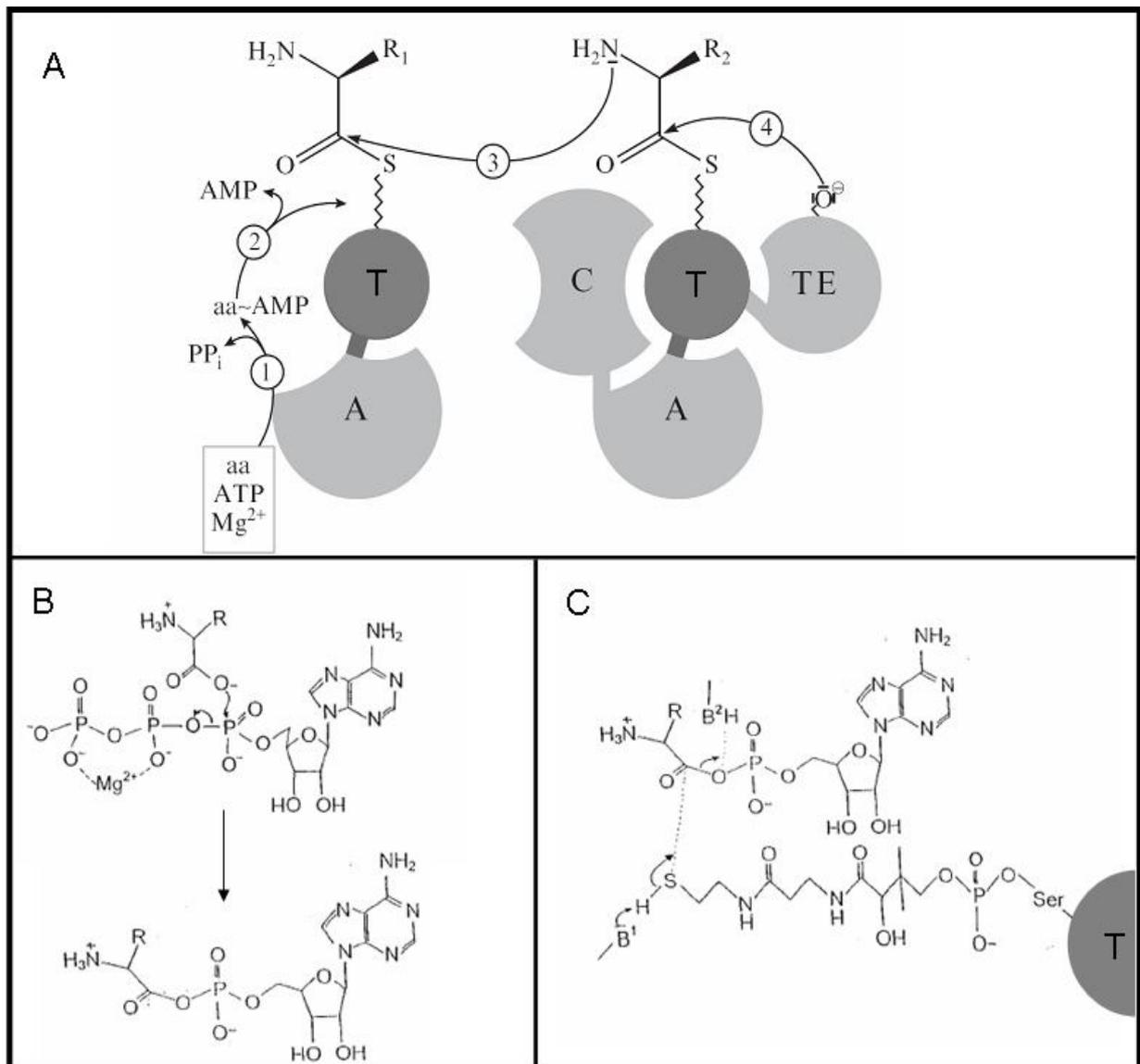


Figure 1.2 –Function of core domains

(A) Overview of the reactions catalysed by each of the core NRPS domains. (B) Detail of aa adenylation catalysed by the A-domain (reaction 1 in panel A). The adenosine moiety of ATP is linked to the substrate aa via its hydroxyl group, liberating PP_i (C) Detail of aa transfer to the 4' PP cofactor of the T-domain (reaction 2 in panel A). The adenosine moiety is displaced as a result of nucleophilic attack by the activated sulphur group of the 4' PP cofactor resulting a thioester linkage of the substrate aa to the T-domain. Panel A reproduced from [3], panels B and C adapted from [23].

1.3 – Adenylation domain structure and catalytic mechanism

A-domains consist of approximately 550 residues. To date the crystal structures for 3 separate A-domains have been solved, shedding some light on their mode of action. The first structure to be elucidated was that of the L-phenylalanine activating gramicidin synthetase module PheA [25], following this the structure of the 2,3-hydroxybenzoate (2,3-DHB) specific DhbE was solved [26] and most recently the structure of the D-alanine activator DltA has been determined [27]. Comparison of these three structures revealed a highly conserved fold for this enzyme family (Figure 3A), in spite of low sequence identity [3]. Although they catalyse a similar reaction, A-domains are structurally and evolutionarily unrelated to aa-tRNA synthetases [28]; instead the fold of A-domains is similar to that of other adenylation forming enzymes such as firefly luciferase from *Photinus pyralis* and acetyl-CoA synthetase from *Salmonella enterica* [1, 3]. Interestingly, firefly luciferase and PheA share only 16% aa identity, but have the same general structure [25]. Since other NRPS A-domains exhibit much higher identity than this, the structure elucidated for PheA is likely to represent a prototype for the structure of all NRPS A-domains [29], an assumption that has been supported by the elucidation of subsequent A-domain structures [26, 27].

There are 10 motifs (A1-A10) that are highly conserved in all NRPS A-domains. Prior to the elucidation of the first A-domain crystal structure, site directed mutagenesis (SDM) [30-33] and photo-affinity labelling [34] experiments were used to probe the role of these conserved motifs in A-domain function. The results of these experiments suggested that these were the residues important for ATP binding, substrate α -amino and α -carboxyl group binding and ATPase activity; processes common to all A-domains. In the crystal structure of PheA, most of the 10 core A-domain motifs were found to cluster around the substrate binding site, consistent with the putative role of these motifs in α amino/carboxyl recognition and ATP binding/hydrolysis.

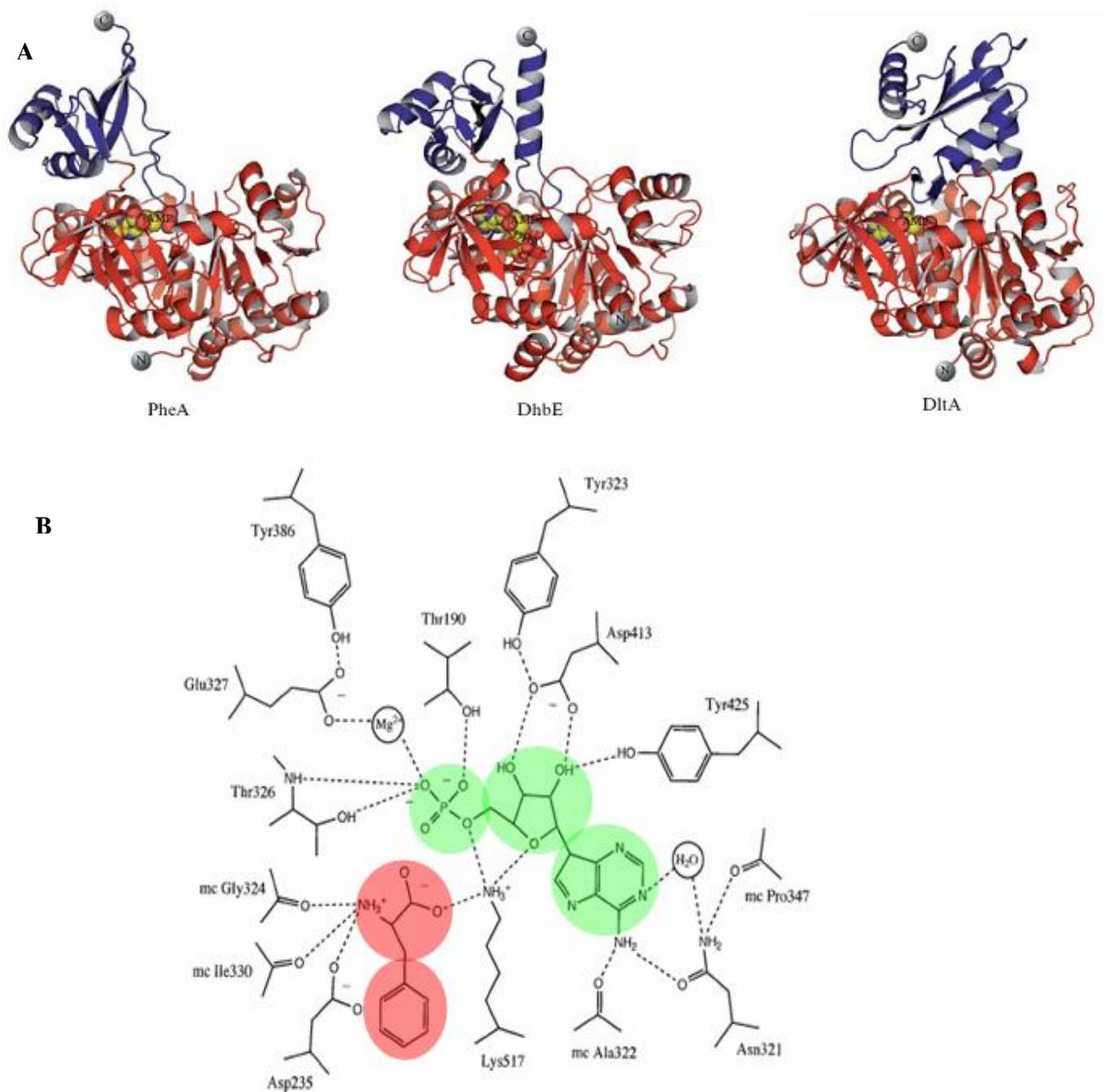


Figure 1.3 – Adenylation domain structure

A – Domain organisation and secondary structure elements of three adenylation domains are shown. In each case the active site residues are located between the large N-terminal (red) and small C-terminal (blue) sub-domains. PheA and DltA are shown with only ATP bound in the active site, DhbF is shown with its adenylation substrate (2,3-DHB-AMP) at the active site.

B – Hydrogen bond network for aa adenylation binding in PheA*. The hydrogen bonds shown are those involved in interactions with the adenylation group (green) and the α -amino and α -carboxyl groups of the aa (red). These interactions are common to all α -aa activating A-domains, and accordingly, the residues involved are highly conserved. Panel A reproduced from [3], panel B adapted from [25].

***In the crystal structure solved for this enzyme the aa adenylation was hydrolysed to phe and AMP**

The structures of all A-domains thus far elucidated consist of two sub-domains, a large N terminal domain of ~ 450 aa and a small C terminal domain of ~ 100 aa. The two sub-domains are linked by a hinge region of 5-10 aa (Figure 1.3A), and the active site of the A-domain is located at the interface of these two sub-domains [25-27]. A-domains are dynamic entities, the relative orientation of the large and small sub-domains changes during the catalytic cycle and the domain adopts three distinct conformations. Based on structural data and information about key residues involved in each of the catalytic steps of the A-domain (substrate binding, adenylation, transfer to the T-domain) a model for A-domain conformation switching and function has been proposed [27]. In this model, outlined in Figure 1.4, the open conformation promotes entry of the substrate aa, and ATP/Mg²⁺ cofactors to the active site cleft. Binding of ATP promotes rearrangement of the large and small sub-domains, bringing all active site residues together and leading to aa adenylation. After adenylation 4'-PPi release is proposed to trigger rotation of the two domains relative to each other by 140 ° bringing in to place residues involved in thiolation. The reactive adenylated substrate is precisely positioned, and presumably stabilised by hydrophobic and van der Waals interactions as well as an extensive network of hydrogen bonds (Figure 1.3b) [25-27]. There are 10 key residues that form the peptide binding pocket of the A-domain, and the identity of these residues plays a key role in determining which aa an A-domain will activate. In the majority of cases the identity of these residues can be used to predict A-domain specificity [25, 29, 35, 36].

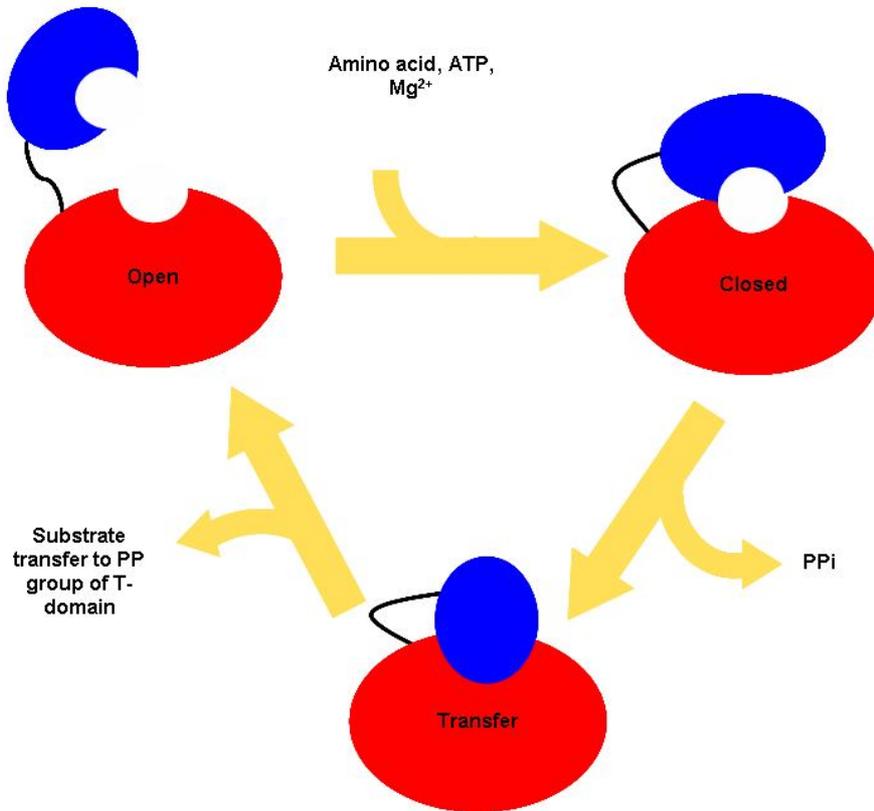


Figure 1.4 – The A-domain reaction cycle

In the open conformation, entry of the aa substrate, ATP and Mg²⁺ into the active site is promoted. Recruitment of the correct substrate triggers a conformational switch to the closed state, in which the C-terminal sub-domain (blue) covers the substrate binding cleft of the N-terminal (red) sub-domain bringing together residues involved in substrate adenylation. Release of P_{Pi} results in rotation of the C-terminal sub-domain relative to the N-terminal sub-domain by 140°, this brings together residues involved in the subsequent thiolation reaction, after the aa has been transferred to a cognate T-domain the A-domain is reset to the open, substrate-binding conformation.

1.3.1 – A-domain specificity and the non-ribosomal code

Solving the crystal structure of the first NRPS A-domain, allowed Conti and co-workers [25] to determine exactly which residues lined the phenylalanine binding pocket of PheA. Of these 10 residues, two (asp235 and lys517) are common to all α -aa activating A-domains [25, 29, 35]. These invariant residues are involved in the formation of hydrogen bonds with the α -amino and α -carboxyl groups of the substrate (Figure 1.3B); the remaining eight residues surrounded the phe side chain and were proposed to be involved in substrate selection. It is possible to determine the position of the 10 residues of the aa binding pocket from sequence data alone by virtue of their position relative to the conserved motifs A4 and A5; and once extracted from the sequence, these residues serve as a basis for substrate prediction by comparison with equivalent residues from a database of A-domains of known specificity [25, 35-38].

Following the determination of the binding pocket of PheA Conti *et al* [25] used this process to show that the identity of the eight variable residues of the peptide binding pocket, when aligned for 24 A-domains is correlated with substrate specificity. The idea that the 10 residues of the binding pocket confer specificity was confirmed in a subsequent study performed by Stachelhaus and colleagues [36]. In this work the 10aa signature sequences from 160 A-domains were extracted and aligned against each other. In the resulting phylogenetic tree, signature sequences clustered according to substrate specificity, and were only slightly affected by organism of origin. This result validated the method for binding pocket extraction as well as the role of the 10 aa signature in substrate selection. By determining a consensus sequence for each of the specificity clusters it was possible to derive a non-ribosomal code in which a particular 10 aa signature recognises a particular substrate, or set of structurally related substrates.

In order to further validate the newly derived non-ribosomal code, SDM experiments were employed; among these were mutations that, based on *in silico* predictions, completely switched the specificity of AspA from asp to asn by introducing a single point mutation. A similar study performed by Challis and co-workers [35] developed a two dimensional model of the aa binding pocket, onto which the corresponding residues of other adenylation domains

could be superimposed. This model made it easy to identify residues likely to be involved in key interactions with substrates and led to the development of an algorithm for automated prediction of A-domain specificity using sequence data. One notable difference between the two models is that residues occupying a position equivalent to cys322 of PheA were omitted from the model proposed by Challis *et al* [35] on the basis that the side chain of this residue points away from the binding pocket of PheA [25]. Subsequent to this however, a three-dimensional model of the substrate binding pocket of the L-thr activating enzyme PvdD [37] suggested that the PvdD equivalent of this position (his806) formed hydrogen bonds with the side chain of pvdD bound L-thr. The same study also found that the substrate binding pocket of PvdD would likely be highly selective for L-thr over other threonine isomers; this was supported by biochemical characterisation that showed an unusually high stereo-selectivity for L-thr over D-thr and L-*allo*-thr.

The most recent advancement in prediction of A-domain specificities has been the development of heuristic transductive support vector machines (TSVM) prediction software [39]. This method generates a profile based on the physical and chemical properties of the aas in the entire eight angstrom binding pocket of an A-domain (which includes but is not limited to the 10 residues extracted for use by the previous methods). The program then evaluates which substrate this profile best matches based on what it has “learned” from A-domains of known specificity. In a study encompassing all 1240 A-domains for which sequence data were available, the TSVM based software was able to give predictions for an additional 18% of sequences when compared to the 10 residue extraction methods. The fact that this method was able to give improved prediction power suggests the 10 residues of the non-ribosomal code while certainly the most significant, are not the only contributors to A-domain specificity.

1.4 – The Thiolation domain

4'-PP cofactor attachment to NRPS enzymes occurs at the T-domain. T-domains contain a highly conserved signature motif LGG(HD)S(LI), where the underlined serine residue in this motif is absolutely conserved and is the site of 4'-PP linkage to the T-domain [40]. The attachment of a 4'-PP cofactor to every T domain in an NRPS system is absolutely required

for function as it is the site of substrate aa and peptide intermediate linkage to the enzyme complex during synthesis [6, 40]. The conversion from inactive/apo to active/holo T-domain occurs by nucleophilic attack of the serine β hydroxyl on the pyrophosphate linkage of coenzyme A (CoA), resulting in transfer of a 4'-PP group to the serine and liberation of ADP (Figure 1.5). This reaction is catalysed by a class of enzymes known as phosphopantetheinyl transferases (PPTases)[40-42]. The 4'-PP group, once attached, is proposed to function as a swinging arm that allows flexible positioning of substrates and intermediates during synthesis [1-3].

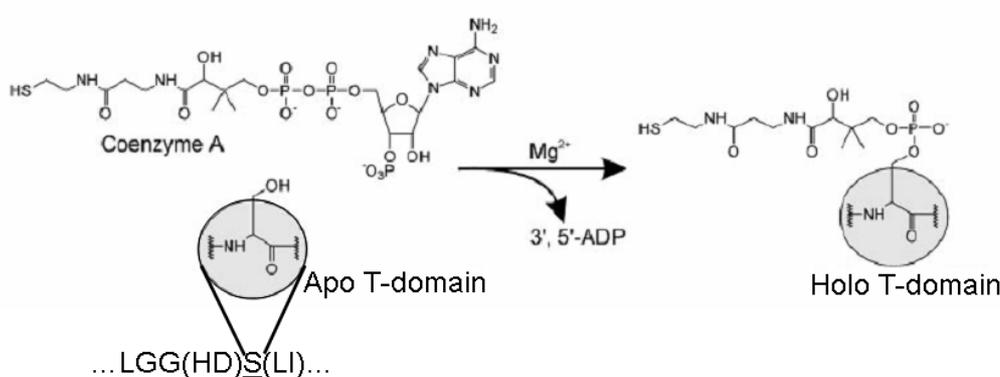


Figure 1.5 – Post translational activation of T-domains

The serine side chain in the core T-domain motif LGG(HD)S(LI) is the site of 4'-PP cofactor attachment. The PP group is derived from CoA and its attachment is catalysed by a PPTase in a Mg²⁺ dependent reaction. Figure reproduced from [40].

NMR studies revealed the solution structure of the TycC3 T-domain from *Bacillus brevis* to be a simple four α -helical bundle extending 37 residues either side of the 4'-PP attachment site. This was the first structural elucidation of a NRPS T-domain comparison, and showed that the overall topology of the T-domain resembled that of Acyl carrier proteins (ACPs) from fatty acid and poly ketide synthetases but with different length, polarity and relative orientation of the helices in each case [43].

The T-domain is required to interact with multiple protein partners at various stages of synthesis (Figure 1.6). First, an **apo** T-domain must be recognised and activated by a PPTase enzyme to give the activated **holo** T-domain. The holo T-domain must then recognise a

specific A-domain which catalyses substrate loading on to the 4'-PP cofactor to give an **acylated** T-domain. Acylated T-domains must then interact with the appropriate C-domain to position substrates for peptide bond formation. Accurate temporal and spatial control of T-domain interactions is therefore essential to NRPS function and fidelity, although it remains unclear how such a simple domain recognises the correct protein partners at the correct time [3, 43]. It is likely that these interactions are modulated by conformational changes in the protein that result in presentation of specific residues involved in partner recognition, and there is strong evidence that this is the case. Investigating the interactions between T-Domains and other NRPS domains became a major focus of this PhD project and is discussed at length in Chapter 6.

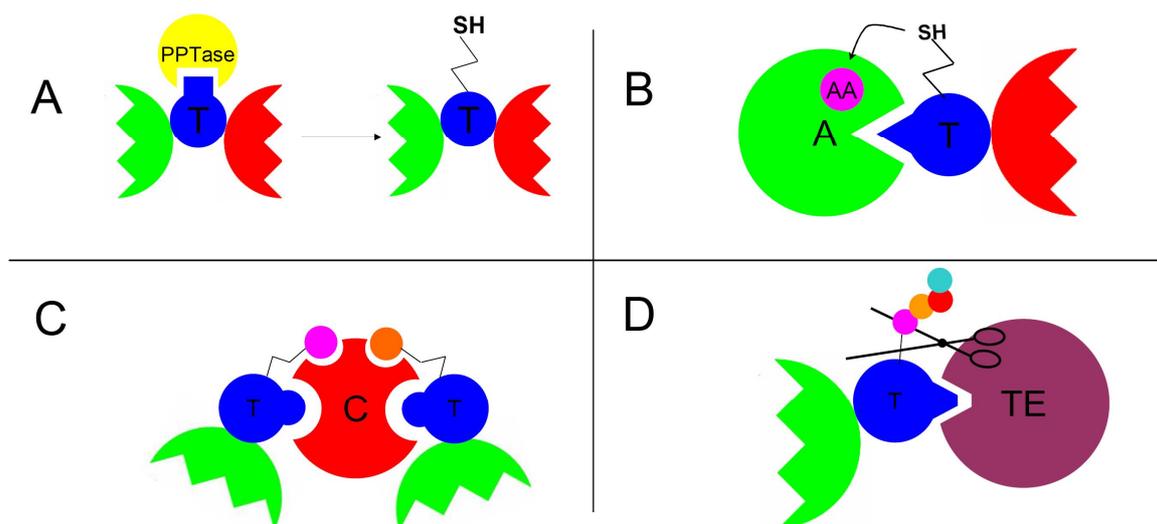


Figure 1.6 – Thiolation domain interactions during synthesis

A: Apo T-domains must be recognised by a cognate phosphopantetheinyl transferase (PPTase) and converted to the active holo form by the addition of a 4'-PP group. **B:** Holo T-domains must be recognised by the correct A-domain to allow transfer of an aa substrate to the 4'-PP group. **C:** Adjacent T-domains must present substrates to a C-domain to facilitate the condensation reaction.

D: T-domains located in a termination module present the final product to a TE-domain for release from the enzyme system.

In a more recent NMR study of TycC3 T-domain structure [44], Koglin *et al.* describe three distinct, slowly inter-converting conformations associated with particular T-domain states.

One of these, the A-state, is seen exclusively in apo T-domains while another, the H-state, is seen only for holo T-domains. The third conformer, the A/H state, is common to both apo and holo T-domains. Of the three states, the A/H state is the most compact, with the four α -helices arranged in a distorted anti-parallel bundle with the third helix protruding from the main globular body of the protein. The A-state is the most flexible conformation. In this state the third helix is not present and the residues instead form an extended loop which is embedded in the core of the protein. In the H-state helix III is again unravelled, and the relative orientation of helices I, II and IV is altered. This alteration results in the 4'-PP cofactor being displaced by 100° relative to its position on the A/H state, corresponding to a 16\AA relocation of the active thiol group toward the C-terminal of the domain (Figure 1.7). The active site serine (S45) is proposed to be important in modulating these conformational changes with the free hydroxyl forming different hydrogen bonds in the A vs. the A/H state (to R47 and H44 respectively). Mutation of the active site serine was found to lock the T-domain in the A state, providing supporting evidence for the importance of S45 in modulating conformational switches [44].

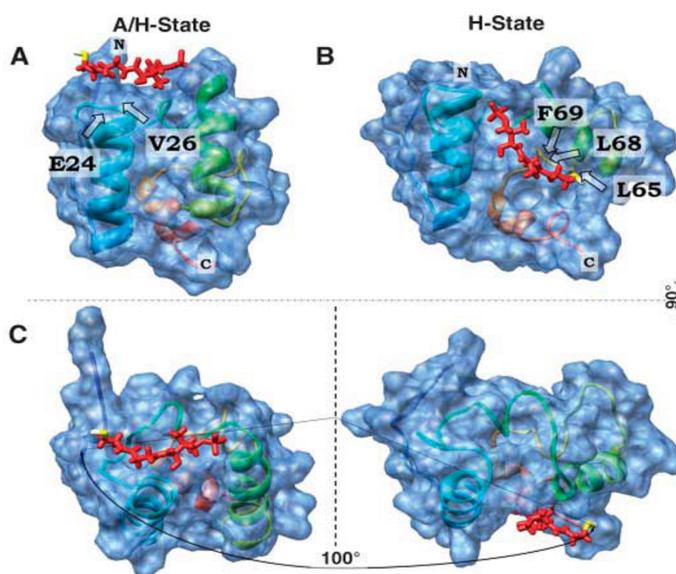


Figure 1.7 – T-domain conformational switching

Switching of the T-domain between the A/H (A) and H states (B) causes the 4'-PP cofactor (red) to move in a 100° arc, resulting in a 16\AA shift of the active thiol group (yellow) toward the C-terminal of the domain. Residues that contact the 4'-PP group in each conformation are indicated. Panel C is the same image rotated 90° to highlight the swinging motion of the 4'-PP group. Figure reproduced from [44].

1.5 – The Condensation domain

Peptide bond formation in NRPS is controlled by the C-domain. During the condensation reaction the α -ammonium group of the downstream 4'-PP linked aa acceptor carries out nucleophilic attack on the carbonyl group of an upstream 4'-PP linked peptide donor. This process breaks the thioester linkage of the upstream peptide and results in its translocation onto the downstream module [3, 45] (Figure 1.8).

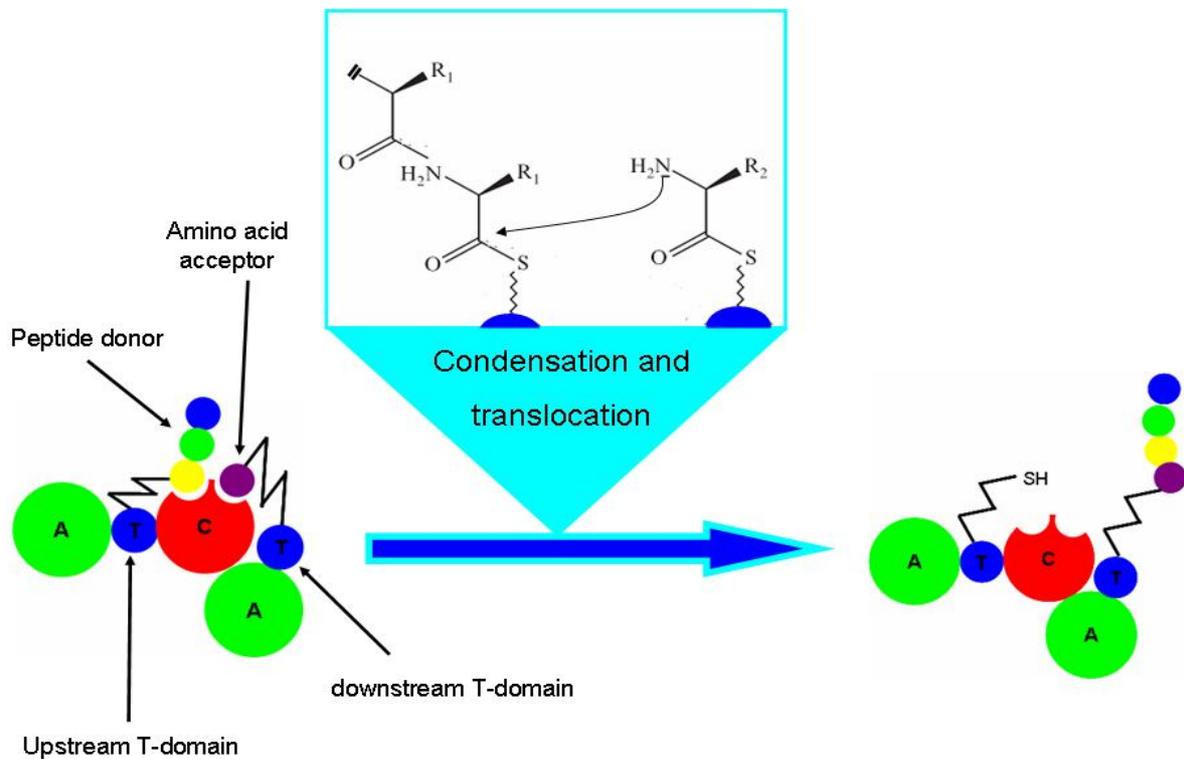


Figure 1.8 – C-domain function

The C-domain catalyses condensation and peptide chain translocation during synthesis. The amino group of the downstream aa acceptor carries out nucleophilic attack on the thioester linked peptide donor, as a result, the thioester linkage is broken and the peptide donor is translocated to the downstream T-domain.

C-domains show no significant sequence homology to any proteins other than NRPS and until 1998 there was no biochemical data available on these domains, in fact they were initially considered to be nothing more than inactive spacers between adjacent AT di-domains [7]. This view changed following identification of a HHxxxDG motif conserved in all known C-domains [46]. This motif is also found in the active site of CoA-dependent acyl transferase proteins, such as chloramphenicol acyl transferase, leading to the proposition that the C-domain is responsible for peptide bond formation and that this reaction proceeds by a similar mechanism to that elucidated for chloramphenicol acyl transferase [46]. In this model, the second histidine of the HHxxxDG motif (H_{CAT}) is essential for catalysis as it acts as a base, deprotonating the α -ammonium group of the acceptor peptide.

The putative role of the C-domain in peptide bond formation was subsequently confirmed when it was found that the C-domain of *tycB1* was able to catalyse peptide bond formation between two T-domain linked substrates *in vitro* [24]. In the same study the essential role of H_{CAT} was also supported by mutational analysis. More detailed mutational analysis of the condensation domain [45] led to a model for peptide bond formation in NRPS that is dependent on acid base catalysis, the mechanism for which is outlined in Figure 1.9A. The role of H_{CAT} as a catalytic base in the C domain has recently been challenged, with newer structural data suggesting that electrostatic interactions, rather than general acid base catalysis, may control peptide bond formation by the C-domain [47]. In the newer model, H_{CAT} is calculated to be protonated at physiological pH and so is unable to act as a base; instead it is proposed instead to participate in hydrogen bonding that stabilises the tetrahedral intermediate of the condensation reaction (Figure 1.9B).

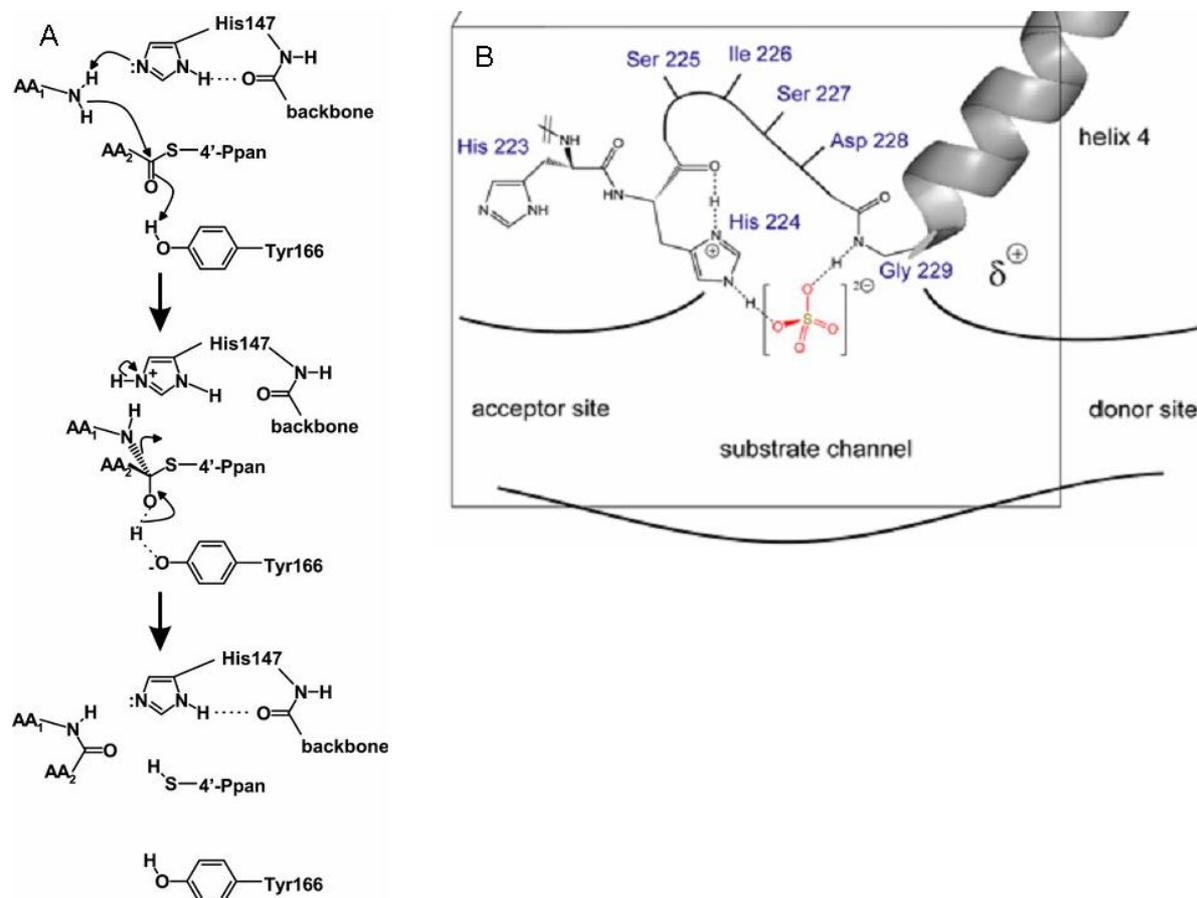


Figure 1.9 – Two proposed mechanisms for peptide bond formation in NRPS:

A – The mechanism proposed by Bergendahl and co-workers. H_{CAT} (his 147) acts as a base, deprotonating the α-ammonium group of the aa acceptor (AA1) and promoting nucleophilic attack on the carboxyl group of AA2. The hydroxyl group of tyr166 serves to stabilise the subsequent tetrahedral intermediate. his147 is invariant in all C-domains, but tyr166 may be replaced by another residue with a free hydroxyl group. Figure reproduced from [45].

B – The mechanism proposed by Samel and co-workers. This mechanism is based on structural data and pKa calculations for the TycC6 C-domain. H_{CAT} (his224) is protonated at physiological pH and therefore unable to act as a base, instead it participates in a hydrogen bonding network that stabilises the tetrahedral intermediate of the reaction. The sulfate ion (red) was found to be co-crystallised at the active site of the domain and is believed to take the place normally occupied by the tetrahedral intermediate of the condensation reaction. Figure reproduced from [47].

1.5.1 – C-domain specificity

The substrate specificity of C-domains is poorly understood compared to that of A-domains; this is mainly due to the difficulty of readily detecting the product of a condensation reaction. There are however some data available on the specificity of C-domains. The first study of C-domain selectivity [48] used a two module system ($A_{\text{phe}}\text{TE}-\text{CA}_{\text{pro}}\text{T}$) in which L-phe and an L-pro specifying A-domain, along with their cognate T-domains and a C-domain, catalyse the production of free diketopiperazines or enzyme bound dipeptides, both of which can be detected if one of the aas in the condensation reaction has been radio-labelled. The selectivity for donor and acceptor residues was tested by artificial loading of T-domains with native and non-native aa-CoA conjugates. This study found low selectivity at the peptide donor site and high specificity at the aa acceptor for both chirality and size of the aa side-chain (Figure 10A). One limitation of this study was the inclusion of an Epimerisation (E) domain in the donor module ($A_{\text{phe}}\text{TE}$). As discussed in Section 1.8, E-domains catalyse the conversion of aas from L to D-isomers, so inclusion of this domain would have resulted in any T-linked donor substrates being converted to their D-isomers before condensation. As such, no conclusion regarding stereo-specificity for donor residues can be drawn from this study. A subsequent study [49] used non-native aminoacyl-N-acetylcysteamine thioesters (aminoacyl-SNACs) as free substrates for the same condensation domain ($\underline{\text{C}}\text{A}_{\text{pro}}\text{T}$), thereby eliminating the need for the first ($A_{\text{phe}}\text{TE}$) module and preventing the epimerisation associated with using enzyme linked substrates. Results of this second study confirmed high selectivity at the acceptor site but also found stereo-selectivity at the donor site that had not previously been apparent. Overall, the results from both of these experiments suggest that condensation domains display stringent selectivity for both aa side-chain and chirality of the aminoacyl acceptor and are selective only for chirality at the donor site. It should be noted however that in both cases the *in vitro* assays employed may not be representative of *in vivo* synthesis conditions; this is particularly relevant to the use of aminoacyl-SNACs as C-domain substrates, as these are not linked to an upstream T-domain.

1.5.2 – C-domain evolution and specialisation

Phylogenetic analysis of available C-domain sequences showed that five broad categories could be identified based on sequence at key positions. The categories reflected specialisation

of the C-domain to a particular function in an NRPS assembly line and were as follows: Starter C-domain, located in the first module of a synthetase system; dual E/C domains, those which catalyse not only condensation but also epimerisation; heterocyclisation domains, those which catalyse both condensation and heterocyclisation of substrates; LCL C-domains, those which catalyse condensation between an L isomer at both the donor and acceptor site; and DCL C-domains, those which catalyse the condensation of a D-isomer at the donor site to an L-isomer at the acceptor site [50]. It is not immediately obvious why distinct LCL and DCL C-domains have evolved, but this may stem from the nature of epimerisation reactions (1.8) in NRPS. DCL C-domains are located immediately after an epimerisation domain however E-domains generate a mixture of both D and L-isomers. Discrimination between the two isomers by a DCL C-domain is therefore essential for incorporation of the correct (D) isomer in the final peptide product [1, 50, 51].

1.5.3 – C-domain structure

To date there are three published structures for NRPS C-domains. The first to be solved was for the stand alone C-domain VibH [52], then subsequently the structure of a T-C di-domain [47] was elucidated. Most recently the structure of a C-domain in the context of an entire NRPS termination module was solved [53]. All of the structures revealed that the C-domain consists of two structurally similar sub-domains, each of which resemble the well known chloramphenicol acyl transferase fold. C-domains do not possess a deep substrate binding pocket, but rather have a substrate cleft formed by a V-shaped arrangement of the two sub-domains that contains the active site H_{CAT} [47, 52] (Figure 1.10). The structure of the C-domain possesses distinct donor (peptide) and acceptor (aa) substrate faces, which are the sites of interaction with upstream and downstream T-domains respectively [47, 52]. While structural data has allowed a detailed understanding of the mechanism of peptide bond formation in NRPS, there is still no definitive structural basis for the selectivity exhibited by the C-domain.

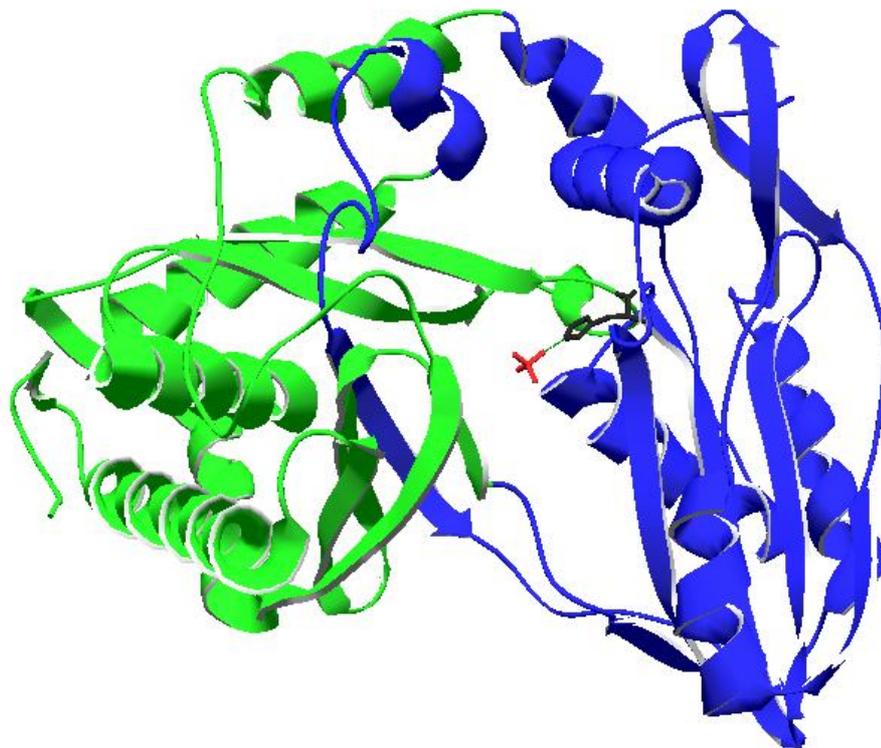


Figure 1.10 – The C-domain of TycC6

The two sub-domains have been coloured green and blue. H_{CAT} (black) is located in the cleft between these two domains. In this structure a sulfate (red) was found at the active site. This sulfate ion is believed to be representative of the tetrahedral intermediate of a condensation reaction. The image was generated from pdb file 2JGP [47] using Deepview.

1.6 – The Thioesterase Domain

TE-domains are found only once in an NRPS system and are always located in the terminal module[1]. TE-domains catalyse the release of the peptide product from its 4'PP cofactor via cleavage of the thioester bond. The role of the TE-domain in product release was initially assigned based on sequence similarity with TE-domains from FAS and PKS, and later confirmed by mutagenesis and module rearrangement experiments [54]. Relocation of the surfactin synthetase TE-domain from the terminal module to various internal modules was shown to result in generation of truncated products, the identity of which corresponded to the

new location of the TE-domain. Deletion of the same thioesterase domain resulted in complete ablation of surfactin synthesis, both results being consistent with the proposed role of the TE-domain in chain termination and product release [54].

1.6.1 – Mechanism of product release by TE-domains

NRPS TE-domains contain a highly conserved ser-asp-his catalytic triad. This combination of nucleophile-acid-histidine is a well characterised motif common to enzymes of the α/β -hydrolase family [54-58]. Mutagenesis studies using the TE-domain from the *E. coli* NRPS EntF found that both the ser and his residues were essential for catalysis whereas the asp was not [59], indicating that this enzyme contained a catalytic dyad, rather than the typical triad. As illustrated in Figure 1.11, product release by NRPS TE-domains involves formation of a peptidyl-O-TE intermediate in which the peptide chain to be released is transferred from the terminal T-domain to the side chain of the catalytic ser residue in the TE-domain. During this process the catalytic his residue acts as a base activating the ser for nucleophilic attack on the thioester linked substrate (Figure 1.11, inset box). Most commonly the TE-linked intermediate is subsequently cleaved from the enzyme by intra-molecular cyclisation to generate a macrocyclic lactone (where the hydroxyl group acts as a nucleophile) or lactam (where the amino group acts as a nucleophile). Less commonly, hydrolysis of the intermediate results in release of a linear peptide [56, 57, 59-61]. If the nucleophile in the cyclisation reaction is the N-terminal amino group (as in Figure 1.11) a cyclic lactam will be generated, whereas use of an amino or hydroxyl group located in the side-chain of an internal aa leads to the generation of a branched cyclic lactam or lactone, respectively.

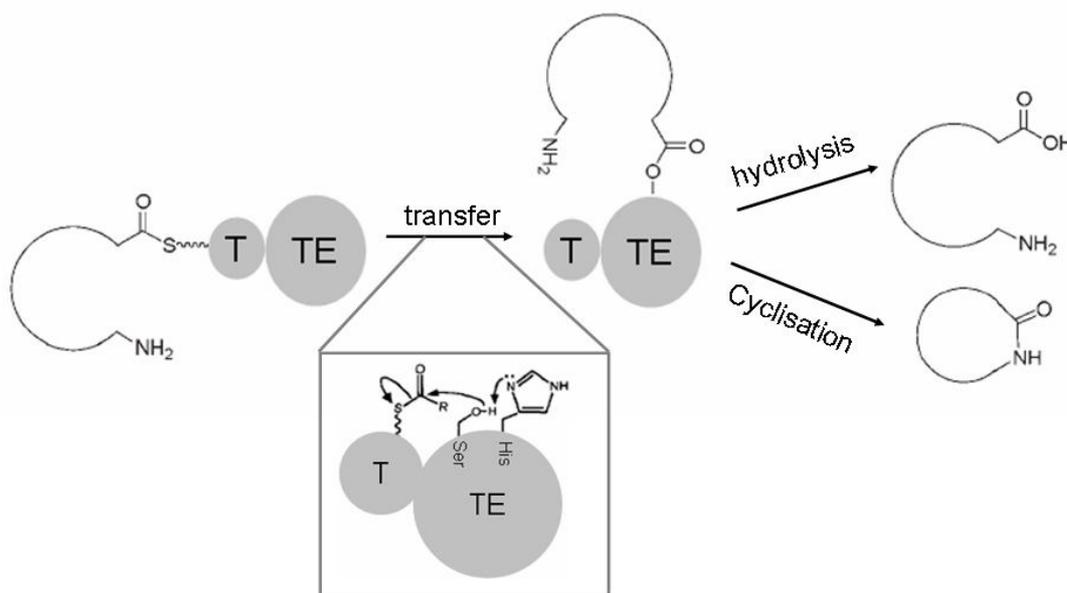


Figure 1.11 – Product release by TE-domains

A peptide chain linked to the terminal T-domain of an NRPS system is transferred to the side chain of a serine group in the TE-domain to generate an O-linked intermediate, which can then be released by either hydrolysis or cyclisation. In this example the nucleophile in the cyclisation reaction is the N-terminal amino group and the macrocycle generated is a lactam. The role of the catalytic ser and his residues during substrate transfer is shown in the inset box.

1.6.2 – TE-domain structure

To date, the structures for two TE-domains have been elucidated by X-ray crystallography [55, 62]. In each case, as expected, the overall fold of the domain resembled that of the α/β -hydrolase family. Like C-domains, TE-domains possess a deep substrate cleft that can accommodate linear peptide intermediates; this cleft is lined with hydrophobic and aromatic residues and contains the ser-his-asp catalytic triad. The hydrophobic environment of the substrate binding cleft is presumed to exclude water molecules and thereby prevent aberrant hydrolysis of linear peptide intermediates [55, 62, 63]. There have been two alternative models proposed for exclusion of water from the TE substrate binding cleft. Bruner and co-workers proposed a model in which the globular bulk of the TE-domain forms a bowl-like structure which is sealed by a protruding helix after substrate entry [55]. A subsequent model has been

proposed by Samel and co-workers in which part of the peptide chain not directly involved in cyclisation forms a lid that seals the substrate cleft from water [62]. The latter of these models is more consistent with existing biochemical data which has shown that TE-domains exhibit low specificity for residues that do not directly participate in a cyclisation reaction, as would be expected if these residues were excluded from the substrate binding cleft [3, 64, 65].

1.6.3 – Type II TE-domains

Type II TE-domains, unlike those discussed in the preceding sections, do not catalyse product release from an NRPS enzyme. The role of the type II TE-domain is to repair misprimed T-domains which would otherwise terminally inhibit an enzyme system [66-68]. In NRPS, the two types of T-domain mispriming that can occur are: attachment of acetylated 4'-PP derived from acetyl-CoA; or transfer of an incorrect substrate by an A-domain. Both of these events are believed to be common. In the case of acetylated 4'-PP attachment, poor discrimination between CoA and acetyl-CoA by a PPTase enzyme is the cause of mispriming, whereas promiscuous activation of non-cognate aa substrates by an A-domain is the cause of incorrect substrate attachment. Type II TE-domains are therefore essential for efficient product synthesis by NRPS enzymes, as without them assembly lines would rapidly stall due to the inability of misprimed T-domains to be processed by downstream partners.

1.7 – Phosphopantetheinyl transferases

PPTases catalyse the post translational attachment of a 4'-PP group to carrier proteins (T-domains) of NRPS, PKS and FAS [40, 42]. PPTases can be divided into two broad classes based on primary aa sequence and substrate specificity. The first of these classes to be discovered was the AcpS type PPTases. These proteins are approximately 120 aa in length and catalyse post translational activation of acyl carrier proteins (ACPs) found in FAS and PKS [40, 69]. The second class, the Sfp type PPTase are approximately 240 aa in length and preferentially activate T-domains found in NRPS and PKS clusters. Sfp type PPTases can be further divided into broad specificity and narrow specificity types. Broad specificity PPTases such as PcpS [69] of *P. aeruginosa*, and the prototypical Sfp of *B. subtilis* [40, 70] can activate a wide range of T-domains of both primary and secondary metabolic pathways. Narrow

specificity PPTases such as EntD of the *E. coli* enterobactin pathway are specialised, and act only on T-domains of proteins found in a particular secondary metabolite cluster [42].

1.8 – NRPS tailoring domains

In addition to the core domains already described, NRPS can be further embellished with a number of tailoring domains. These are domains that catalyse chemical modification of aa constituents that are essential for biological activity of the final product of an assembly line. Tailoring domains can be integrated into specific points in a NRPS assembly line where they modify particular enzyme linked intermediates, or they can act as stand alone enzymes that modify a peptide after it is released from an assembly line [71, 72]. An exhaustive discussion of tailoring domains is beyond the scope of this thesis, however a brief outline of some of the more common tailoring domains and reactions is given below.

The most commonly occurring of the tailoring domains is the **epimerisation (E)** domain; E-domains are usually inserted after a T-domain to give a module with the domain order CATE. E-domains catalyse racemisation of an L-aa to yield a mixture of D and L-isomers [71, 72] and act preferentially on aas immediately after their incorporation into a peptide chain [73, 74]. Although E-domains produce a racemic mixture, selectivity for the D-isomer by the next C-domain ensures incorporation of only the D-isomer into the final product [50, 51, 73].

Methyl transferase (MT) domains catalyse methylation of aas and are typically integrated into an assembly line after an adenylation domain to give a module with the domain order CA(MT)T. MT-domains can catalyse attachment of a methyl group to a nitrogen (N-MT), carbon (C-MT) or oxygen (O-MT) atom of a specific aa [72]. N-MT domains (Figure 1.12) are most common of the MT domains; for example, synthesis of the immunosuppressant cyclosporin requires methylation of seven of the 11 aas in the molecule [75]. N-methylation of aas may also be catalysed by a stand-alone MT enzyme following release of a peptide from an assembly line, as is the case with synthesis of the vancomycin group antibiotic chloroeremomycin [76].

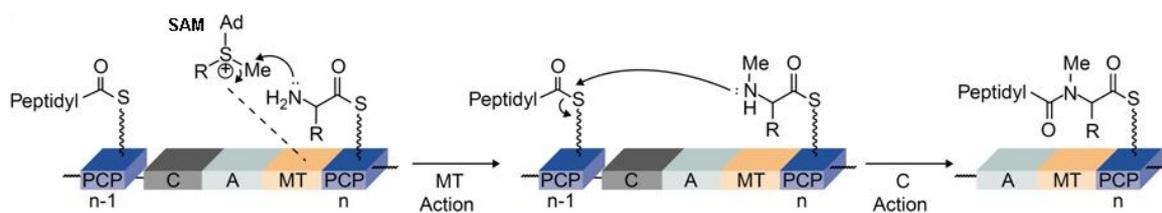


Figure 1.12 – N-methylation of amino acids by MT-domains in an NRPS module

Transfer of a methyl group to the α -ammonium of an enzyme linked intermediate is catalysed by an N-MT domain, with methylation occurring prior to peptide bond formation and requiring a S-adenosyl methionine (SAM) cofactor. Figure reproduced from [72].

Cyclisation (Cy) domains are variants of condensation domains that catalyse heterocyclisation prior to product release from an NRPS complex. Cy domains can act on the side chain of either cysteine or serine/threonine residues, with the former yielding a five-membered thiazoline and the latter a five-membered oxazoline (Figure 1.13).

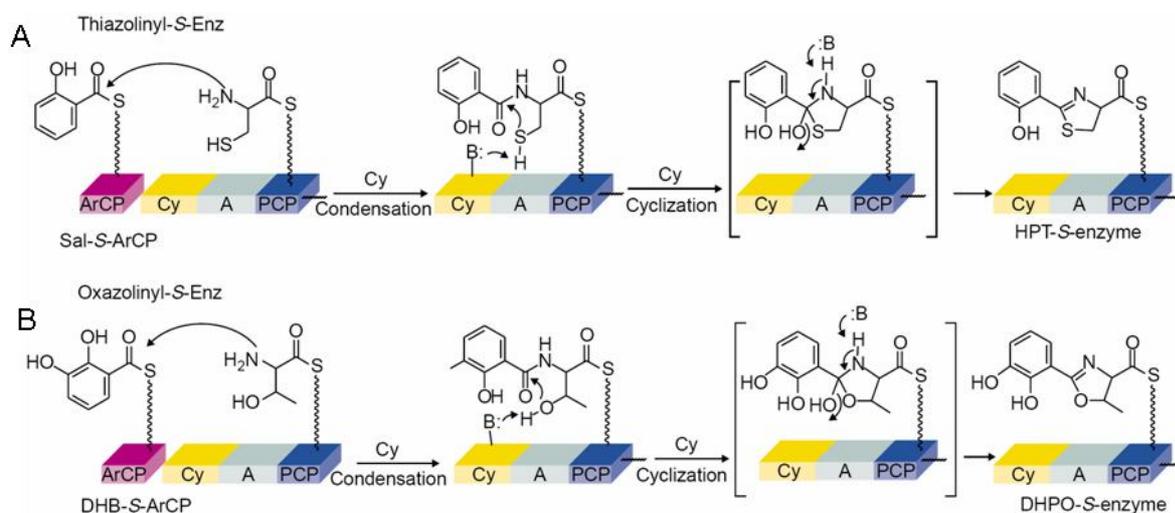


Figure 1.13 – Formation of thiazoline and oxazoline groups by Cy domains

A – In this example the Cy domain first catalyses a condensation reaction between salicylic acid (Sal) and cysteine, and subsequently intra-molecular cyclisation of the cysteine side chain to yield a five-membered thiazoline group.

B – In this example the condensation reaction is between di-hydroxybenzoate (DHB) and threonine. Subsequent cyclisation yields a five-membered oxazoline group. Figure reproduced from [72].

β -Hydroxylation of aas contained within NRPS-derived secondary metabolites is another common tailoring reaction. This most commonly occurs at tyrosine residues, as in synthesis of

the antibiotics nikkomycin and vancomycin [72, 77]; aspartic acid residues, as in synthesis of the phytotoxin syringomycin [78]; or histidine residues, as in the synthesis of the antibiotic bleomycin. β -Hydroxylation is carried out by standalone enzymes and can occur either before or after product release from an NRPS complex [72, 78].

Glycosylation of NRPS-derived peptides by dedicated glycosyltransferase enzymes (Gtfs) occurs after a nonribosomal product is released from the NRPS template. The most prominent example of this is the addition of various deoxy sugars to vancomycin type antibiotics [72].

Finally, there are a number of examples of FMN dependent **oxidation** (Ox) domains that are integrated into NRPS assembly lines. Ox domains are found either immediately after an A-domain as seen in the bleomycin synthetase system [79], or in some cases integrated into the A-domain [80]. In the case of bleomycin synthesis, oxidation is coupled to cyclisation, with the thiazoline formed by a Cy domain being oxidised to a thiazole by a subsequent Ox domain. A particularly interesting example of an integrated oxidation domain is found in the enzyme BpsA, synthesising the pigment indigoidine. In this enzyme, although the Ox domain is integrated into the A-domain, it appears to act on its substrate (L-gln) after this residue has been released from the NRPS enzyme [80].

1.9 – Manipulation of NRPS

Of all the potential biotechnological applications of NRPS research by far the most interesting – and potentially lucrative – is the construction of hybrid enzyme systems. The goal of such endeavours is either to alter the structure of an existing natural product – and in doing so modify its activity – or to generate entirely new compounds in a random combinatorial fashion [13, 14, 81-83]. The former strategy would be particularly useful for modification of NRPS-derived antibiotics – such as daptomycin [15, 16] and surfactin [20, 84] – to modulate activity or circumvent bacterial resistance mechanisms, whereas the latter strategy could be used to generate libraries of new molecules that could be screened for novel biological activities [13, 14, 65, 82]. One needs only to look at the variety of molecules currently known to be synthesised by NRPS complexes to see that the diversity of structures that can be created by

varying combinations of the number, type, order and configuration of aas in a peptide chain is astounding [9, 12]. Whether or not these complexes can be easily manipulated to unlock this potential remains to be seen. As discussed in the preceding sections of this chapter, much progress has been made toward understanding the function and structure of NRPS enzymes; however this is only beginning to be translated into truly robust platforms for the production of novel bioactive molecules. The synthesis of novel daptomycin derivatives (outlined 1.9.2) is the only example to date of NRPS manipulation that has allowed the generation of novel bioactive compounds with efficiency that could allow for large scale production.

1.9.1 – Module swaps and site directed mutagenesis

The general strategy for modification experiments in prokaryotes thus far has been to create deletion mutants lacking a particular domain or module that can subsequently be replaced by a modified DNA sequence via homologous recombination [20, 84], or complemented with a plasmid based expression vector encoding a modified protein [85]. Early attempts at engineering NRPS systems by A-domain swapping succeeded in generating altered products but often at a drastically reduced rate. For example, Stachelhaus *et al.* [20] successfully altered the structure of the lipopeptide antibiotic surfactin by substitution of the seventh and final AT bi-domain from surfactin synthetase for AT bi-domains from the gramicidin and ACV synthetase enzymes [20]. Using this strategy it was possible to recover modified surfactin in which the terminal leu residue had been exchanged for a cys (Figure 1.14), however the yield of altered product was only about 0.5-1.0% of the wild type enzyme system. This result was subsequently expanded upon by the same group, who conducted further manipulation experiments using the surfactin biosynthesis system [84]. In this later work, an internal leu-specifying AT bi-domain was exchanged for foreign modules. Using this strategy it was possible to derive modified surfactin in which the second leu residue was replaced with orn, however the products of the new enzyme system were unexpectedly truncated and the modified surfactin was again synthesised at a greatly reduced rate relative to wild type.

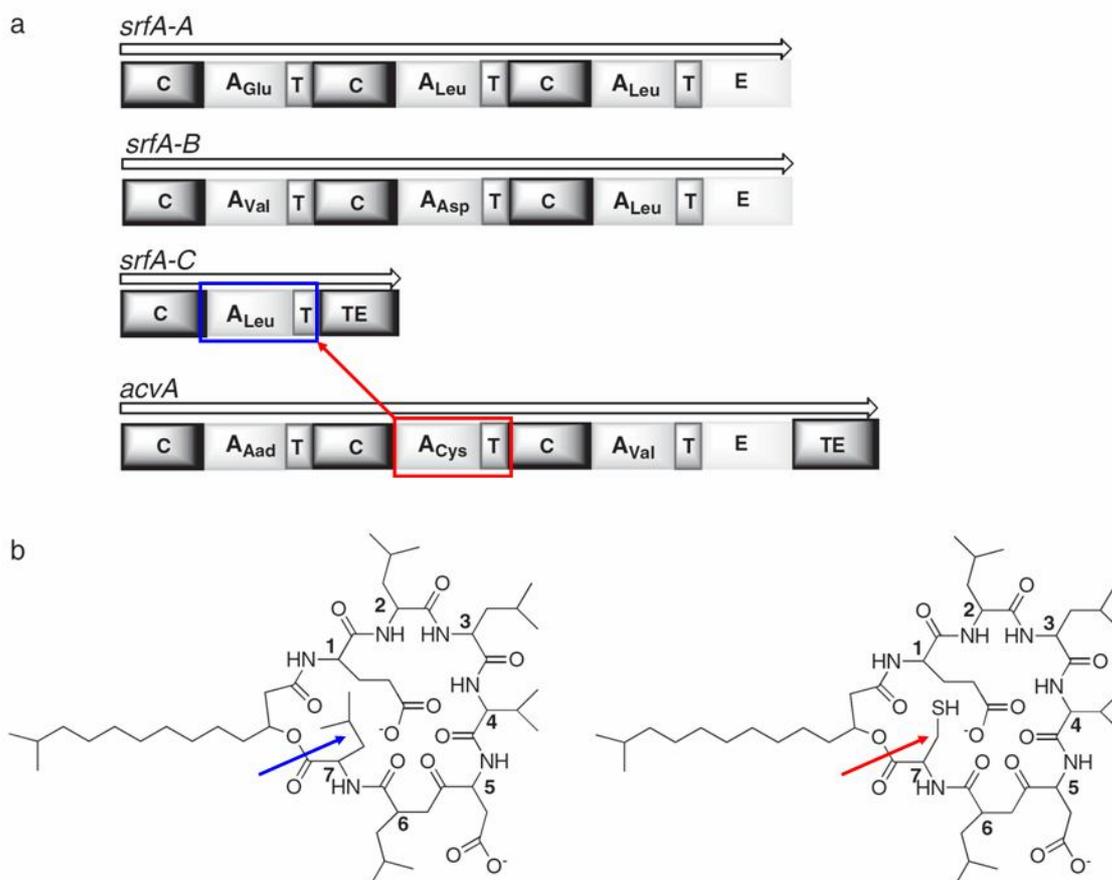


Figure 1.14 – Module swapping in the surfactin biosynthesis system

Swapping of a leu-specifying minimal module (AT) for one specifying cys resulted in the production of modified surfactin in which the seventh and final aa had changed from a leu to a cys.

A – Overview of the module swapping scheme.

B – Products of the modified (red arrow) and un-modified (blue arrow) synthetase systems. Arrows identify the side chain of the aa which was modified and are colour coded to match the boxed modules that were responsible for their incorporation into surfactin. Figure adapted from [13].

As discussed in Section 1.3.1, elucidation of the crystal structure of a phe activating A-domain [25] has allowed determination of specificity conferring residues and proposition of a “non-ribosomal code” [29, 35-39, 86]. This, in turn, has allowed an additional strategy for NRPS modification by targeted alteration of specific residues within an A-domain, to alter or relax its specificity and affect incorporation of a novel aa. Modification of existing A-domains by site-directed mutagenesis (SDM) is an alternative strategy for generation of novel NRPS derived molecules. Eppelmann and colleagues, for example [86], altered the specificity of two A-

domains within the surfactin biosynthesis system using SDM. The specificity of an L-glu activating A-domain was altered to L-gln by targeted modification of putative specificity conferring residues, resulting in an A-domain which activated L-gln as efficiently as its wild type progenitor activated L-glu. They also altered the specificity of an asp-activating A-domain to asn, but this time the new enzyme exhibited a 10-fold decrease in aa activation activity compared to wild type. Analysis of the products of the modified enzyme systems showed unexpected levels of wild type peptide product as well as the anticipated modified peptide. The pioneering work outlined in the previous two paragraphs showed that manipulation of NRPS to produce novel bio-active molecules was possible, but also indicated that efficient production of novel molecules would require consideration of more than just A-domain specificity.

1.9.2 – Combinatorial biosynthesis using the daptomycin synthetase system

More recently domain swapping experiments have been conducted by Cubist Pharmaceuticals, Inc., using the daptomycin biosynthetic cluster of *Streptomyces roseosporus*[16, 87]. These experiments aimed at modifying the structure of daptomycin in four different ways. First the terminal synthetase gene, *dptD* encoding an NRPS composed of two modules was deleted and complemented *in trans* by chromosomal integration of equivalent genes from related antibiotic synthetase systems [88, 89]. Each of the genes used to complement the *dptD* deletion specified a different aa in the final module but were nonetheless able to complement the mutation, resulting in production of modified daptomycin derivatives with a new aa incorporated (Figure 1.15). Remarkably, the modified synthetase system generated by this simple subunit exchange strategy resulted in strains producing modified daptomycin with up to 69% the efficiency of wild type.

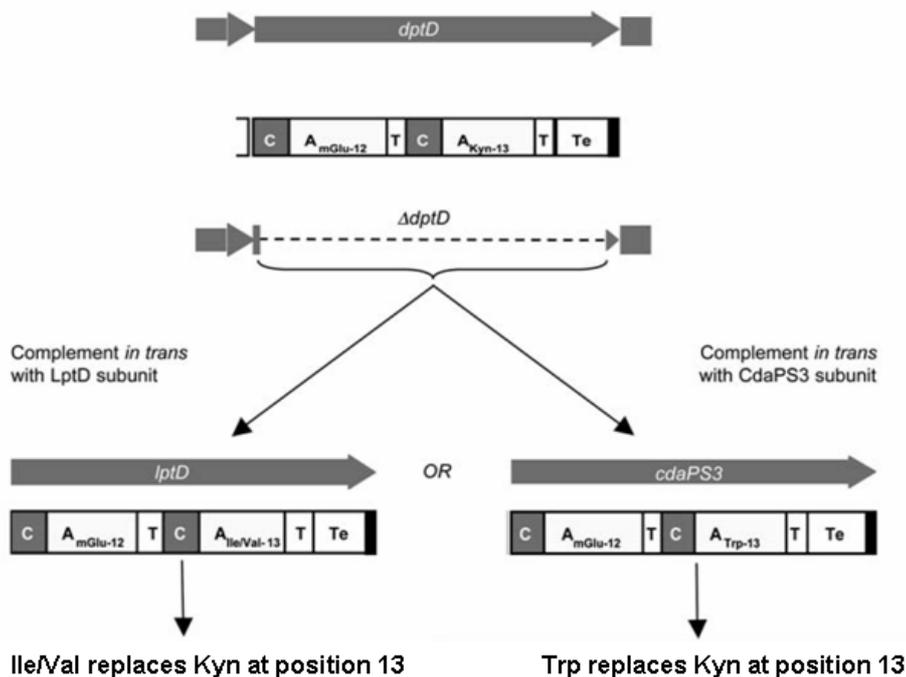


Figure 1.15 – Subunit exchange to produce novel daptomycin derivatives

The *dptD* gene encodes the final two modules of the daptomycin synthetase system in *Streptomyces roseosporus*. It is a bi-modular gene with the domain order CAT-CATTE. Deletion of this gene followed by complementation with bi-modular termination genes from related synthetase systems (*iptD* or *cdaPS3*) results in production of daptomycin in which the final aa has changed from kynurenine (kyn) to ile, val or trp. Figure reproduced from [16].

A second strategy was one of module exchange within the gene *dptBC* [88, 90]. The third and sixth modules of *dptBC* have the domain order CATE with A-domains specifying alanine and serine respectively. Hybrid genes were created in which module three was replaced by a duplicate of module six and vice versa (Figure 1.16). In each case the expected product was generated with good yield.

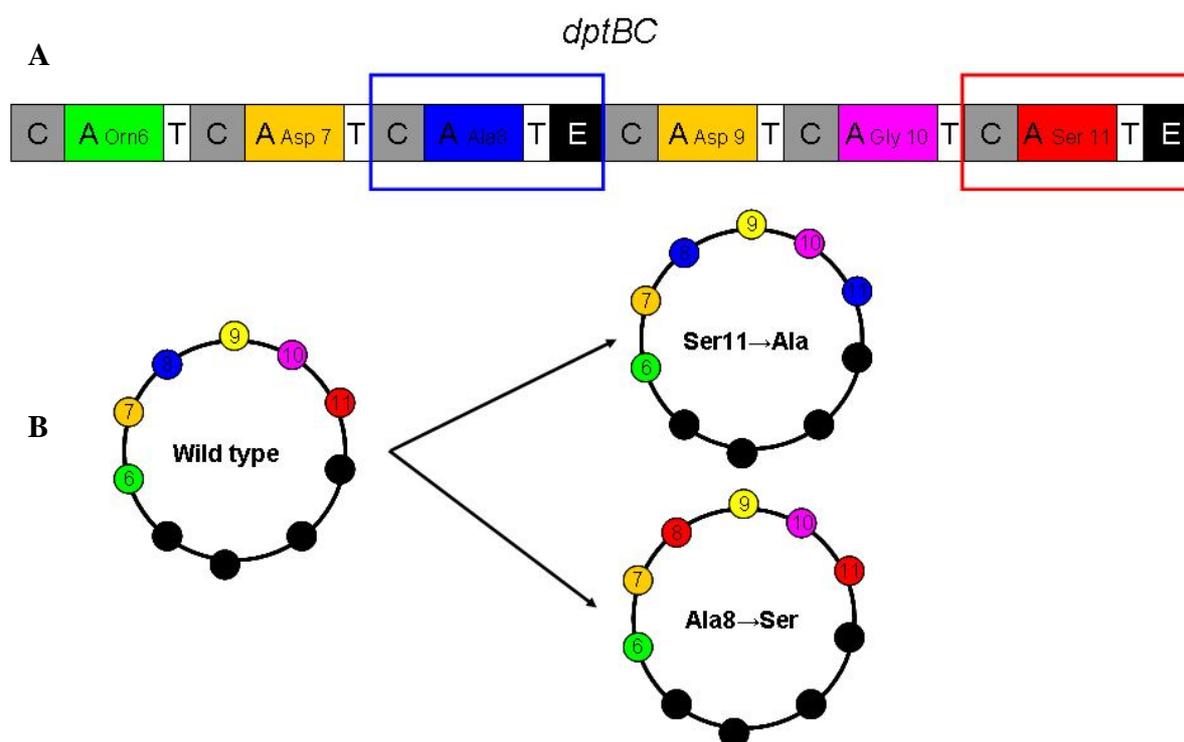


Figure 1.16 – Module swapping to produce novel daptomycin derivatives in *Streptomyces roseosporus*

Replacement of the ala 8 module of *dptBC* with a duplicate of the ser 11 module, and vice versa, resulted in production of the expected modified daptomycins with good yield.

A – A schematic representation of the *dptBC* gene, A-domains are colour coded to match the structures in the bottom panel.

B – Simplified representation of the cyclic portion of wild type and modified daptomycins with the aas incorporated by the DptBC protein colour coded, aas incorporated by other synthetase proteins are shown in black.

The third modification technique was deletion of *dptI*, a C-MT enzyme responsible for methylating the third carbon of the glutamic acid residue in daptomycin. As expected, deletion of *dptI* from the *S. roseosporus* genome resulted in a strain producing non-methylated daptomycin. In addition to these genetic modification strategies, the natural propensity for *S. roseosporus* to incorporate variant fatty acid side chains into daptomycin further increases the potential of this system for combinatorial biosynthesis. Using different combinations of the three modifications outlined above, it is theoretically possible to derive 72 different daptomycin derivatives. This combinatorial approach has been tested, and each of the 72 possible derivatives was indeed detected [91]. The minimal inhibitory concentration of each derivative with *Staphylococcus aureus* was determined and in some cases found to be

comparable to that of wild type daptomycin. The experiments conducted using the daptomycin system are a significant step toward unlocking the potential of NRPS enzymes for combinatorial biosynthesis, and demonstrate how a number of modification strategies can be combined to generate a diverse library of compounds with novel biological activity.

A fourth modification technique, that of module swapping, was also investigated using the daptomycin system [15]. These experiments were particularly interesting as they identified flexible linker regions within the final module of the synthetase system that were tolerant of aa substitutions, insertions and deletions. Perhaps the most interesting outcome of this research was development of a hypothesis that an intimate link exists between the TE-domain of the final module and its preceding T-domain. Disruption of the native T-TE linkage in the final module was found to result in a completely non-functional synthetase system while disruption of the C-A and the A-T linkers were well tolerated. This result is consistent with a variety of structural and biochemical data indicating that T-domains must make specific spatial and temporal interactions as they shuttle intermediates between catalytic centres in an enzyme [11, 53, 92-94].

1.9.3 Development of an NRPS tool box

As far as biotechnological applications are concerned, the ultimate goal of NRPS research is to develop a set of rules which can be used to freely recombine the diverse set of catalytic activities present in various NRPS modules, domains and enzymes. In just over a decade the field has progressed from proof of principle in the surfactin synthetase system to practical application in the daptomycin system; however there is still no consensus on exactly how NRPS manipulation should be conducted. What is becoming clear is that NRPS domains can no longer be considered simply as a set of discrete enzymes. Data from structural, biochemical and manipulation experiments has shown that precisely coordinated interaction between modules and domains occurs, and that in order to realise the potential applications of these fascinating enzymes this interaction must be examined in detail.

1.10 – Aims of this study

The initial aim of this work was to characterise the NRPS proteins involved in synthesis of the siderophore (iron carrier) pyoverdine in *Pseudomonas syringae*. Although the ultimate goal of this characterisation attempt was to provide raw material for subsequent domain swapping experiments, it also afforded an opportunity to investigate the biosynthesis and physiological significance of siderophores in *P. syringae*.

Characterisation of NRPS modules involved in pyoverdine synthesis in *P. syringae* enabled them to be used in module swapping experiments with the aim of probing C-domain specificity for the peptide donor in a condensation reaction. It was hypothesised, based on previous experiments [85], that C-domain specificity for the donor substrate in a condensation reaction might result in inactivity following module swapping. This hypothesis was tested in the experiments described in Chapter 5 of this thesis.

The results from the module swapping experiments described in Chapter 5 provided evidence against the initial hypothesis of C-domain donor specificity and prompted an investigation into the role of interdomain interactions in NRPS function. Subsequent experiments, described in Chapter 6, aimed to test the hypothesis that NRPS T-domains enter into specific interactions with other domains and that disruption of these interactions is a possible cause of inactivity in recombinant NRPS proteins. A recently characterised pigment-synthesising NRPS, BpsA, was employed as a reporter gene in domain swapping and directed evolution experiments which were conducted in order to determine key residues involved in T-domain interactions with other domains in BpsA.

BpsA proved to be a powerful reporter gene for directed evolution experiments by virtue of its ability to synthesise a readily detectable pigment in *E. coli*. Synthesis of this pigment is dependent on BpsA being activated by 4'-PP attachment to its T-domain, a reaction which can only be catalysed by PPTase capable of recognising the BpsA T-domain. The dependence of BpsA activity on PPTase activation suggested that this enzyme might also serve as a reporter for characterisation and discovery of PPTase enzymes. The experiments described in Chapter

7 of this thesis aimed to derive systems for PPTase characterisation and discovery based on the unique properties of BpsA.

Chapter 2 – Material and methods

2.1 – General reagents and materials

Unless otherwise noted, all chemicals were obtained from Sigma Aldrich and were of analytical quality. ^{32}P Pi was obtained from Perkin Elmer. HisbindTM protein purification resin and buffers were obtained from Novagen. AlkphosTM kit, CDP StarTM detection reagent and Hybond N+TM for Southern blotting were obtained from Amersham. EnzcheckTM Pyrophosphate detection kit was obtained from Invitrogen. IPTG (isopropyl β -beta-D-thiogalactoside) and Xgal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) were obtained from Bioline. EDDHA (ethylene-diamine-di-hydroxyphenylacetic acid) was kindly supplied by Professor Ian Lamont of Otago University.

2.2 – Enzymes

Restriction endonucleases, calf intestinal phosphatase (CIP), yeast inorganic pyrophosphatase (YIPP), T4 DNA polymerase and PhusionTM high fidelity polymerase were obtained from New England Biolabs. Bioline RedTM Taq polymerase mastermix was obtained from Bioline. T4 DNA Ligase was obtained from Fermentas.

2.3 – Bacterial strains and plasmids

2.3.1 – Bacterial strains

The strains of *E. coli*, *P. syringae*, *P. aeruginosa* and *P. putida* used in this study are given in Tables 2.1-2.3, respectively.

Table 2.1 – E.coli strains used in this study

Strain	Relevant characteristics*	Source
DH5 α	<i>supE44 DlacU169</i> ($\text{\O}80$ <i>lacZ</i> DM5) <i>hsdR17</i>	Invitrogen
DH5 α - λ pir	<i>supE44 DlacU169</i> ($\text{\O}80$ <i>lacZ</i> DM5) <i>hsdR17</i> λ (pir)	ATCC
BL21	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B m_B)</i> λ (DE3)	Novagen

* Standard genotype abbreviations are used, a list of these can be found at http://openwetware.org/wiki/E._coli_genotypes#Nomenclature_.26_Abbreviations

Table 2.2 – *P. aeruginosa* strains used in this study

Strain	Relevant characteristics	Source
PAO1	Wild-type	ATCC
PAO1Δ <i>pvdD</i>	<i>pvdD</i> deletion mutant, Pvd ⁻	[85]
Δ <i>pvdD</i> , Pa(Thr)Thr	PAO1Δ <i>pvdD</i> with Pa(Thr)Thr inserted at <i>attB</i>	This study
Δ <i>pvdD</i> , Pa(fOHOrn)Thr	PAO1Δ <i>pvdD</i> with Pa(fOHOrn)Thr integrated at <i>attB</i>	This study
Δ <i>pvdD</i> , Ps(Asp)Thr	PAO1Δ <i>pvdD</i> with Ps(Asp)Thr integrated at <i>attB</i>	This study
Δ <i>pvdD</i> , Ps(Thr)Thr	PAO1Δ <i>pvdD</i> with Ps(Thr)Thr integrated at <i>attB</i>	This study
Δ <i>pvdD</i> , Ps(Thr)Ser	PAO1Δ <i>pvdD</i> with Ps(Thr)Ser integrated at <i>attB</i>	This study

Table 2.3 – *P. syringae* strains used in this study

Strain	Relevant characteristics	Source
1448a	Wild-type	[95]
1448aΔ <i>pspph1911</i>	<i>pspph1911</i> deletion mutant, Pvd ⁻	This study
1448aΔ <i>pspph1923</i>	<i>pspph1923</i> deletion mutant, Pvd ⁻	This study
1448aΔ <i>pspph1924</i>	<i>pspph1924</i> deletion mutant, Pvd ⁻	This study
1448aΔ <i>pspph1925</i>	<i>pspph1925</i> deletion mutant, Pvd ⁻	This study
1448aΔ <i>pspph1926</i>	<i>pspph1926</i> deletion mutant, Pvd ⁻	This study
1448aΔ <i>acsA</i>	<i>acsA</i> deletion mutant, Acr ⁻	This study
1448aΔ <i>pspph1924</i> Δ <i>acsA</i>	Deletion of <i>pspph1924</i> and <i>acsA</i> , Pvd ⁻ / Acr ⁻	This study
1448aΔ <i>pspph1924</i> Δ <i>acsA</i> Δ <i>ybt</i>	Deletion of <i>pspph1924</i> , <i>acsA</i> and <i>ybt</i> , Pvd ⁻ / Acr ⁻	This study

2.3.2 – Plasmids

Table 2.4- Plasmids used in this study

Plasmid	Relevant characteristics	Reference
pET28a(+)	<i>LacI^f</i> , T7prom, Kan ^R , ColE1ori	Novagen
pCDFDuet1	<i>LacI^f</i> , T7prom, spec ^R , CDFori	Novagen
pDM4	Suicide vector for gene deletion-Chl ^R , <i>sacB</i> ,	[96]
pDM4::1911	pDM4 + <i>pspph1911</i> mutagenic fragment	This study
pDM4::1923	pDM4 + <i>pspph1923</i> mutagenic fragment	This study
pDM4::1924	pDM4 + <i>pspph1924</i> mutagenic fragment	This study
pDM4::1925	pDM4 + <i>pspph1925</i> mutagenic fragment	This study
pDM4::1926	pDM4 + <i>pspph1926</i> mutagenic fragment	This study
pDM4::acsA	pDM4 + <i>acsA</i> mutagenic fragment	This study
pDM4::ybt	pDM4 + <i>ybt</i> mutagenic fragment	This study
pAT23	pET28a(+) with AT domains form <i>Pspph1923</i> (NdeI-XhoI)	This study
pAT241	pET28a(+) with AT domains form module 1 of <i>Pspph1924</i> (NdeI-HindIII)	This study
pAT242	pET28a(+) with AT domains form module 2 of <i>Pspph1924</i> (NdeI-HindIII)	This study
pAT251	pET28a(+) with AT domains form module 1 of <i>Pspph1925</i> (NdeI-EcoRI)	This study
pAT252	pET28a(+) with AT domains form module 2 of <i>Pspph1925</i> (NdeI-HindIII)	This study
pAT261	pET28a(+) with AT domains form module 1 of <i>Pspph1926</i> (NdeI-HindIII)	This study
pAT262	pET28a(+) with AT domains form module 2 of <i>Pspph1926</i> (NdeI-HindIII)	This study
pXAT241	pAT241 with alternative AT domain boundaries	This study
pXAT251	pAT251 with alternative AT domain boundaries	This study
pSW196	Integration proficient vector allowing arabinose regulated expression of genes introduced into chromosomal <i>attB</i> site-Gm ^R	[97]
pSMC	pSW196 based staging vector allowing introduction of foreign CA domains	This study

	into second module of <i>pvdD</i> via SpeI and NotI restriction sites	
Pa(Thr)Thr	pSMC + CA domains from second module of <i>pvdD</i>	This study
Pa(fOHOrn)Thr	pSMC + CA domains from first module of <i>pvdD</i>	This study
Ps(Asp)Thr	pSMC + CA domains from second module of <i>pspph1924</i>	This study
Ps(Thr)Thr	pSMC + CA domains from first module of <i>pspph1925</i>	This study
Ps(Thr)Ser	pSMC + CA domains from second first module of <i>pspph1925</i>	This study
pET::BpsA	pET28a + <i>bpsA</i> (NdeI-HindIII)	This study
pSX::BpsA	pSX + <i>bpsA</i> (NdeI-HindIII)	This study
pBPSA1	pET28a based staging vector allowing introduction of foreign T-domains into <i>bpsA</i> using substitution scheme one	This study
pBPSA2	pET28a based staging vector allowing introduction of foreign T-domains into <i>bpsA</i> using substitution scheme two	This study
pLBPSA3	pCDFduet based staging vector allowing introduction of foreign T-domains into <i>bpsA</i> using substitution scheme three	This study
oBPSA	pBPSA1 + T-domain from <i>bpsA</i>	This study
oDhbF	pBPSA1 + T-domain from first module of <i>B. subtilis dhbF</i>	This study
oPvdD	pBPSA1 + T-domain from first module of <i>P. aeruginosa pvdD</i>	This study
oPST	pBPSA1 + T-domain from second module of <i>P. syringae pspph1926</i>	This study
oEntF	pBPSA1 + T-domain from <i>E. coli entF</i>	This study
oEntB	pBPSA1 + T-domain from <i>E. coli entB</i>	This study
nBPSA	pBPSA2 + T-domain from first module of <i>B. subtilis dhbF</i>	This study
nEntF	pBPSA2 + T-domain from <i>E. coli entF</i>	This study
nPvdD	pBPSA2 + T-domain from first module of <i>P. aeruginosa pvdD</i>	This study
nPST	pBPSA2 + T-domain from second module of <i>P. syringae pspph1926</i>	This study
slBPSA	pBPSA3 + T-domain from <i>bpsA</i> , expresses WT BpsA protein	This study
slPvdD	pBPSA3 + T-domain from first module of <i>P. aeruginosa pvdD</i>	This study
slEntF	pBPSA3 + T-domain from <i>E. coli entF</i>	This study
slEntB	pBPSA3 + T-domain from <i>E. coli entB</i>	This study
slDhbF	pBPSA3 + T-domain from first module of <i>B. subtilis dhbF</i>	This study
pSX	pUCP22 backbone with expression region from pUCX, allows IPTG regulated expression of his-Tagged proteins in <i>Pseudomonas</i> species	This study
pUCP22	Source of pSX vector backbone – Ori1600, Gm ^R	[98]
pUCX	<i>Lac^I</i> , tacProm, Amp ^R , ColE1ori	[99]
pSX::231	pSX + AT domains from <i>pspph1923</i>	This study
pSX::23a	pSX + A domain from <i>pspph1923</i>	This study
pSX::251	pSX + AT domains from first module of <i>pspph1925</i>	This study
pSX::252	pSX + AT domains from second module of <i>pspph1925</i>	This study
pNEO	pUCX in which the original colE1 ori has been replaced with a p15 ori	This study
pNEO::KT	pNEO + <i>P. putida</i> KT2440 PPTase	This study
pNEO::sfp	pNEO + <i>sfp</i> (PPTase from <i>B. subtilis</i>)	This study
pNEO::pcpS	pNEO + <i>pcpS</i> (PPTase from <i>P. aeruginosa</i>)	This study
pET::KT	pET28a + <i>P. putida</i> KT2440 PPTase	This study
pET::sfp	pET28a + <i>sfp</i>	This study
pET::pcpS	pET28a + <i>pcpS</i>	This study
pCDF::KT	pCDF + <i>P. putida</i> KT2440 PPTase	This study
pCDF::sfp	pCDF + <i>sfp</i>	This study
pCDF::	pCDF + <i>pcpS</i>	This study
pSX::1923	pSX + entire <i>Pspph1923</i> gene, for complementation of gene deletion	This study
pSX::1925	pSX + entire <i>Pspph1925</i> gene, for complementation of gene deletion	This study
pUCP22::PvdS*	pUCP22 with <i>pvdS</i> gene of <i>P. aeruginosa</i> ligated into EcoRI site of MCS	Lamont lab

*Constructed in laboratory of Iain Lamont, Otago University, New Zealand

2.4 – Oligonucleotide primers

Primers were designed using Vector NTI® and ordered from IDT custom oligonucleotide service. Primers were reconstituted to a final concentration of 100 µM in 1 x TE buffer for

storage at -20 °C. Working stocks were prepared by dilution to 10 µM with sterile ddH₂O. The names and sequences of all primers used in this study are given in Table 2.5.

Table 2.5 – Primers used in this study

Restriction sites are indicated in bold, homology regions for overlap PCR are underlined.

Primer name and function	Sequence (5'→3')
<u>Primers for mutagenic fragments</u>	
1911_Lup_XhoI	GGGGCT CGAGG CCTTGCACACGCCGGATC
1911_Ldown	AGACGCACTGTCCAGCGTATCG
1911_Rup_hom	<u>CGTCGATACGCTGGACAGTGCCTAGCGTGCCGGTGTTCGTG</u>
1911_Rdown_XbaI	<u>CCCCTCTAGACTTGCGGTCAAGCTTGCC</u>
1923_Lup_EcoRI	GGGGGG ATT CGATGTACGTGCGTAACTACGGC
1923_L_dwn	GTTTCATGGGTCGATCCTTGC
1923_Rup_Hom	<u>ATTCAGCAAGGATCGACCCATGAACGCATGACCGGGACCACCG</u>
1923_Rdown_BamHI	GGGGGG ATC CTTCGGTTTCAGCACGGCCG
1924_Lup_XhoI	GGGGCT CGAG ATGACCGGGACCACCGCT
1924_Ldown	AGCCAGACCGGACAAATCAATG
1924_Rup_Hom	<u>CAGCATTGATTTGTCCGGTCTGGCTCTGAAGACCGAGCTCCGTGC</u>
1924_Rdown_XbaI	<u>CCCCTCTAGAGTCCAGCGCATGGTCACTTTC</u>
1925_L_Up_SacI	GGGGGG AGCT CACAGTTGCTTGACTCCGTTA
1925_L_down	GCACTTCCTGATCGACATAA
1925_R_up_hom	<u>GAGCGTTATGTTCGATCAGGAAGTGCAAACACTGCCGATTACAT</u>
1925_R_dwn_HindIII	GGGGGA AGCT TTTTTTTCAATATCGGACAAGC
1926_Lup_XhoI	GGGGCT CGAG ATGGACAAGTCTGCAGCAGAGC
1926_Ldown	TTCGATGTCGGCTGGCGC
1926_Rup_Hom	<u>CTCGGTTGCGCCAGCCGACATCGAAGTGGGGCTTATCGAAACCCG</u>
1926_Rdown_XbaI	<u>CCCCTCTAGATCACACAAGCCGCTCGGC</u>
acsA_Lup_XhoI	GGGGCT CGAG ATGAACTTCACTTCACTCGCCG
acsA_Ldown	CTCCAGAAACAGCGCTATGCC
acsA_Rup_Hom	<u>AAAGGGCATAGCGCTGTTTCTGGAGTGGCTTGAGCGTAACGGCA</u>
acsA_Rdown_XbaI	GGGGT CTAG ATCAGGTTGCACGGCGCAA
YBKO_Lup_XhoI	CCCC CTCG AGATGTACACATCGACCAGAAAAGC
YBKO_Ldown	CTGATACAGATGACAGCC
YBKO_Rup_hom	<u>TGGAGTCGGCTGTCATCTGTATCAGCTGGCTGAGCAACCCAG</u>
YBKO_Rdown_XbaI	<u>CCCCTCTAGATCATGAACCACCTCCGCT</u>
<u>Inserts for AT/A domain overexpression constructs</u>	
231_fwd_NdeI	GGGG CATAT GGAGCGCGCTCAGTTGCTGGATA
231_rev_XhoI	GGGG CTCG AGTCATGCGCCAAATTCCTCCA
23a_rev_XhoI	GGGG CTCG AGTCACTGGCTTGAATCGGGCGCTG
241_fwd_NdeI	GGGG CATAT GCCGTTGTTACGCCTCGGTG
241_rev_HindIII	GGGG AAAGCT TTTCATGCCACCGAGGCCAGCTG
242_fwd_NdeI	GGGG CATAT TGGCACGTCTCGGCGAGTTGA
242_rev_HindIII	GGGG AAAGCT TCTAAGCAATCAAATTATC
251_fwd_NdeI	GGGG CATAT GAGCGTCGAAGAACGGCAAGA
251_rvw_EcoRI	GGGG GAATT CTCACGCAGCCTGGGCGAAAGCT
252_fwd_NdeI	GGGG CATAT TGCCGAACAGCAGCAGATTGT
252rev_HindIII	GGGG AAAGCT TTTCAGATACCCGCCATTTTCATTC
261_fwd_NdeI	GGGG CATAT TGCGACTTGCCGATCAGCCG
261_rev_HindIII	GGGG AAAGCT TTTCATGCCACCGAGGCCAACTG
262_fwd_NdeI	GGGG CATAT TGGAACAGCAGCAGATTGTTTCGCG
262_rev_HindIII	GGGG AAAGCT TTTCACATGAAATCGGCCAGCTCAC

EntF_A_Fwd_NheI	GGGGGCTAGCGATATTATGCTGCCAGGTGAGT
EntF_A_Rev_HindIII	GGGGAAGCTTTCACCTTCAGTTCAGGCAACGGTA
<u>Inserts for AT-domain overexpression constructs with alternative domain boundaries (pXAT)</u>	
x241fw_ndeI	GGGGCATATGATTCCCGAAATGCTTGCG
x241rev_HindIII	GGGGAAGCTTTCACAACGGCAGCGAGCCACT
x251fw_NdeI	GGGGCATATGTGCCGCGACGCCAAC
x251rev_HindIII	GGGGAAGCTTTCACAGCGGCTGGCTGCGGTC
<u>Primers for construction of pSMC staging vector</u>	
pvdDmod1_Fwd_EcoRI	CCCCGAATTC CATATGCAAGCACTCATAGAGAAGGTG
pvdDmod1_Rev_SpeI	CCCCACTAGTCAATCCCTGGGCGAACGC
pvdDTTE_Fwd_NotI	CCCCGCGGCCGCGTCCGAGCAGGCCTATCGAGCGC
pvdDTTE_Rev_SacI	CCCCGAGCTCTCAGCGCCCGGCACGCTCCAGGGCGGCCCGGA
<u>Primers for introduction of CA domains into pSMC</u>	
pvdD-CA1_Fwd_XbaI	CCCCCTCTAGAGGTGTCAATCTCTTCGAG
pvdD-CA2_Fwd_XbaI	GGGGTCTAGAACGACGGATGCGGTCTCGACGA
pvdD-CA 1+2_Rev_NotI	CCCCGCGGCCGCGATCCGGTTGCGGCAACGCCTGC
CA Ser(Thr)_Fwd_XbaI	CCCCCTCTAGAGGGCAGGGCAATGCTGCG
CA Ser(Thr)_Rev_NotI	CCCCGCGGCCGCGTCTGGTGTGCGCAGGGC
CA-Thr(Thr)_Fwd_XbaI	CCCCCTCTAGAAATCTCTACGGGGTACACCGA
CA-Thr(Thr)_Rev_NotI	CCCCGCGGCCGCGATCCGGCGCGGGCAGCGCCTTG
CA-Thr(Asp)_Fwd_XbaI	CCCCCTCTAGAACGACGGATGCGGTCTCGACGATACCGCTTGC
	CGATCGGCAGGAAGATATCTACCCGCTGTGCG
CA-Thr(Asp)_Rev_NotI	CCCCGCGGCCGCGATCCGGCGCGGGCAGCGCCTTG
<u>Primers for creation of pBPSA1</u>	
pBPSA1_Lup	CCCCCATATGACTCTTCAGGAGACCAGCGT
pBPSA1_Ldwn	GGGGGAATTCGGTACCCTCCGTCTCCGTGCGCGGGG
pBPSA1_Rup	CCCCAAGCTTGCTAGCGAGCGCGAGGTCGCCAGGAGTCC
pBPSA1_Rdwn	CCCCCTCGAGTCACTCGCCGAGCAGGTAGC
<u>Primers for substitution of T-domains into pBPSA1</u>	
oBPSAT_Fwd	GGGGCATATGGGTACCAAGGAGATCGCGGCGGTCTG
oBPSAT_Rev	GGGGGCTAGCCAGGCGGGCCAGCTTCTCGA
oPvdDT_Fwd	CCCCCATATGGGTACCCAGCGCATCGCAGCGATC
oPvdDT_Rev	CCCCGCTAGCCAATCCCTGGGCGAACGC
oDhbFT_Fwd	GGGGCATATGGGTACCGAGATATTGTGTGACTTGTGTTGCAG
oDhbFT_Fwd_Rev	GGGGGCTAGCAAGATGGGCAGCGAGTCCGGCTACT
oEntFT_Fwd	CCCCCATATGGGTACCACGATTATCGCCGCGGCA
oEntFT_Rev	CCCCGCTAGCAATAATCGTTGCCAGTTT
oEntBT_Fwd	CCCCCATATGGGTACCAGCGCGCTGCGTGAGGTG
oEntBT_Rev	CCCCGCTAGCGAGTAGCTTCCACCAGGC
oPsT_Fwd	CCCCCATATGGGTACCCAACGTCTTGCCGCGTTG
oPsT_Rev	CCCCGCTAGCCATGAAATCGGCCAGCTC
<u>Primers for creation of pBPSA2</u>	
pBPSA2_Lup	CCCCCATATGACTCTTCAGGAGACCAGC
pBPSA2_Ldwn	CCCCGAATTCGGTACCGGGGCGCTCGACGAGCTC
pBPSA2_Rup	CCCCGGTACCCCGCTAGCCGCGAGGTCGCCAGGAG
pBPSA2_Rdwn	CCCCAAGCTTTCACCTCGCCGAGCAGGTA
<u>Primers for substitution of T-domains into pBPSA2</u>	
nBPSAT_Fwd	CCCCCATATGGGTACCTTCGTGCCCCGCGCACG
nBPSAT_Rev	CCCCGCTAGCCTCCAGGCGGGCCAG
nPvdDT_Fwd	CCCCCATATGGGTACCTATCGAGCGCCCGGTAGC
nPvdDT_Rev	CCCCGCTAGCTTCCAATCCCTGGGCGAA
nEntFT_Fwd	CCCCCATATGGGTACCGGGCGTGCGCCGAAAGCG
nEntFT_Rev	CCCCGCTAGCATCAATAATCGTTGCCAG
nPsT_Fwd	CCCCCATATGGGTACCCACGTGCGGCCACGTACG
nPsT_Rev	CCCCGCTAGCACGCATGAAATCGGCCAG
<u>Primers for creation of pBPSA3</u>	

pBPSA3_Lup	CCCCGGATCCGATGACTCTTCAGGAGACCAGCG
pBPSA3_Ldown	AGCTAAGCTTAGCTATGCATTGACCTGGTCGGAGGCGG
pBPSA3_Rup	AGCTAAGCTTAGCTACTAGTCGCTTCGTCCGCCTGCACG
pBPSA3_Rdown	CCCCCTCGAGTCACTCGCCGAGCAGGTAGC
<u>Primers for substitution of T-domains into pBPSA3</u>	
slBPSAT_Fwd	AGCTCTGCAGAGCTCGTCGAGCGCCCCCTTCGTGCGCCCCGCGCACG
slBPSAT_Rev	AGCTTCTAGACTCCTGGGCGACCTCGCGCTCCAGGCGGCGGGCCAG
slPvdDT_Fwd	AGCTCTGCAGAGCTCGTCGAGCGCCCCATCGAGCGCCCGGTAGC
slPvdDT_Rev	AGCTTCTAGACTCCTGGGCGACCTCGCGTTCCAATCCCTGGGCGAA
slDhbFT_Fwd	AGCTCTGCAGAGCTCGTCGAGCGCCCCGATCGGGCCCCGCGGACT
slDhbFT_Fwd_Rev	GCTTCTAGACTCCTGGGCGACCTCGCGATCAAGATGGGCAGCGA
slEntFT_Fwd	AGCTCTGCAGAGCTCGTCGAGCGCCCCGGGCGTGCGCCGAAAGCG
slEntFT_Rev	AGCTTCTAGACTCCTGGGCGACCTCGCGATCAATAATCGTTGCCAG
slEntBT_Fwd	AGCTCTGCAGAGCTCGTCGAGCGCCCCCTGCCAGCACCTATCCC
slEntBT_Rev	AGCTTCTAGACTCCTGGGCGACCTCGCGGAGAGTAGCTTCCAC
<u>Primers for PPTase amplification</u>	
PP_PPTase_fwd_NdeI	GGGGCATATGAACACACTTCCCCGCTGC
PP_PPTase_rev_SalI	GGGGGTCGACTCAGACGCTGACCAGGCTCA
pcpS_fwd_NdeI	GGGGCATATGCGCGCCATGAACGACCG
pcpS_rev_SalI	GGGGGTCGACTCAGGCGCCGACCGCCACCA
sfp_fwd_NdeI	GGGGCATATGAAGATTTACGGAATTTA
sfp_rev_SalI	GGGGGTCGACTTATAAAAAGCTCTTCGTACG

2.5 – Media

Unless otherwise noted all media components were dissolved in ddH₂O and sterilised by autoclaving.

LB

LB was obtained as a premixed powder from Sigma or HiMedia and when reconstituted contained: 5 g/L yeast extract, 10 g/L bacto tryptone, 10 g/L NaCl.

King's B medium

20 g/L bacto peptone (DIFCO), 1 % v/v glycerol, 1.5 g/L K_2HPO_4 , 6.1 mM $MgSO_4$. $MgSO_4$ was added from a sterile 1 M stock after autoclaving to avoid precipitation.

M9 minimal medium

5 x M9 salts were obtained as a premixed powder from Sigma and when reconstituted contained 64 g/L Na_2HPO_4 , 15 g/L KH_2PO_4 , 2.5 g/L NaCl, 5 g/L NH_4Cl . M9 Minimal medium was prepared immediately prior to use and contained (per litre): 778 mL sterile ddH_2O , 200 mL 5 x M9 salts, 2 mL sterile 1M $MgSO_4$, 20 mL sterile 20 % w/v glucose, 100 μ l sterile 1 M $CaCl_2$.

TYM medium

2 % bacto-tryptone, 0.5 % yeast extract, 100 mM NaCl, 10 mM $MgCl_2$. $MgCl_2$ added after autoclaving from a sterile 1 M stock.

ZYP5052 autoinduction medium

ZY Base: 10 g/L N-Z Amines AS, 5 g/L yeast extract

50 x 5052: 250 g/L glycerol (=200 mL/L), 25 g/L D-glucose, 100 g/L α -lactose

20 x NPS: 66 g/L $(NH_4)_2SO_4$, 136 g/L KH_2PO_4 , 142 g/L Na_2HPO_4

ZYP5052 was prepared immediately prior to use and contained (per litre): 929 mL ZY base, 20 mL 50 x 5052, 50 mL 20 x NPS, 1 mL 1M $MgSO_4$.

2.5.1 – Media supplements

All antibiotic stocks were made to 1000 x the media concentration indicated for *E. coli* in Table 2.5. IPTG stocks were prepared to a final concentration of 100 mg/mL. With the following exceptions, media supplements were dissolved in ddH_2O and filter sterilised using a 0.22 μ M filter: Tetracycline was dissolved in 100 % ethanol, rifampicin was dissolved in 50 % methanol with sufficient NaOH to facilitate solubility, chloramphenicol was dissolved in 50 % ethanol, Xgal stocks were prepared to a final concentration of 24 mg/mL in 100 % dimethyl formamide, EDDHA stocks were prepared to a final concentration of 50 mg/mL in ddH_2O

with sufficient NaOH to facilitate solubility and filter sterilised using a 0.22 μM filter. Unless otherwise noted final media concentration of antibiotics is as listed for each organism in Table 2.6.

Table 2.6 – Antibiotics used in this study

Antibiotic	E. coli ($\mu\text{g/mL}$)	P. syringae ($\mu\text{g/mL}$)	P. aeruginosa ($\mu\text{g/mL}$)	P. putida ($\mu\text{g/mL}$)
Ampicillin	100	100	-	-
Kanamycin	50	-	-	-
Spectinomycin	50	-	-	-
Chloramphenicol	34	-	-	-
Gentamycin	20	20	100	50
Tetracycline	10	10	100	50
Rifampicin	-	100	-	-
Carbenicillin	-	100	300	100

2.6 – Growth and maintenance of bacteria

LB was used for routine growth and maintenance of all bacterial strains. Growth temperatures used were typically: 37 °C for *E. coli* and *P. aeruginosa*, 30 °C for *P. putida* and 28 °C for *P. syringae*. For liquid cultures aeration was provided by shaking at 200-300 rpm. Medium-term maintenance was achieved by storing agar plates containing bacteria at 4 °C for up to two weeks. For long-term maintenance of strains, an aliquot of overnight culture was mixed 1:1 with 80 % glycerol and stored at -80 °C. To ensure maintenance of plasmids with antibiotic resistance markers, the appropriate antibiotic(s) were added to liquid and solid media.

2.7 – Routine molecular biology

2.7.1 – PCR protocols

Standard PCR reactions

Where PCR products were to be used in downstream cloning applications, amplification was conducted using the NEB Phusion™ high fidelity polymerase kit. For screening purposes and

splicing by overlap, PCR Bioline Biomix Red™ Taq polymerase mastermix was employed. Standard reactions set up according to manufacturer's directions were usually sufficient for good amplification. In rare instances, systematic variation of the concentration of one or all of the following components was required to achieve amplification: MgCl₂, DNA template, primers, DMSO.

Overlap PCR for mutagenic fragment generation

For the generation of mutagenic fragments, two separate standard PCR reactions were carried out to amplify ~400 b.p from the 5' and 3' end of the gene to be deleted. Following cleanup 100-200 ng of each of these fragments was added to a standard 50 µL Biomix red reaction without primer or genomic template addition. Splicing of the two fragments, mediated by a 25 b.p homology region on one of the primers was then carried out by conducting three primerless cycles – as indicated in the overlap amplification protocol – after three cycles primers were added to the concentration recommended by the manufacturer and the remainder of the amplification protocol continued.

Amplification protocols

The following thermocycler protocols were used for amplification using Phusion or Biomix red. In rare cases it was necessary to systematically vary one or all of the cycle parameters to achieve amplification.

Phusion PCR protocol

98° C 60 s

98° C 15 s

72° C 30 s (-1° C /cycle)

72° C 30 s per kilobase

10 cycles

98° C 15 s

62° C 30 s

72° C 30 s per kilobase

25 cycles

72° C 5 min

12° C on hold

Biomix red PCR protocol

95° C 5 min

95° C 30 s

56 ° C 30 s

72° C 1 min per kilobase

10 cycles

95° C 30 s

52 ° C 30 s

72° C 1 min per kilobase

20 cycles

72° C 10 min

12° C on hold

Overlap PCR protocol

95° C 5 min

95° C 30 s

50° C 30 s

72° C 1 min per kilobase

3 cycles – no primers

Primers added to reaction held at 95 °C

95° C 30 s

56 ° C 30 s

72° C 1 min per kilobase

10 cycles

95° C 30 s

52 ° C 30 s

72° C 1 min per kilobase

20 cycles

72° C 10 min

12° C on hold

2.7.1 – Isolation purification and manipulation of DNA

Isolation and purification

Small-scale preparation of plasmid DNA was carried out using either the Zymogen Zyppy™ or Qiagen Minispin™ kit. Preparation of genomic DNA was achieved using a Qiagen DNeasy™ Kit. Deproteinisation and buffer exchange of PCR products, restriction digests and ligation mixtures was carried out using Zymogen Zymospin™ columns. In all cases the manufacturer's protocol was followed. Where necessary, DNA samples were quantified and assessed for purity using a Nanodrop™ spectrometer or by agarose gel electrophoresis.

Restriction digests

Unless otherwise noted, restriction digests were carried out according to the manufacturer's directions. Before use in ligations, digests were heat inactivated and cleaned using a Zymospin™ column.

Ligations

Ligations were typically set up using a 1:6 molar ratio of vector:insert with enzyme and buffer concentration, as specified in the manufacturer's directions. Total DNA concentration was kept below 10 ng/μL. For transformation by electroporation, ligations were cleaned using a Zymospin™ column. Negative controls containing no insert DNA were set up and transformed in parallel to assess vector self-ligation and residual uncut vector DNA.

2.7.2 – Preparation and transformation of competent cells

2.7.2.1 – Preparation and transformation of chemically competent *E. coli*

This protocol is based on the method described in [100].

Reagents

TFBI: 30 mM KOAc, 50 mM MnCl₂, 100 mM KCl, 10 mM CaCl₂ 15 % w/v, glycerol. pH adjusted to 5.8 with 0.2 M acetic acid, solution sterilised using a 0.22 μM filter.

TFBII: 10 mM Na-MOPS pH7.0, 10 mM KCl, 75 mM CaCl₂ 15% w/v glycerol. Solution sterilised using a 0.22 µM filter.

Preparation

The desired strain was streaked onto TYM agar and incubated overnight at 37 °C. A single colony was then inoculated into 3 mL TYM broth and incubated at 37 °C/200 rpm for 12-16 h. 1 mL of this overnight was then used to inoculate 40 mL of TYM in a baffled, sterile 250 mL conical flask. The resulting culture was grown at 37 °C/200 rpm until an OD 600 of 0.35-0.4 was reached at which point the culture was transferred into a sterile 50 mL tube and placed on ice for 5 min. Cells were then collected by centrifugation (2500 rcf, 4 °C, 10 min) and supernatant decanted off. The resulting cell pellet was resuspended in 2 mL ice-cold TFB I by gentle pipetting before addition of a further 38 mL ice-cold TFB I. The cells were then kept on ice for 2 h and again harvested by centrifugation (2500 rcf, 2 °C, 10 min). The resulting cell pellet was resuspended in 4 mL ice cold TFB II and 100 µL aliquots distributed into pre-chilled 1.5 mL microfuge tubes on ice. Cell aliquots were then snap-frozen using liquid nitrogen or a metal tube block cooled to -80 °C. Frozen aliquots of cells were stored at -80 °C until needed. When appropriate for plasmid maintenance, antibiotics were added to TYM broth and agar.

Transformation

Cell aliquots were removed from storage and placed on ice until thawed. ~ 10 ng Plasmid DNA or up to 200 ng ligation mixture was then added in a volume not exceeding 1/10th that of the cell aliquot. Following addition of DNA, the tube was flicked gently to mix and kept on ice for 30 min. Cells were then heat shocked by placing in a water bath at 42 °C for 2 min. Following heat shock cells were placed immediately on ice for 5 min, before the addition of 9 vols sterile LB broth. Transformations were then incubated at 37 °C/200 rpm for 45-60 min to recover. After recovery cells were plated on LB agar containing appropriate antibiotics. For plasmids aliquots of 100 µL were plated. For ligations an aliquot of 100 µL was plated followed by collection of remaining cells by centrifugation (13,000 rcf, 20 s). The supernatant was then decanted leaving ~ 100 uL residual medium, in which the cell pellet was resuspended and then plated.

2.7.2.2 – Preparation and transformation of electrocompetent *E. coli*

This protocol is based on the method described in [101].

Reagents

GYT: 0.25 % bacto tryptone, 0.125 % yeast extract, 10 % v/v glycerol. Sterilise by autoclaving.

Preparation

The desired strain was streaked onto LB agar and incubated overnight at 37 °C. A single colony was then inoculated into 25 mL LB broth and incubated at 37 °C/200 rpm for 12-16 h. 10 mL of the overnight culture was then used to inoculate 500 mL LB medium in a baffled, sterile 2000 mL conical flask. The resulting culture was grown at 37 °C/200 rpm until an OD 600 of 0.35-0.4 was reached, at which point the culture was transferred into 10 sterile 50 mL tubes and placed on ice for 15-30 min with occasional swirling. Cells were then collected by centrifugation (1000 rcf, 30min, 2 °C) and the pellets resuspended in a total volume of 500 mL ice cold sterile ddH₂O by gentle pipetting. Following this first washing step cells were again collected by centrifugation (1000 rcf, 30min, 2 °C) and the resulting pellet resuspended in 250 mL ice-cold sterile 10 % v/v glycerol. Cells were again pelleted by centrifugation (1000 rcf, 30min, 2 °C) and the resulting pellet resuspended in 125 mL ice cold sterile 10 % v/v glycerol. Cells were again collected by centrifugation and the pellet resuspended in 500 µL ice cold sterile GYT. The OD 600 of a 1/100 dilution of the cell mixture was then determined and GYT added to give a final concentration of 2-3 x 10¹⁰ cells/mL (OD 600 1.0 = 2.5 x 10⁸ cells/mL). 40 µL aliquots were then distributed into pre-chilled 1.5 mL microfuge tubes on ice. Cell aliquots were then snap-frozen using liquid nitrogen or a metal tube block cooled to -80 °C. Frozen aliquots of cells were stored at -80 °C until needed. When appropriate for plasmid maintenance, antibiotics were added to LB broth and agar for growth steps.

Transformation

Cell aliquots were removed from storage and placed on ice until thawed. 50 ng ligation/40 µL cell mixture was then added in a volume not exceeding 1/10th that of the cell aliquot.

Following addition of DNA, the tube was flicked gently to mix and the contents transferred to an ice cold, sterile 2 mm gap electroporation cuvette. Cells were then electroporated (2.5 kV, 25 μ F, 100 Ω) and 10 vols. SOC broth was immediately added and the mixture transferred to a sterile 15 mL centrifuge. Cells were then incubated at 37 °C/200 rpm for 1 h before plating on LB medium containing appropriate antibiotics. When likely transformation efficiency was unknown, 100 μ l of neat transformation as well as 100 μ L of a 1/10 and 1/100 dilution were plated to ensure single colonies were obtained.

2.7.2.2 – Preparation and transformation of electrocompetent *Pseudomonas*.

This protocol is based on the method described in [102] and was used for the preparation of electrocompetent *P. aeruginosa*, *P. syringae* and *P. putida*.

A 50 mL culture of the desired strain was grown until stationary phase under the appropriate conditions. For each competent cell aliquot to be prepared 6 mL culture was aliquoted into 4 x 1.5 mL microfuge tubes. Cells were then pelleted by centrifugation (13,000 rpm, 30 s) and each pellet washed twice in 1 mL sterile 300 mM sucrose. After the final wash step the four cell pellets were resuspended in a final volume of 100 μ L sterile 300 mM sucrose and transferred to a sterile 2 mm gap electroporation cuvette. For transformation, 50 ng replicative or up to 500 ng non-replicative (KO) plasmid was added in a volume not exceeding 5 μ L and the cuvette gently flicked to mix. Cells were then electroporated (2.5 kV, 25 μ F, 200 Ω) and 1mL LB broth immediately added. Following transfer to a sterile 15 mL tube, cells were allowed to recover for 1 h (strain appropriate growth conditions) before plating 100 μ L aliquots on LB agar containing appropriate antibiotics. For maximum efficiency of transformation all preparation steps were conducted at room temperature [102].

2.7.2.3 – Identification of recombinant clones

Colony PCR

The first step for identification of clones harbouring desired inserts on a plate arising from transformation of a ligation was colony PCR. Standard biomix red PCR reactions with a final volume of 15 μ L were set up on ice for each clone to be screened. Where possible, a primer

targeting plasmid sequence was used in combination with a primer targeting insert sequence to avoid non-specific amplification. Colonies to be screened were picked using a sterile pipette tip and streaked onto a numbered plate containing appropriate antibiotics. The residual bacteria on the tip were then transferred into a PCR reaction and released by gentle swirling. Reactions were then subjected to the Biomix red amplification protocol described in Section 2.7.2. Typically, two controls were included in each screen these were: a reaction in which a clone from a vector-only control plate was used as template and a reaction inoculated with a pipette tip that had been touched to a region of the vector + insert agar plate not containing any colonies. The former assessed non-specific amplification from plasmid and genomic DNA sequences, the latter amplification of residual PCR product spread onto the surface of a plate. Following amplification 7 μ L aliquots of each reaction were analysed by agarose gel electrophoresis and 3 mL overnight cultures set up for clones showing amplification of the correct sized product.

Confirmation by restriction digest

Plasmid DNA from PCR positive clones was prepared from overnight cultures as described in 2.7.1. Digests of plasmid DNA from each clone were then set up as described in Section 2.7.1, using a combination of enzymes that would generate a diagnostic fragment in insert positive clones. Following incubation, digest were assessed for liberation of diagnostic fragment by agarose gel electrophoresis and plasmid DNA from positives sent for sequence analysis.

Sequencing of DNA

Sequencing was performed by Macrogen inc. DNA samples and primers for sequencing were prepared according to the company's specifications of concentration and purity. Sequence quality was assessed using Contig Express™ and insert identity confirmed by alignment of sequences against a template obtained from genbank.

2.7.3 – In frame deletion of genes from the chromosome of Ps1448a

Primers for generation of mutagenic fragment were designed using VectorNTI™ such that the resulting 800 b.p overlap PCR product would contain 400 b.p from each end of the gene, and lack the internal region. These fragments were designed to maintain the original reading frame

of the gene in order to avoid polar effects following chromosomal integration. Mutagenic fragments were restriction digested and ligated into cut pDM4 vector following the protocols outlined in Section 2.7.1. Ligations were transformed into chemically competent DH5 α - λ pir cells as described in 2.7.2.1 and recombinant clones identified as described in 2.7.2.3.

Once a mutagenic plasmid for deletion of a target gene had been obtained, this was delivered to cells using the electroporation protocol described in Section 2.7.2.2. Primary integration events were selected on plates containing 70 μ g/mL Chl and 100 μ g/mL Rif and presence of the mutagenic plasmid confirmed by colony PCR as described in 2.7.2.3. The second recombination event was then forced by growing integrants overnight in LB medium containing 5 % w/v sucrose and plating aliquots on LB medium containing 10 % w/v sucrose. After development, 20-40 colonies were patched onto LB medium with and without 70 μ g/mL Chl. Colonies which had reverted to Chl sensitivity were taken to be double crossover events. Double crossovers were subjected to an initial screen for presence of the mutagenic fragment using colony PCR (Section 2.7.2.3). Genomic DNA was then prepared from positive clones and again assessed the presence of the truncated gene and the absence of intervening sequence by standard PCR using Biomix red. Final conformation was obtained by southern blotting using an Amersham AlkphosTM kit according to the manufacturer's protocol. Analysis by Southern blot was carried out for Δ 1926 and Δ 1924.

2.8 – Siderophore characterisation

2.8.1 – CAS agar assays for iron uptake

One hundred mL Chromeazurol S (CAS) dye for the detection of siderophores [103] was made by dissolving 60.5 mg CAS powder (Sigma) in 50 mL distilled water. To this 10 mL of a 1 mM solution of FeCl₃ was added. The entire solution was then poured slowly with stirring into 40 mL distilled water containing 72.9 mg dissolved HDTMA (Sigma) and autoclaved to sterilise. To make agar plates, freshly autoclaved KB agar was cooled to 60 °C before adding 1 part CAS dye to 9 parts medium. Plates were immediately poured, and at this point exhibited a dark green colour. Strains were inoculated into dried CAS plates by picking a large colony with a sterile 100 μ l pipette tip and piercing the tip approximately 5 mm into the surface of the

agar plates. Duplicate plates were then incubated upside down at 28 °C for 24 h. After 24 h incubation in the 28 °C condition, one of the plates was removed from the incubator and maintained at 22 °C. Plates were photographed with minimal exposure to temperature change at 24, 48 and 72 h. The entire assay was repeated three times; results presented are from a single assay and are representative of all repeats.

2.8.2 – CAS media assays for iron uptake

Strains were inoculated in triplicate into 200 µl KB media in a 96 well plate to an initial OD₆₀₀ of 0.1, with outer wells filled with sterile H₂O to minimize evaporation. Replicate plates were then covered but not sealed and incubated for 24 h at 28 °C or 22 °C with shaking. The next day cells were pelleted by centrifugation (4000 g, 15 min) and 150 µl of supernatant was transferred to fresh wells in a flat bottomed 96 well plate. To each well 30 µl of CAS dye (prepared as described above) was added using a multi-channel pipette. Plates were immediately placed into the plate reader and OD 655 values recorded every 5 min for 50 min, then again at 65 min and 125 min.

2.8.2 – EDDHA Inhibitory Concentration (IC₅₀) assays

A 2-fold serial dilution series of KB media containing from 200-0.195 µg/mL of the iron chelator EDDHA was established in 96 well plates. Strains were inoculated in quadruplicate to an initial OD 600 of 0.1 from cultures synchronised by sub-inoculation over two nights, giving a final volume of 125 µl per well. Unsealed plates were then incubated for 24 h at 28 °C or 22 °C with shaking. Wells were diluted 1:1 with KB in order to be within the linear range of the plate reader, and OD 600 values were measured. For each temperature the assay was repeated twice with consistent results.

2.8.3 – Ps1448a pathogenicity tests in *Phaseolus vulgaris*

Single colonies from fresh 48 h KB agar plates were picked using a sterile hypodermic needle. Strains were then inoculated into snap bean pods (*Phaseolus vulgaris*) by piercing the surface of the bean approximately 5 mm. Each strain was inoculated in triplicate together with a WT positive control. Bean pods were then placed in a sealed humid container or alternatively, for on plant assessment, pods were left attached to parental plants growing indoors at 20-25 °C.

Results were recorded every 24 h. Development of water soaked lesions similar to those of WT strain was taken as a positive result. The assay was repeated in triplicate.

2.8.4. – Achromobactin purification

The protocol for achromobactin purification was adapted from that described in [104].

Reagents

Solvent A: 9:1:10 v/v methanol:H₂O:ethyl acetate

Solvent B: 9:1 v/v methanol:H₂O

Procedure

Two hundred mL of M9 minimal medium, in which succinic acid was the carbon source, was inoculated with 10 mL pyoverdine null *P. syringae* from a stationary phase culture grown in the same medium. The resulting culture was grown for 72 h (22 °C, 200 rpm) before removal of cells by centrifugation (5000 rcf, 30 min). The supernatant was then sterilised by passing through a 0.22 µm filter and then the volume reduced to 20 mL by rotary evaporation (temperature not exceeding 45 °C). 180 mL methanol was then added, at which point salt from the culture medium precipitated out of solution. Precipitate was removed by centrifugation (12,000 rcf, 20 min) followed by filtration (0.45 µm). The solution was then mixed 1:1 with ethyl acetate and 100 mL of the resulting solution applied to a glass chromatography column containing 40 cc silica beads pre-equilibrated with solvent A. 100 mL Solvent A was then applied to the column, followed by 100 mL solvent B. The eluate from the solvent B step was captured in 10 mL fractions. Siderophore activity of the fractions was then assessed by adding 30 µL CAS reagent to a 150 µL aliquot of each fraction and incubating for 10 min at room temperature. The fraction which resulted in the greatest discolouration of the CAS dye was then reduced in volume to 2 mL by rotary evaporation (temperature not exceeding 40 °C) and 1 mL of the solution removed. The remaining 1 mL was evaporated to dryness and resuspended in 1 mL ddH₂O. Both of these 1 mL samples were then sent to the Centre for Protein Research at the University of Otago for MS analysis.

2.8.5 – Pyoverdine purification

Pyoverdine purification was achieved using the method described in [105]. Pyoverdine containing culture supernatant was prepared as described in Section 2.8.4 except an achromobactin null *P. syringae* was used. The pH of the resulting 200 mL culture supernatant was adjusted to 6.0 using concentrated HCl. Approximately 40 mL wet Amberlite XAD-4 resin, which had been previously activated according to the manufacturer's directions, was then added to the acidified culture supernatant. The mixture was then shaken for ~ 90 min at 200 rpm, after which the beads were discernibly green, indicating pyoverdine adsorption. The supernatant was then discarded and the beads washed five times with 200 mL ddH₂O by shaking at 200 rpm for 15 min. After this the beads were washed with 500 mL ddH₂O (5 min, 200 rpm), then 500 mL 15 % v/v methanol (5 min, 200 rpm). Pyoverdine was then removed from the beads by shaking with 100 mL 50 % v/v methanol (200 rpm, 2h) and the resulting solution freeze-dried. Purified pyoverdine was then resuspended in 1 mL ddH₂O and, following confirmation of siderophore activity by CAS assay, sent to the Centre for Protein Research at the University of Otago for MS analysis.

2.8.6 – Assessment of pyoverdine production in wild type and recombinant *P. aeruginosa* strains

Qualitative assessment of pyoverdine production on agar plates

Strains were streaked onto KB agar plates with and without 0.4 % arabinose and incubated for 16 h at 37 °C. Plates were then photographed under natural and UV light.

Quantitative assessment of pyoverdine production

Synchronised overnight cultures were used to inoculate triplicate cultures for each strain in a 96 well plate to a uniform density in a final volume of 200 µl. Following 16 h growth (37 °C 200 rpm) cells were pelleted by centrifugation (4000 rcf, 10 min) and supernatant removed to fresh wells. Pyoverdine concentration was analysed by measuring the optical density at 405 nm and/or emission at 450 nm after excitation at 403 nm.

2.8.7 – Determination of intracellular pyoverdine levels by flow cytometry

Triplicate cultures for each strain were established in a 96 well plate as described above. The growth medium used was KB supplemented with 0.4 % arabinose. 50 µl from each replicate was then collected and pooled in a single 1.5 mL microfuge tube. Cells were then washed twice with 1 mL sterile PBS by centrifugation and resuspension. Following the second washing step cell suspensions were analysed using an LSRII flow cytometer. Excitation was with a violet laser (405nm) and emission detected with a 450/50 bandpass filter. A total of 5000 events were collected for each condition and events graphed against relative fluorescence.

2.9 – Protein expression and purification

2.9.1 – Protein expression

2.9.1.1 – Expression under standard conditions

For expression under standard conditions, an overnight culture (37 °C/200rpm) of the expression strain was grown in LB broth containing appropriate antibiotics. An expression culture using the same medium was inoculated with this overnight to a final OD 600 of 0.1 grown at 37 °C/200 rpm until an OD 600 of between 0.6 and 1.0 was reached. At this point IPTG was added to a final concentration of 0.5 mM to induce expression. Expression was allowed to proceed for 4-6 h before collection of cells by centrifugation (400 rcf, 4 °C, 15 min) for solubility analysis by SDS PAGE and/or purification.

2.9.1.2 – Betaine sorbitol expression protocol for improving solubility

For improved solubility of recombinant NRPS proteins, the following modifications were applied to the standard expression protocol described in Section 2.9.1.1: 1 M D-sorbitol and 2.5 mM glycine betaine were included in the overnight starter and expression cultures. The overnight culture was grown at 37 °C/200 rpm, the expression culture was grown at 18 °C/200rpm. Under these conditions it took between 24 and 36 h for an expression culture to

reach an OD 600 of 0.6-1.0, at which point expression was induced as previously described. Expression was allowed to proceed for a further 24-48 h at 18 °C/200 rpm before harvest of cells.

2.9.1.3 – Expression in *Pseudomonas* hosts

For expression from *Pseudomonas* strains, the following amendments were made to the standard expression protocol. Growth of starter cultures was conducted using LB broth and strain appropriate incubation conditions. Growth of expression cultures was conducted at 25 °C. A final concentration of 5.0 mM IPTG was used to induce expression. Expression was allowed to proceed for 16-24 h before harvest of cells.

2.9.1.4 – Co-expression of molecular chaperone proteins

Protein expression from a molecular chaperone background was conducted as for standard expression except after reaching an OD 600 of 0.6-1.0; cultures were placed at 18 °C/200 rpm for 45-90 min before addition of IPTG to induce expression. Expression was allowed to proceed for an additional 26-24 h at 18 °C/200 rpm before harvest of cells. Table 2.7 outlines the combinations of molecular chaperone proteins expressed by each chaperone strain used in this study.

Table 2.7 – Chaperone strains 1-4[106]

Strain	Plasmids	Chaperone proteins expressed
1	pBB530 + pBB535	dnaK, dnaJ, grpE
2	pBB540 + pBB535	dnaK, dnaJ, grpE, clpB
3	pBB540 + pBB542	dnaK, dnaJ, grpE, clpB, groESL high concentration
4	pBB540 + pBB550	dnaK, dnaJ, grpE, clpB, groESL low concentration

2.9.1.5 – Auto induction protocol

An overnight starter culture of the desired strain was grown in LB medium containing appropriate antibiotics and 0.4 % w/v D-glucose. This starter culture was then used to inoculate ZYP5052 medium expression culture containing appropriate antibiotics. The volume of starter culture used to inoculate was 1/50 the amount of ZYP5052 used in the expression culture. The expression culture was then incubated at 16 °C/250 rpm until an OD 600 of over 2.5 was reached, at this point expression levels of target protein were determined by SDS

PAGE analysis. If expression was deemed sufficient for subsequent purification, the culture was harvested by centrifugation (4000 rcf, 4°C, 15 min).

2.9.2 – Protein purification by Ni-NTA affinity chromatography

2.9.2.1 – Cell lysis and fraction separation

For protein purification, cell pellets were gently resuspended in 1 x Hisbind™ Binding buffer on ice and lysed by either French press or sonication. The amount of buffer used to resuspend pellets was 1/20 of the expression culture volume. For lysis by French press three passages at 40,000 psi were conducted, the French press chamber was chilled to 2 °C prior to lysis. For sonication, resuspended cells were placed in an ice bath and sonicated at 70% maximum output, with 50% duty cycle and 10 s bursts until consistency indicated sufficient lysis (usually 2-3 min). Following lysis, soluble and insoluble fractions were separated by centrifugation (17,000 rcf 2°C, 20 min) and the soluble fraction stored on ice until purification.

2.9.2.2 – Standard Ni-NTA affinity chromatography

Unless otherwise noted, purification of 6his-tagged proteins was achieved using a Novgen Hisbind™ Ni-NTA chromatography kit, according to the manufactures directions, a final volume of 1.5 mL settled resin was used for purification unless otherwise noted. Eluted protein was captured in 1.5 mL fractions for analysis by SDS PAGE. The first time any protein was purified, flow-through from lysate application and all washing steps was also analysed by SDS PAGE to evaluate binding efficiency and wash stringency.

2.9.2.3 – Purification of denatured proteins followed by refolding treatment

For refolding of denatured proteins, inclusion bodies containing target protein were dissolved in binding buffer containing 8 M urea and 1 mM β-mercaptoethanol, passed through a 0.45 μM filter and applied to a 1.5 mL Hisbind column pre-equilibrated with the same buffer. The recommended washing steps were then carried out using buffers with the addition of 8 M urea and 1 mM β-mercaptoethanol. A final washing step of 3 mL 1 x binding buffer with 8 M urea and 1 mM β-mercaptoethanol was then conducted. After washing a linear concentration

gradient of 8-0 M urea was applied to the column over a period of 24 h using the apparatus illustrated in Figure 4.15. The total flow through volume used to achieve this gradient was 200 mL. After urea removal, refolded protein was eluted from the column using standard elution buffer according to the manufacturer's (Novagen Hisbind kit) directions and captured in 1.5 mL fractions for analysis by SDS PAGE.

2.9.2.4 – Special amendments to purification procedures

For purification of BpsA it was found that protein binding to the column was weak resulting in protein elution during the second more stringent wash step. In order to achieve protein of good purity, it was necessary to fully saturate the column's binding capacity by applying an excess of cell lysate. To achieve this, a 2000 mL expression culture was set up using the auto induction protocol described in Section 2.9.1.5. The entire soluble fraction (100 mL) resulting from this culture was then applied to a 7 mL pre-equilibrated Hisbind resin. Flow-through from the soluble fraction was collected and again applied to the column. After this, purification was conducted according to the manufacturer's protocol except the volume of the first wash step was tripled and the second wash step eliminated.

For purification of PPTase enzymes it was found that proteins precipitated immediately once eluted from the column. To prevent this 10 % v/v glycerol was added to all His bind buffers and all buffers kept ice cold during the purification. Following elution all fractions were immediately desalted and stored at -20 °C in 50 mM sodium phosphate buffer (pH 7.8) containing 50 % w/v glycerol.

2.9.2.5 – Buffer exchange, storage and quantification of recombinant proteins

Buffer exchange was achieved using GE Healthcare Hitrap™ desalting columns, according to the manufacturer's protocol. Storage buffer for proteins was 50 mM Tris-Cl 50% w/v glycerol, the pH of which was the same as that of the assay solution in which proteins were to be employed. Proteins were stored at -20 °C and kept on ice at all times when in use. Quantification of recombinant proteins was achieved using Biorad DC Kit according to the manufacturer's directions.

2.9.3 – SDS-polyacrylamide gel electrophoresis

SDS-PAGE was performed using 12 % SDS-polyacrylamide gels prepared according to the method of Laemmli [107]. Gels were cast and run using the Biorad Protean II TM system, according to the manufacturer's directions. Preparation of samples, as well as gel staining and destaining, was carried out as described by Sambrook and Russell [108].

2.9.3.1 – Molecular weight markers

Two molecular weight markers for the estimation of protein size were employed in this study. Gels pictured in Chapter 4 used Fermentas broad range molecular weight marker while those pictured in Chapter 7 used NEB broad range molecular weight marker. The molecular weight of each band present in these markers is shown in Figure 1.1.

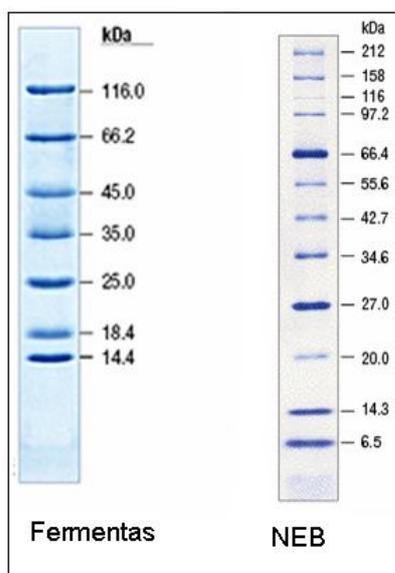


Figure 2.1 – Molecular weight markers used in this study

2.10 – A-domain specificity assays

2.10.1 – ATP/³²PPi exchange assays

These protocols are based on the original method described in [109].

Reagents

Stop solution: 3.5 % w/v perchloric acid, 100 mM PPi, 1.6 % w/v activated charcoal

PPi washing solution: 100 mM PPi, pH 8.0

Elution buffer: 0.3 M ammonia in 50 % v/v methanol

EA1 protocol

Triplicate 100 μ L reactions were set up for each substrate in 1.5 mL microfuge tubes. Each reaction contained 2.5 mM ATP, 5 mM aa substrate, 0.5 mM EDTA, 12.5 mM MgCl₂, 2.5 mM ³²PPi [0.0028 μ Ci/mM], 2.5 μ g enzyme, 25 mM tris pH 7.7. Enzyme was added last to initiate reactions. Triplicate no-enzyme and no-enzyme/no-substrate controls were also run. Reactions were incubated at 28 °C for 30 min before termination by addition of 400 μ L stop solution. Terminated reactions were mixed briefly and charcoal collected by application to a 0.45 μ M filter by syringe. Collected charcoal was then washed three times with 3 mL PPi washing solution and two times with 5 mL ddH₂O. ATP was then eluted from the charcoal by application of 6 mL elution buffer, which was collected in a scintillation vial containing 9 mL ddH₂O. Cerenkov counts were then measured directly from scintillation vials using a Wallac 1400 liquid scintillation counter.

EA2 protocol

Triplicate 100 μ L reactions were set up for each substrate in 1.5 mL microfuge tubes. Each reaction contained: 2 mM ATP, 5 mM aa substrate, 0.5 mM EDTA, 10 mM MgCl₂, 100 mM NaCl, 0.1 mM ³²PPi [7 μ Ci/mM], 2.5 μ g enzyme, 50 mM tris pH 7.7. Enzyme was added last to initiate reactions. Reactions were incubated at 28 °C for 30 min before termination by addition of 200 μ L stop solution. Terminated reactions were mixed briefly and charcoal collected by centrifugation (13,000 rpm 1 min). Charcoal was then washed two times with 1 mL ddH₂O. After the final wash step charcoal pellets were resuspended in 1 mL ddH₂O and transferred to scintillation vials containing 14 mL ddH₂O. Vials were left for 30 min to allow

charcoal to settle. Cerenkov counts were then measured directly from scintillation vials using a Wallac 1400 liquid scintillation counter.

2.10.2 – PPI release assay RA1

This assay is based on the method described in [110] and measures PPI liberated from ATP by and A-domain in the presence of an amino acid substrate. The rate of PPI release is indicative of substrate preference. PPI is detected in a two step enzymatic reaction. Firstly PPI is cleaved into two equivalents of Pi by inorganic pyrophosphatase (IP), Pi created in this fashion is then linked to 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) by purine nucleoside phosphorylase (PNP). The addition of Pi to MESG results in a shift of absorbance maximum from 330 to 360 nM. Absorbance at 360 nM is therefore an indirect measure of the amount PPI liberated from ATP by an A-domain.

End point assay

Reagents

Detection reagent: 0.5 mM MESG, 1 U PNP 0.03 U IP, 50 mM Tris-HCl (pH 7.7) in sterile ddH₂O.

Triplicate reactions for each substrate were established in a 96 well plate and contained, in a final volume of 100 μ L: 50 mM Tris-HCl (pH 7.7), 10 mM MgCl₂, 1 mM ATP, 5-20 μ g enzyme and 1.25 mM aa substrate. Reactions were then incubated for 60 min at 28 °C, after which 100 μ L detection reagent was added. Plates were then incubated at room temperature for 20 min before measurement of A₃₆₀ using a microplate reader. For PPI standard curve generation, 100 μ L detection reagent was added to mock reactions, without enzyme or substrate, containing a triplicate two fold serial dilutions series (40-2.5 μ M) of sodium pyrophosphate followed by incubation and A₃₆₀ measurement steps.

Continuous release assay

Triplicate reactions for each substrate were established in a 96 well plate and contained in a final volume of 100 μ L: 50 mM Tris-HCl (pH 7.7), 10 mM MgCl₂, 1 mM ATP, 5-20 μ g

enzyme, 0.25 mM MESG, 1 U PNP, 0.03 U IP and 2 mM aa substrate. Reactions were initiated by addition of enzyme and the plate incubated in a microplate reader set to take A_{360} measurements every 60 s.

2.10.3 – PPi release assay RA2

This assay is based on the method described in [111]. Like RA1 it detects A-domain activity by measuring levels of PPi released from ATP and then cleaved into two equivalents of Pi by IP. Addition of the detection reagents results in the formation of a green complex, the amount of which is proportional to Pi levels in the sample.

Detection Reagents

PPgreen reagent A: 4.2 % w/v sodium molybdate in 4 M sulfuric acid

PPgreen reagent B: 0.135 % w/v brilliant green dye in sterile ddH₂O

Triplicate reactions for each substrate were established in a 96 well plate and contained in a final volume of 100 μ L: 50 mM Tris-HCl (pH 7.7), 10 mM MgCl₂, 1 mM ATP, 5-20 μ g enzyme, 0.03 U IP and 1.25 mM aa substrate. Plates were then incubated for 30 min at 28 °C, after which reactions were terminated by addition of 20 μ L PPgreen reagent A per well. Plates were then tapped gently to mix, and 20 μ L PPgreen reagent B added to each well. Plates were again mixed by tapping and A_{650} measured immediately using a microplate reader. For PPi standard curve generation, mock reactions, without enzyme or substrate, containing a triplicate two-fold serial dilutions series (50 – 0.25 μ M) were used.

2.11 – Directed evolution Protocols

The protocols used for directed evolution were continually refined throughout the course of this study and the final optimised version of each aspect is given below. These protocols allow the generation of libraries of $1-5 \times 10^6$ insert containing clones from a single epPCR reaction. Libraries of this size were more than sufficient for recovery of multiple improved clones.

2.11.1 – Library generation

2.11.1.1 – Vector preparation

For preparation of vector for directed evolution, 16 µg of plasmid DNA (prepared as described in Section 2.7.1) was heated to 70 °C for 20 min to relieve super-coiling and then digested with 50 U of each restriction enzyme in a final volume of 400 µL. Buffer composition and incubation temperature was as recommended by the enzyme manufacturer. After 5 h incubation, an additional 20 U of each enzyme was added, and the digest incubated for a further 12-14 h. Following this, digests were heat inactivated (80 °C, 20 min) and purified using a 20 µg capacity zymospin® clean up column. Digested vector was eluted from the clean up column using 30 µL sterile ddH₂O and the concentration of the resulting solution determined. This was generally between 250-350 ng/µL. A 200 ng sample of vector was then analysed by agarose gel electrophoresis to ensure no degradation had occurred. Aliquots of 5 µL were then prepared and stored at – 20 °C.

2.11.1.2 – Determination of vector quality

Before being employed in directed evolution experiments the quality of prepared vector was assessed. To achieve this, the native T-domain of *bpsA* was amplified using Phusion® high fidelity polymerase and prepared for ligation as described in Section 2.11.1.4. A control ligation was then set up as described in 2.11.1.5, except with digested *bpsA* T-domain in place of epPCR. The ligation was then transformed into chemically competent cells (2.7.2.1) harbouring an activating PPTase and aliquots of the transformation mixture plated on pigment development agar (LB, 100 mM L-gln, 0.5 mM IPTG). Plates were then incubated for 12-14 h at 37 °C before induction of protein expression and as described in 2.11.2. Plates were then left at room temperature for 12-16 h before determination of transformation efficiency and percentage of colonies containing insert. Determination of insert percentage was achieved by counting blue and white colonies on an entire plate (containing at least 200 colonies). Blue colonies indicated correct ligation of the T-domain insert into the prepared vector, white colonies indicated vector self ligation or out-of-frame ligation. If the percentage of blue

colonies was 70 or higher, the vector was deemed of sufficient quality for use in directed evolution experiments.

2.11.1.3 – Error prone PCR

Error prone PCR was carried out using a Stratagene Mutazyme II® kit, according to the manufacturer's directions. Optimal error rate (as assessed by number of improved clones recovered) was achieved using 100 ng of purified PCR product as template per 50 µL reaction, with 30 amplification cycles. Template was prepared by amplification of the appropriate T-domain sequence using Phusion polymerase. Prior to thermocycling, reactions were divided into four to reduce the number of clonal mutations in the final library. Amplicon size and quality was assessed by agarose gel electrophoresis of a 3 µL aliquot.

2.11.1.4 – Insert preparation

50 µL epPCR reactions were purified using a 5 µg capacity zymospin® column, with elution achieved using 20 µL sterile ddH₂O. The entire eluent was then digested for 5 h with 30 U of each enzyme, in a final volume of 50 µL. Buffer composition and incubation temperature was as recommended by the enzyme manufacturer. Following digest, reactions were heat inactivated (80 °C, 20 min) and purified using 5 µg capacity zymospin® column (elution volume 20 µL, sdH₂O).

2.11.1.5 – Ligation

Ligations were set up as described in 2.7.1 with the following amendments: Total amount of DNA per ligation was 1.0-1.5 µg. Ligations were incubated overnight at 16 °C, after which an additional 2 U of ligase were added, and the reaction incubated for a further 8 h at 16 °C. The optimal molar ratio of vector to insert for achieving maximum ligation efficiency was found to be 1:6.

2.11.1.6 – Control reactions

One positive and two negative control reactions were run for each directed evolution experiment. For the positive control, the native T-domain of BpsA was amplified using high-fidelity polymerase, digested and ligated and transformed using the same conditions as for ep-

library generation. The purpose of the positive control was to give a reliable estimate of percent insert containing clones using the process described in Section 2.11.1.2. The first negative control reaction was set up in the same way, except using the T-domain to be evolved, amplified with high fidelity polymerase. Inclusion of this negative control was important to check for contamination of pcr reagents with WT *bpsA* DNA, which would result in a large number of false positives. The second negative control was a plasmid containing the original recombinant *bpsA* gene which the experiment aimed to improve. Colonies arising from transformation of this control were used as a reference point for determination of improved clones.

2.11.1.7 – Transformation and library storage

Prior to transformation, ligations were purified using a 5 µg capacity Zymospin column (elution volume 20 µL, sdH₂O). Purified ligations were then transformed into electrocompetent *E. coli* BL21 cells, harbouring a plasmid for expression of an activating PPTase. Electrocompetent cells were prepared and transformed as described in Section 2.7.2.2, with the following amendments: The volume of cells used was 60 µL per 50 ng ligation to be transformed. Cells and ligations were mixed on ice and then aliquoted into chilled cuvettes (80 µL/cuvette). Following electroporation, 1 mL SOC medium was immediately added to each cuvette. All cells for a single ligation were then pooled in a single 15 mL tube and recovered for 1 hr (37 °C, 200 rpm). Following recovery, cells were mixed 1:1 with sterile 80 % glycerol and 0.5 mL aliquots dispensed. These aliquots were then placed at -80 °C. A single aliquot was then thawed on ice and serial dilutions plated on pigment development agar (LB, 100 mM L-gln, 0.5 mM IPTG) containing appropriate antibiotics in order to determine the optimal plating volume for screening. Libraries stored as frozen transformation aliquots remain viable for at least 3 months. For evolution of oDhbF, libraries were transformed into chemically competent cells. The use of electrocompetent cells was subsequently adopted to improve transformation efficiency.

2.11.2 – First tier screening

For screening, cells were thawed on ice and plated on LB agar containing 100 mM L-gln and strain appropriate antibiotics. The volume of cells used per plate was adjusted so that ~ 10,000 insert containing clones would be present. After 12 h of incubation at 37 °C, expression was induced in plates. Induction was achieved by IPTG addition using the method described in Section 6.5.1. Plates were then incubated at ~ 20 °C and monitored for colour development. Colonies which developed colouration before negative controls were taken to be hits. These putative improved clones were picked from plates and small scale overnight growth cultures set up for preparation of glycerol stocks. Due to the high density of colonies on a single plate, it was often impossible to recover improved clones without risk of contamination. In such cases, clones were picked as accurately as possible using a toothpick and resuspended in 100 µL GYT. Single colonies were obtained by streaking this suspension on a portion of an agar plate. Subsequent induction of this plate allowed recovery of improved clones without contaminants. Improved clones were then grown overnight in LB medium supplemented with 0.4 % glucose and appropriate antibiotics. Duplicate glycerol stocks of improved clones were then prepared in a 96 well plate.

2.11.3 – Second tier screening

For second tier screening, up to 45 improved clones, along with a positive and negative control, were stored as duplicate glycerol stocks in a 96 well plate. From this plate duplicate overnight cultures were established in LB supplemented with 0.4 % glucose and appropriate antibiotics. Set up of overnight cultures was achieved in a 96 well plate using a 96-pin inoculating tool. Overnight cultures were grown for 16-20 h (37 °C 200 rpm) before transfer of 20 µL of each culture into a fresh well containing 130 µL LB supplemented with 115 mM L-gln, 0.6 mM IPTG and appropriate antibiotics. Assay plates set up in this fashion contained duplicate cultures for each clone from two separate overnights. Assay plates were then wrapped in foil and incubated for 6-24 h (18 °C, 200 rpm), depending on activity of improved clones. After incubation cells were collected at the corner of each well by centrifugation and OD 590 values determined using a microplate reader. Assays were repeated at least twice before determination of clones for sequence analysis.

2.12 – Indigoidine quantification

In vivo pigment synthesis efficiencies for recombinant BpsA proteins described in Section 6.6 required multiple assays using the protocol outlined in Section 2.11.3. Each of these assays was terminated at a different time point between 3 and 36 h. This was necessary due to the fact that pigment synthesis for more efficient proteins was found to saturate before activity of slower proteins was detectable. The values presented are derived as follows: quadruplicate values for sIBpsA (WT) and oBpsA were determined after 3 h, at which point pigment production was not saturated for WT. A separate quadruplicate assay was then run to obtain a value for sEntF by comparison to oBpsa after 12 h. Quadruplicate values for oPsT and nBpsa were then obtained in a separate assay by comparison to sEntF after 36 h. The activity of oDhbF and oEntF was too low to detect by measurement of supernatant OD 590 however these variants did produce discernable pigmentation after 48 h incubation on pigment development agar. The remaining variants did not produce discernable pigmentation, appearing identical to an empty plasmid control on pigment development agar. These variants were deemed completely inactive.

2.13 – Enzyme kinetics

2.13.1 – Activation of BpsA by 4'-PP attachment

Reactions for pre-activation of BpsA prior to use in kinetic assay (2.13.2) contained: 0.83 μ M BpsA, 1 μ M PP1183 (PPTase), 1 mM MgCl₂, 0.1 mM CoA and 50 mM sodium phosphate buffer pH 7.8. Reactions were incubated for 30 min at 28 °C.

2.13.2 – Determination of kinetic parameters for BpsA

For derivation of kinetic parameters a two-fold serial dilution series from 20-0.156 mM ATP was established in triplicate in a final volume of 100 μ L in a 96 well plate. Each well also contained: 4 mM L-gln, 1 mM MgCl₂ and 50 mM sodium phosphate buffer pH 7.8. Reactions were initiated by addition of 100 μ L activated BpsA solution and A590 readings taken every 10 s in a microplate reader. Maximum velocity values for each reaction were determined by

measuring the steepest gradient over six consecutive data points using Softmax Pro®. Derivation of kinetic parameters was achieved by analysis of data using the one site saturation regression function of SigmaPlot®.

2.13.3 – Determination of kinetic parameters for PPTases

Reactions were established for a duplicate two-fold serial dilution series from 25.0-0.78 μM CoA in a 96 well plate. In addition to CoA, each reaction contained, in a final volume of 200 μL : 4 mM L-gln, 1 mM ATP, 1.64 μM BpsA (inactive), 1.03 μM PPTase. BpsA was added last to initiate reactions and A590 readings taken every 10 s in a microplate reader. Average velocity of the pigment synthesis reaction between every three data points (20 s) was determined using Softmax Pro®. Velocity values were plotted using Microsoft Excel® and a maximum acceleration value determined from the linear portion of the resulting plot. Maximum acceleration values were analysed using Sigmaplot® to derive kinetic parameters. The units for the Vmax values derived from this analysis were “change in gradient per second”; the calculations used to convert these into “amount of BpsA activated per second” are outlined in Appendix 2.

2.14 – Screening of soil derived eDNA library for PPTase genes

The protocol used for eDNA library screening is essentially the same as that outlined in Sections 2.11.1.7 and 2.11.3 except electrocompetent BL21 cells did not contain an activating PPTase and instead contained the plasmid for expression of WT BpsA. Following recovery, plasmid DNA was prepared from hits and isolation of the pRSET plasmid containing the eDNA insert for sequencing was achieved by transformation into chemically competent *E. coli* cells followed by selection on medium containing ampicillin only. Inserts were sequenced using both T7 promoter and T7 terminator primers which anneal to plasmid sequences immediately upstream and downstream of the insert respectively.

2.15 –*In silico* manipulation using Vector NTI and DeepView

2.15.1 – Design of scheme three for seamless T-domain substitution into BpsA

Substitution scheme three (Section 6.5.3) achieves a seamless transition between BpsA sequence and the sequence of a substituted T-domain by utilising silent restriction sites that do not alter aa sequence of the protein arising from expression constructs. The program VectorNTI® (VNTI) was used to identify silent restriction site candidates and design primers for generation of swapping plasmid as follows:

- 1) A list of restriction enzymes that do not cut the vector (pCDFduet1), BpsA gene or T-domain inserts was generated using the restriction report function of VNTI.
- 2) The resulting enzymes were then saved as a subset in the restriction enzyme database of VNTI. Additional hybrid sequences which arise from the ligation of two separate restriction enzyme cuts producing compatible ends were entered into the database manually and were also saved in this subset.
- 3) Positions at which these restriction sites could be introduced were then determined using the mutagenesis function of VNTI. This resulted in the identification of a hybrid NsiI-PstI site upstream and a hybrid XbaI-SpeI site downstream of the region of T-domain substitution. Introduction of these hybrid sites did not alter aa sequence encoded at the splice point.
- 4) Primers were then designed to amplify the regions up and downstream of the T-domain of BpsA using the amplify selection function of VNTI. The products of amplification by these primers were cloned into pCDFduet to generate a staging vector.
- 5) For amplification of T-domains for substitution into the staging vector the amplify selection function was again used. The BpsA sequence between the silent splice points and the desired point of T-domain introduction was added manually to the 5' end of the up and downstream primers as illustrated in Figure 2.2. The exact location of the transition points between BpsA sequence and that of an introduced T-domain is illustrated in Figure 6.6.

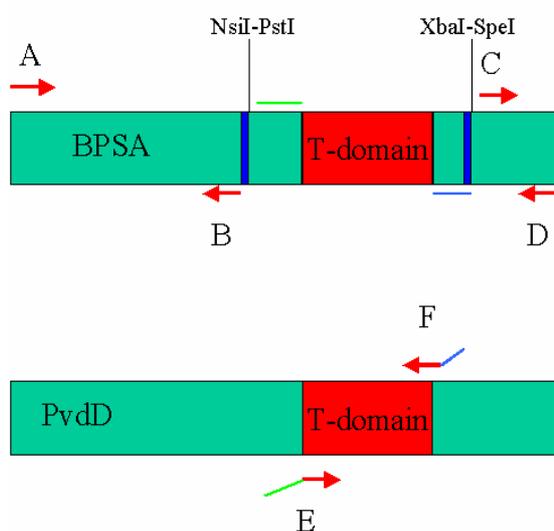


Figure 2.2 – Substitution scheme three cloning strategy

Amplification of the regions up and downstream of the BpsA T-domain using primers A+B and C+D respectively allowed generation of a construct in which the BpsA gene lacked its native T-domain and contained an NsiI and a SpeI site into which foreign domains could be introduced. Primers E and F for amplification of substitute T-domains (PvdD T-domain in this example) had 5' sequence elements added that would replace the missing BpsA sequence indicated by the blue and green lines. Ligation of T-domains into BpsA resulted in an upstream NsiI-PstI hybrid site and a downstream XbaI-SpeI hybrid site, indicated by dark blue bars. Introduction of these sites did not alter the aa sequence encoded at these points. The result of this is a seamless transition between BpsA sequence (light blue) and T-domain sequence (red).

2.15.2 – Structural modelling

The structural model of the T-TE interaction region of BpsA presented in Chapter 6 was generated using the first conformer of PDB file 2roqA (solution structure of the T and TE domains of *E. coli* EntF) as a template. Determination of suitability of this template and generation of an alignment was achieved using the template identification tool of the Swiss Model SIB server (<http://swissmodel.expasy.org/>). The resulting alignment was visualised in DeepView and subjected to preliminary visual assessment to ensure that conserved elements of the T-domain in the target and template structures occupied equivalent positions. The preliminary model was then submitted to the online server for refinement. Following refinement it was found that the loop region between the T and TE-domains was poorly fitted to the 2roq template, alternative conformations of this region were generated using the iterative magic fit, explore alternative conformation and energy minimisation functions of DeepView. Alternative fits were subjected to preliminary analysis using DeepView and those containing residues making clashes or buried charges not present in 2roq were discarded. This resulted in a three separate models in which the loop region between the T and TE-domains of BpsA was in a similar conformation to that seen in 2roq. These models were again submitted for refinement. The quality of the refined models was then assessed using the Qmean [112] online server. The model which returned the highest score from this assessment is presented in Chapter 6.

Chapter 3 – Siderophores in *Pseudomonas syringae*: Synthesis and physiological significance

3.1 – Summary

Pseudomonas syringae pv. *phaseolicola* 1448a (Ps1448a), the causative agent of bean halo blight, is a bacterium capable of occupying diverse biological niches. Under conditions of iron starvation Ps1448a synthesises siderophores for active uptake of iron. The primary siderophore of Ps1448a is pyoverdine, a fluorescent molecule that is assembled from aa precursors NRPS. Five putative pyoverdine NRPS genes in Ps1448a were identified and characterised *in silico* and their role in pyoverdine biosynthesis was confirmed by gene knockout. Creation of pyoverdine null Ps1448a enabled identification of a previously uncharacterised temperature-regulated secondary siderophore, achromobactin, which is not synthesised by NRPS and has a lower affinity for iron. Pyoverdine and achromobactin null mutants were characterised in regard to iron uptake, virulence and growth in iron limited conditions. Results of these studies show that although achromobactin has potential to contribute to fitness under iron-limiting conditions, these effects are masked by the presence of pyoverdine, which is a significantly more effective siderophore. Neither pyoverdine nor achromobactin appear to be required for Ps1448a to cause bean halo blight, indicating that these siderophores are not promising targets for crop protection strategies. Ps1448a also has the genetic potential to synthesise a third siderophore, yersiniabactin although no evidence was found that this siderophore was produced during iron starvation, and deletion of a yersiniabactin NRPS from the chromosome of Ps1448a did not impair pathogenicity. Both achromobactin and pyoverdine were purified from the culture supernatant of iron starved Ps1448a for analysis by mass spectrometry. MALDI TOF and MS/MS analysis allowed the assignment of structures for both of these molecules. The characterisation work described in this chapter allowed a functional role to be assigned to each of the modules in the pyoverdine biosynthetic NRPS of Ps1448a. This was an important prerequisite for the module swapping experiments described in Chapter 5 of this thesis.

3.2 – Introduction

Acquisition of iron is essential for bacterial growth. However, due to insolubility at neutral pH the bioavailability of iron is extremely low in most environments. To circumvent this problem many bacteria respond to iron starvation by synthesising high affinity iron-chelating molecules known as siderophores. These siderophores are secreted into the extracellular environment where they bind ferric iron and are then actively transported back into the cell via specific ferric-siderophore receptors [113]. Siderophores are especially important in the biology of fluorescent *pseudomonads*, a genus renowned for occupying a very wide range of environmental niches. These bacteria synthesise the peptide-derived molecule pyoverdine as their primary siderophore, together with secondary siderophores which have lower affinity for iron [114]. Although *pseudomonads* are not obligate pathogens, many species are capable of causing disease in a wide variety of hosts [115, 116]. As iron restriction is a key host defence mechanism, pyoverdine is frequently identified as an important virulence factor [114, 117].

3.2.1 – Pyoverdine synthesis in fluorescent *pseudomonads*

Pyoverdine is synthesised from aa precursors by NRPS [118, 119]. It is pyoverdine that provides the fluorescent *Pseudomonas* species with their defining fluorescence and yellow-green pigmentation under conditions of iron limitation [120]. These properties derive from an invariant dihydroxyquinoline chromophore, to which is attached an acyl moiety and a strain-specific peptide side chain [121]. More than 50 different pyoverdine structures have been described to date [122] and the variability of the peptide side chain of pyoverdines from different strains reflects rapid evolution of both the NRPS that synthesise this side chain and the outer membrane receptors that recognise ferric pyoverdine [123]. Analysis of the pyoverdine locus of different *P. aeruginosa* strains has shown that it is the most divergent region in the core genome and that its evolution has been substantially shaped by horizontal gene transfer [105, 123, 124]. The diversification of pyoverdine structures is particularly interesting when viewed in the context of recent NRPS manipulation experiments [1, 14, 81] – the barriers to variation of pyoverdine structures that have evidently been surmounted by genetic recombination and evolution of the corresponding NRPS modules are the same barriers that often preclude artificial modification of biotechnologically relevant NRPS

systems [85, 125]. The wide variety of pyoverdine structures that derive from a limited pool of NRPS modules arranged in different combinations – together with the genetic tractability of many *Pseudomonas* species and the ability to detect pyoverdine production at nanomolar levels by UV-fluorescent screening [126] – make the pyoverdine synthetases potentially a very attractive model system to study NRPS recombination. However, in terms of providing ‘raw material’ for such work, experimental studies of pyoverdine NRPS genes to date have primarily focused on just a single species, *P. aeruginosa*. This work aimed to extend this focus to another fluorescent pseudomonad, *Pseudomonas syringae* pv. *phaseolicola* 1448a (Ps1448a), which secretes an alternative form of pyoverdine to PAO1. The NRPS modules characterised in this work provide the raw material for subsequent manipulation experiments described in Chapter 5.

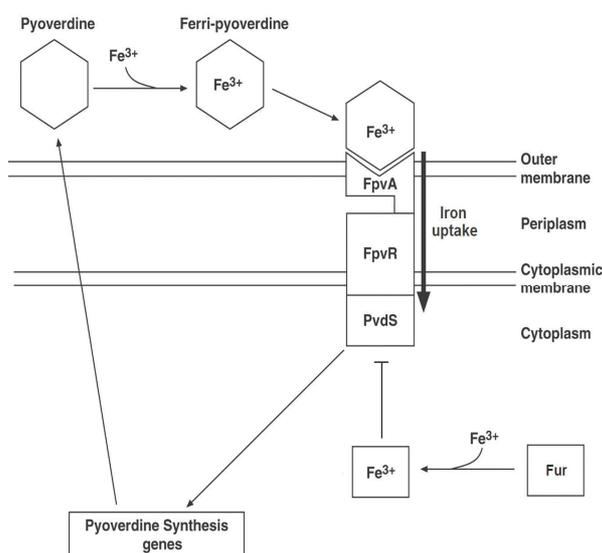
3.2.2 – Regulation of pyoverdine synthesis

Regulation of pyoverdine synthesis has been most extensively studied in *P. aeruginosa*, however the proteins involved in this regulatory mechanisms are found in other fluorescent *pseudomonads* [127, 128] and the regulatory pathways elucidated for *P. aeruginosa* are thought to be representative of all fluorescent *Pseudomonas* species [129]. Pyoverdine synthesis is regulated by Fur (ferric uptake regulator) protein, the master regulator of iron responsive genes in *Pseudomonas* and other bacteria. When complexed with Fe^{2+} , Fur represses the expression of genes containing a “Fur box” in their promoter sequence by binding to this element. During iron starvation, Fur is not complexed with Fe^{2+} due to low intracellular concentration. As a result Fur mediated repression of genes containing a Fur box is relieved [127, 128].

As with many iron responsive genes, those involved in pyoverdine biosynthesis and export/import are not directly regulated by Fur and do not contain Fur boxes in their promoters. Rather, regulation is indirectly mediated by Fur via expression of the sigma factors *pvdS* and *fpvI* and the anti-sigma factor *fpvR*. During iron starvation Fur repression of these genes is relieved, leading to their expression. The presence of PvdS enables expression of pyoverdine genes by association of this sigma factor with the core subunit of RNA polymerase

(RNAP). Once in association with PvdS, RNAP is able to initiate transcription of pyoverdine genes at an element in their promoter known as the iron starvation (IS) box [119, 127, 128].

Expression of pyoverdine transport and biosynthetic genes is also positively regulated by binding of ferric (holo) pyoverdine to its outer membrane receptor FpvA. When pyoverdine is not successfully scavenging iron from the environment, apo-pyoverdine is bound to FpvA. In this state the anti-sigma factor FpvR which is associated with the membrane receptor complex sequesters PvdS at the inner face of the bacterial membrane, down-regulating expression of pyoverdine transport and biosynthetic genes. Conversely, the binding of holo pyoverdine to



FpvA causes the release of PvdS and up-regulation of expression. Due to this dual regulatory mechanism, pyoverdine biosynthesis and transport is maximally induced only when cells are starved of iron and pyoverdine is able to successfully sequester iron from the environment [119, 127, 128, 130]. Figure 3.1 outlines pyoverdine regulation in *P. aeruginosa*.

Figure 3.1 – Regulation of pyoverdine synthesis and transport in *P. aeruginosa*

When intracellular iron concentration is low, Fur repression of *pvdS* and *fpvR* expression is relieved. Association of PvdS with the core subunit of RNAP enables transcription of pyoverdine biosynthesis and transport genes. An additional layer of regulation is provided by the anti-sigma factor FpvR. In the absence of ferri-pyoverdine binding to its receptor FpvA, PvdS is sequestered at the cytosolic face of the cell membrane by FpvR. Binding of ferri-pyoverdine to FpvR induces release of PvdS by an unknown mechanism resulting in up-regulation of expression of pyoverdine biosynthesis and transport genes. Figure reproduced from [130].

3.2.3 – NRPS independent siderophores

During the course of this study pyoverdine null mutants were generated, revealing that Ps1448a produces achromobactin as a secondary siderophore, and enabling us to genetically and functionally characterise this siderophore also. In contrast to pyoverdine, achromobactin is

synthesised by a mechanism that is entirely independent of NRPS enzymes. NRPS-independent siderophores have been studied much less intensively than their NRPS-dependent counterparts, and their mechanisms of synthesis have only recently begun to be deciphered. Three types (A, B and C) of NRPS-independent siderophore synthetase enzymes have been identified to date, each responsible for the attachment of a different functional group to a citric acid backbone [131, 132]. The achromobactin biosynthetic pathway is a particularly valuable resource for the study of these enzymes as it relies on the action of all three types of synthetase [131, 133]. Achromobactin has been shown to be important for virulence and growth under iron limitation in *Dickeya dadantii* (formerly *Erwinia chrysanthemi*) [134], but the contribution of achromobactin to these activities in Ps1448a has not previously been characterised. We therefore examined the roles of both achromobactin and pyoverdine in virulence of Ps1448a, as well as their relative contribution to iron uptake and growth under more precisely defined conditions.

3.3 – Identification and *in silico* characterisation of the pyoverdine NRPS genes

The biosynthesis of pyoverdine has been extensively studied in *P. aeruginosa* PAO1; most, if not all, of the genes required for pyoverdine synthesis in this strain have now been identified [121, 127, 135]. Unlike *P. aeruginosa*, *P. syringae* does not appear to exhibit a high degree of variability in pyoverdine structure from strain to strain; all fluorescent *P. syringae* strains examined thus far produce an identical pyoverdine molecule [136, 137]. We made use of the publicly available genome sequence for Ps1448a [138] (GenBank accession numbers CP000058 to 60) to carry out an *in silico* analysis of Ps1448a pyoverdine genes. This was achieved by conducting a BLASTP search of individual sequences from proteins known to be involved in pyoverdine synthesis in PAO1 against the Ps1448a protein database. Results of this analysis are summarised in Table 3.1, with PAO1 functions as described in [127]. The genomic organisation of pyoverdine genes in Ps1448a is similar, but not identical, to that of PAO1 (Figure 3.2). The only PAO1 gene that clearly lacks an ortholog in Ps1448a is PvdF, an enzyme required for generating the N⁵-formyl-N⁵-hydroxyornithine residues that are present in the PAO1 (but not Ps1448a) pyoverdine sidechain. In its place Ps1448a contains a gene

(*Pspph1922*; marked * in Figure 1) with a predicted protein product that is 37 % identical to the aspartate hydroxylase SyrP, required for synthesis of the NRPS-derived phytotoxin syringomycin in *Pseudomonas syringae* *pv. syringae* [78]. *Pspph1922* very likely catalyses β -hydroxylation of the two hydroxyaspartate residues that play an equivalent iron-chelating role to the N⁵-formyl-N⁵-hydroxyornithine residues of PAO1 pyoverdine.

Table 3.1 – Summary of PAO1 and Ps1448a pyoverdine gene alignment results

	PAO1 gene	Function in <i>P. aeruginosa</i> PAO1	Ps1448a ortholog(s) [†]
1	<i>pvdY</i>	Regulatory protein	1515
2	<i>pvdX</i>	Regulatory protein	3568
3	<i>pvdS</i>	ECF iron sigma factor	1909
4	<i>pvdG</i>	Thioesterase (34% identity with GrsT thioesterase from <i>Bacillus brevis</i>)	1910
5	<i>pvdL</i>	Chromophore peptide synthetase	1911
6	<i>pvdH</i>	Aminotransferase	1912
7	<i>Pa2412</i>	MbtH-like protein (no known function)	1913
8	<i>Pa2411</i>	Thioesterase (36% identity with thioesterase GrsT from <i>Bacillus brevis</i>)	1910
9	<i>Pa2403-2410</i>	No known function, however expression of these genes is co-regulated with pyoverdine synthesis genes. 2408 and 2409 are predicted to encode an ABC transporter	1914-1921
*	Not present	(Likely pyoverdine aspartate hydroxylase of Ps1448a)	1922
10	<i>pvdDIJ</i>	Pyoverdine side chain NRPS	1923-1926
11	<i>fpvA</i>	Ferripyoverdine receptor protein	1870, 1927 , 1928
12	<i>pvdE</i>	ABC transporter (secretion)	1929
13	<i>pvdF</i>	N ⁵ -hydroxyornithine transformylase	Not present
14	<i>pvdO</i>	No known function	1930
15	<i>pvdN</i>	26% identity with isopenicillin N epimerase from <i>Streptomyces clavuligerus</i>	1931
16	<i>pvdM</i>	Dipeptidase (23% identity with porcine dipeptidase)	1932
17	<i>pvdP</i>	No known function	1933
18	<i>Pa2391</i>	Porin (over 30% identity with outer membrane factor (OMF) proteins of RND/MFP/OMF-type efflux systems)	1934
19	<i>Pa2390</i>	ABC transporter (over 40% identity with resistance-nodulation-division (RND)-type transporter components of RND/MFP/OMF-type efflux systems)	1935/macB
20	<i>Pa2389</i>	Periplasmic protein (over 30% identity with periplasmic membrane fusion proteins (MFP) of RND/MFP/OMF-type efflux systems)	1936
21	<i>fpvR</i>	Antisigma factor for PvdS and FpvI	2117, 4764
22	<i>fpvI</i>	ECF sigma factor required for expression of <i>fpvA</i>	4765, 1175, 1093, 2747, 1909
23	<i>pvdA</i>	L-ornithine hydroxylase	2415, 3753
24	<i>pvdQ</i>	Acylase (38% identity with Aculeacin A acylase from <i>Actinoplanes utahensis</i>)	1937

[†]Gene numbers for Ps1448a are as annotated in the *Pseudomonas* genome database. Genes were presumed to be orthologs if they belonged to the same COG group. Hits are listed in order of significance, with those falling within the Ps1448a pyoverdine locus (as pictured in Figure 1) listed in bold.

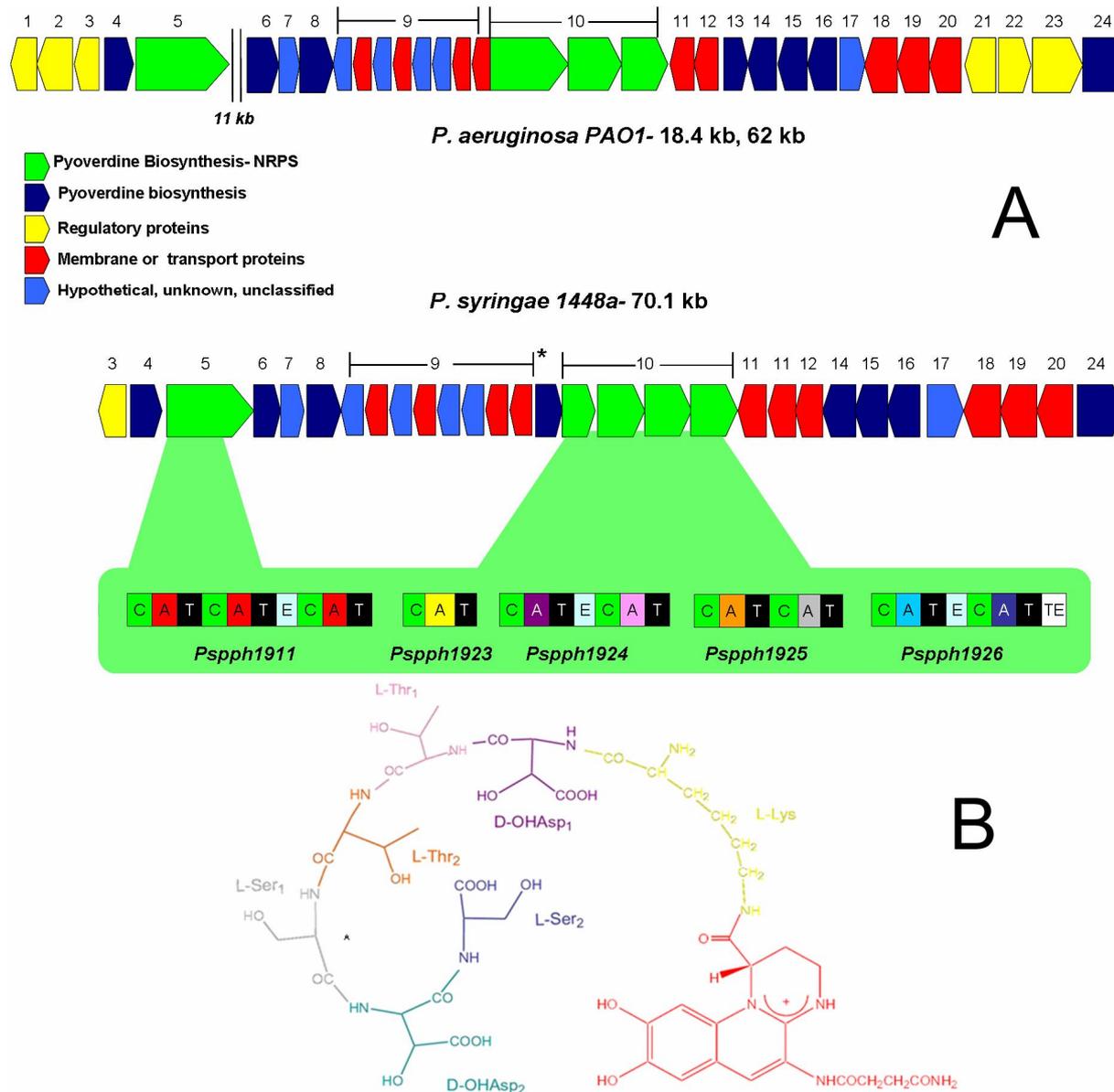


Figure 3.2 – A Comparison of the pyoverdine loci of *P. aeruginosa* PAO1 and Ps1448a

The core PAO1 pyoverdine genes fall into two closely linked clusters, 11 kb apart. In contrast, the core Ps1448a genes form a single contiguous cluster. Genes are colour coded according to their function, as indicated in the key; and orthologous genes in each organism have been assigned the same number, corresponding to the annotations in Table 1. The green highlighted region details the modular structure of the Ps1448a pyoverdine NRPS genes (C = condensation domain, A = Adenylation domain, T = Thiolation domain, E = Epimerisation domain, TE = Thioesterase). A-domains are colour coded to correspond with the aa residue that each incorporates into the Ps1448a pyoverdine molecule (as pictured in B).

Ps1448a contains 5 NRPS genes that lie within the pyoverdine locus (Fig 3.2). The gene *Pspph1911* presumably governs synthesis of the pyoverdine chromophore, as it shares 72.36 % predicted aa identity with the chromophore NRPS gene *pvdL* of *P. aeruginosa* PAO1 and homologs of this gene are present in all fluorescent pseudomonads that have been examined [121, 139, 140]. Likewise, the four contiguous genes *Pspph1923-1926* were strong candidates for encoding the side-chain NRPS of Ps1448a, with the total number of NRPS modules in these genes (7) corresponding exactly with the number of aas in the Ps1448a pyoverdine sidechain. Analysis of these genes using the TIGR online NRPS analysis tool (<http://www.tigr.org/jravel/nrps/>) (which analyses an 8-residue signature sequence that is predictive of the aa substrate specificity for an individual NRPS module [35]) as well as the more recently developed TSVM heuristic prediction software [39] shows that their predicted substrates, assuming the co-linearity that is typical of NRPS clusters [7], are entirely consistent with the residues that are believed to make up the pyoverdine molecule of Ps1448a (Table 2). Substrate specificities predicted by alignment of 10 residue signature sequences [36] against the database of known A-domains found at the TSVM NRPS predictor server were also consistent with the analysis above (Table 2; alignments not shown).

Table 3.2 – *In silico* prediction of A-domain specificity for Ps1448a pyoverdine sidechain NRPS

A domain	8 residue signature alignment	Identity of best match	TSVM prediction congruent?
1923	DGEDHGTV : DAESIGSV	BacB-M1-Lys bacitracin synthetase 2	No: val=leu=ile=abu=iva-like specificity
1924 mod1	DLTKIGHV : : DLTKVGHI	SrfAB-M2-Asp surfactin synthetase B	Yes: asp=asn=glu=gln=aad-like specificity
1924 mod2	DFWNIGMV DFWNIGMV	PvdD-M2-Thr pyoverdine synthetase	Yes: thr=dht-like specificity
1925 mod1	DFWNIGMV DFWNIGMV	PvdD-M2-Thr pyoverdine synthetase	Yes: thr=dht-like specificity
1925mod2	DVWHVSLI DVWHVSLI	PvdJ-M1-Ser pyoverdine synthetase	Yes: ser-like specificity
1926 mod1	DLTKIGHV : : DLTKVGHI	SrfAB-M2-Asp surfactin synthetase B	Yes: asp=asn=glu=gln=aad-like specificity
1926 mod2	DVWHVSLI DVWHVSLI	PvdJ-M1-Ser pyoverdine synthetase	Yes: ser-like specificity

3.4 – Confirmation of pyoverdine NRPS identity by gene knockout

In order to confirm that the candidates identified in 3.3 were in fact the pyoverdine biosynthetic NRPS genes of Ps1448a, an in-frame deletion of each gene was created and the phenotypes of the resulting mutant strains characterised.

3.4.1 – Construction of pyoverdine NRPS mutant strains

In-frame deletion of target NRPS genes from the chromosome of Ps1448a was achieved using a rapid overlap PCR-based method [141], a schematic overview of the deletion procedure is shown in Figure 3.3. A detailed description of the procedure for knockout generation is given in Section 2.7.3.

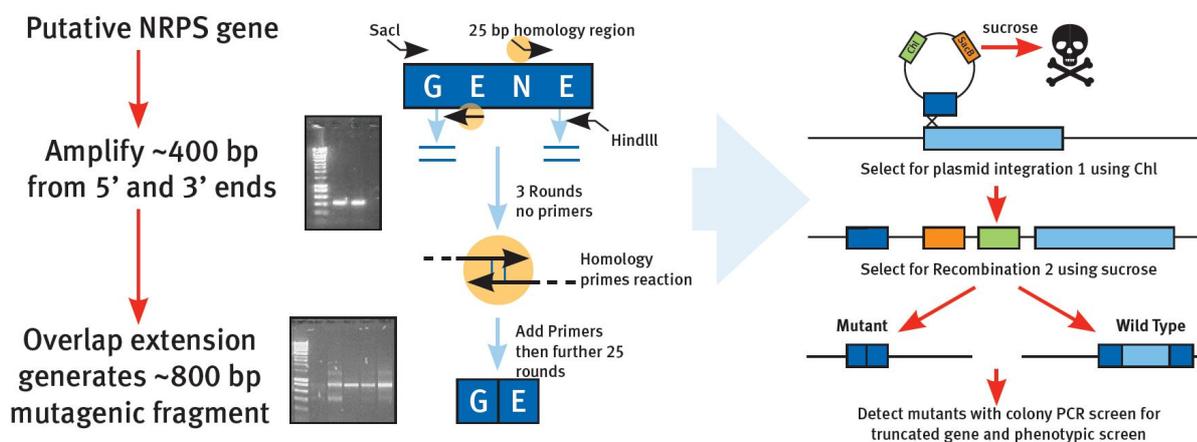


Figure 3.3 – Strategy for in-frame deletion of siderophore biosynthetic genes

A- Overlap PCR was used to generate 800 base pair mutagenic fragments; these fragments were designed so as to maintain the original reading frame of the gene

B- Mutagenic fragments were cloned into the vector pDM4 which is unable to replicate in *P. syringae*. The resulting mutagenic plasmids were delivered to cells by electroporation and homologous recombination events selected on chloramphenicol plates. Selection of cells in which a second recombination event had mediated loss of plasmid sequence from the chromosome were then selected using sucrose containing medium and the identity of double recombinants confirmed by PCR and/or Southern blot, as well as phenotypic analysis.

3.4.1.1 – Confirmation of knockouts by PCR and southern blot

In order to detect and confirm the identity of putative mutant strains a two step PCR screen was employed. In order to detect candidate gene deletion events, colonies isolated following growth on sucrose medium were screened directly by PCR (Figure 3.4a) using the protocol described in 2.7.2.3. In this initial screen primers flanking the gene of interest were used and the amplification conditions set such that colonies containing a truncated version of the gene would result in amplification of an 800 bp product, whereas WT colonies would result in no amplification. Genomic DNA was then prepared from clones positive for the 800 bp truncated gene and subjected to a second PCR screen using primers specific for the putative deleted region of the gene (Figure 3.4b). Clones found testing positive for the truncated gene and negative for intervening sequence were taken to be knockouts and subjected to phenotypic analysis. For $\Delta 1926$ and $\Delta 1924$ confirmation by southern blot was also carried out. Figure 3.4 shows a representative example of knockout confirmation by PCR and southern blot.

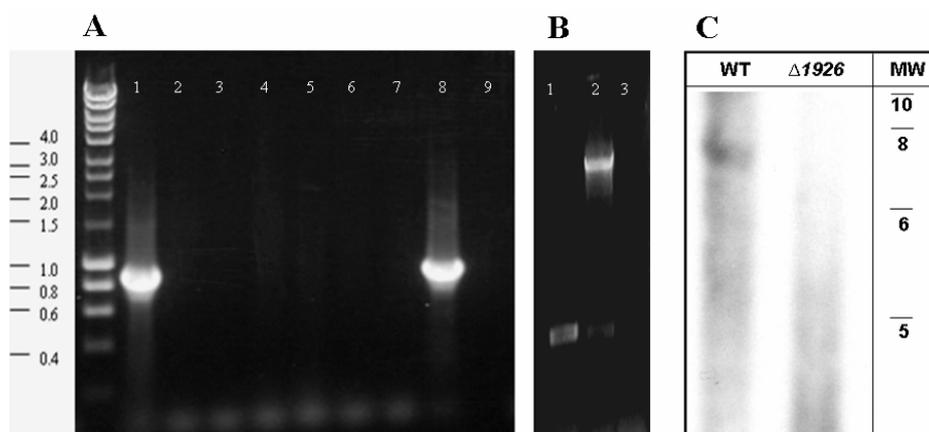


Figure 3.4 – Confirmation of *pspph1926* deletion by PCR and Southern blot

A – Putative knockouts were screened by colony PCR using primers that flank the gene of interest. Colonies one and eight are positive for 800 b.p truncated gene. Lane 9 is a no template control reaction.

B – Confirmation of gene deletion by PCR. Primers targeted to the deleted region of *pspph1926* amplified the expected 3.5 kb product when chromosomal DNA from the WT bacterium (lane 2) was used. This product could be amplified when genomic DNA from $\Delta 1926$ (lane 1) was used. Lane three is a no template control reaction

C – Confirmation by Southern blot. Luminescent probe targeted toward a 6.7 kb genomic DNA EcoRI fragment predicted to be absent in the mutant strain was used. The probe hybridized to WT but not $\Delta 1926$ genomic DNA that had been digested with EcoRI.

3.4.2 – Initial phenotypic characterisation of NRPS knockout strains

Following confirmation by PCR and/or Southern blot, each of the five pyoverdine NRPS knockout strains was subjected to three basic phenotypic tests to assess the capacity of the strain for pyoverdine production. The first of these was growth on iron-limiting King's B (KB) medium [142], a medium known to stimulate pyoverdine production in fluorescent *Pseudomonads*. When grown on KB plates, each NRPS gene deletion strain lacked the UV fluorescence of WT (Figure 3.5A), indicating that pyoverdine was no longer being produced. Siderophore secretion and the iron uptake capacity of mutant strains was also assessed using chromeazurol S (CAS) agar plates. CAS plates contain a dye which is blue-green when complexed with iron and yellow when iron is absent, such that when a siderophore producing organism is grown on CAS agar a yellow halo is observed around colonies as a result of secreted siderophore wresting iron from the dye. Secondary assessment on CAS agar was necessary to confirm the absence of pyoverdine production as pyoverdine fluorescence is pH dependent. As such, mutations that alter intracellular pH without reducing pyoverdine production would also abrogate fluorescence. Following 24 h growth on CAS agar at 28 °C

each of the five NRPS mutant strains lacked the characteristic orange halo of the wild type strain (Fig 3.5B). This result confirmed that the lack of fluorescence observed for the mutant strains was indeed due to pyoverdine deficiency. In order to assess growth under severe iron limitation each mutant strain, along with a wild type control was streaked on KB media containing 200 $\mu\text{g}/\text{mL}$ EDDHA (an iron chelating agent which restricts passive iron uptake). Unlike wild type Ps1448a, the mutant strains were unable to grow under severe iron limitation. Taken together, the phenotypes described above confirmed that none of the gene deletion strains were able to produce pyoverdine.

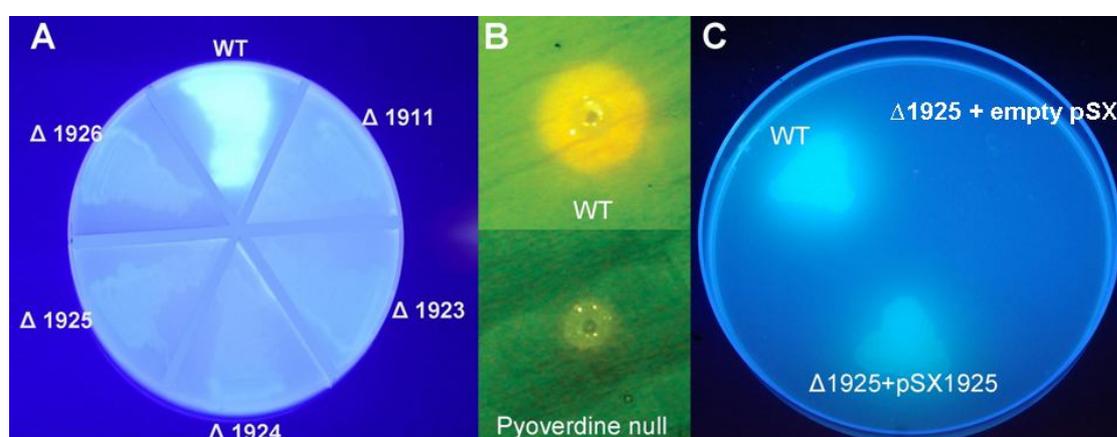


Figure 3.5 – Characterisation of Ps1448a pyoverdine NRPS knockouts

A – Wild type (WT) and pyoverdine NRPS knockouts ($\Delta 1911$, $\Delta 1923$ - 1926) on iron-limiting KB agar viewed under UV light. Only the wild type is able to synthesise fluorescent pyoverdine.

B – Wild type and pyoverdine null strain ($\Delta 1925$) inoculated into KB agar containing CAS dye and incubated for 24 h at 28 $^{\circ}\text{C}$. Only the wild type strain scavenged discernible levels of iron as evidenced by the orange halo surrounding this inoculum. All pyoverdine NRPS knockouts exhibited indistinguishable iron transport deficient phenotypes.

C – Wild type, $\Delta 1925$ and $\Delta 1925$ complemented by pSX:1925 on iron-restricted KB agar containing 200 $\mu\text{g}/\text{mL}$ EDDHA. Complementation by a functional gene copy *in trans* restored pyoverdine synthesis to near wild type levels in each of the NRPS knockout strains.

3.4.3 – Complementation of a pyoverdine NRPS deletion *in trans*

To ensure that the gene deletion strategy produced non-polar mutations it was necessary to complement a pyoverdine NRPS mutant *in trans* using a plasmid encoded copy of the wild type gene. To this end the entire sequence of the gene *Pspph1925* was cloned into the *Pseudomonas* expression vector pSX to generate the complementation plasmid pSX:1925. This plasmid was then delivered to Ps $\Delta 1925$ by electroporation. Introduction of the

complementation plasmid into the mutant cells resulted in a restoration of fluorescence and also allowed growth on KB agar containing 200 $\mu\text{g}/\text{mL}$ EDDA (Figure 3.4C). This successful restoration of pyoverdine production by complementation *in trans* indicated that the mutant phenotypes described in Section 3.3.2 did not result from polar effects. Expression from pSX:1925 is under the control of the IPTG inducible *tac* promoter, interestingly complementation by this plasmid was only observed when bacteria were grown at 18 °C in the absence of IPTG. This indicated that high expression levels or temperature were resulting in the production of inactive protein from pSX:1925. This result is consistent with previous observations that over-expression or elevated temperatures result in misfolded, inactive NRPS proteins [37, 85, 143].

3.5 – Identification and genetic characterisation of achromobactin as a secondary siderophore of Ps1448a

Although the pyoverdine deficient (Pvd^-) strains were unable to discernibly alter the colour of the CAS dye during 24 h growth on agar at 28 °C (Figure 3.5b), i.e. no active iron transport was apparent within this timeframe, some colour change was observed when these plates were subsequently left at room temperature or maintained at 28 °C for an extended duration. These observations suggested that the pvd^- strains were employing an alternative siderophore. Production of this presumed secondary siderophore appeared to be temperature dependent, with the pvd^- strains exhibiting greater iron uptake at 22 °C than at 28 °C (the latter being the optimal laboratory temperature for growth of Ps1448a [144] (Figure 3.6a,b). However, none of the pvd^- strains were able to grow during 72 h incubation at either temperature on solid media containing 200 $\mu\text{g}/\text{mL}$ EDDHA, indicating that this secondary siderophore had a much lower affinity for iron than pyoverdine.

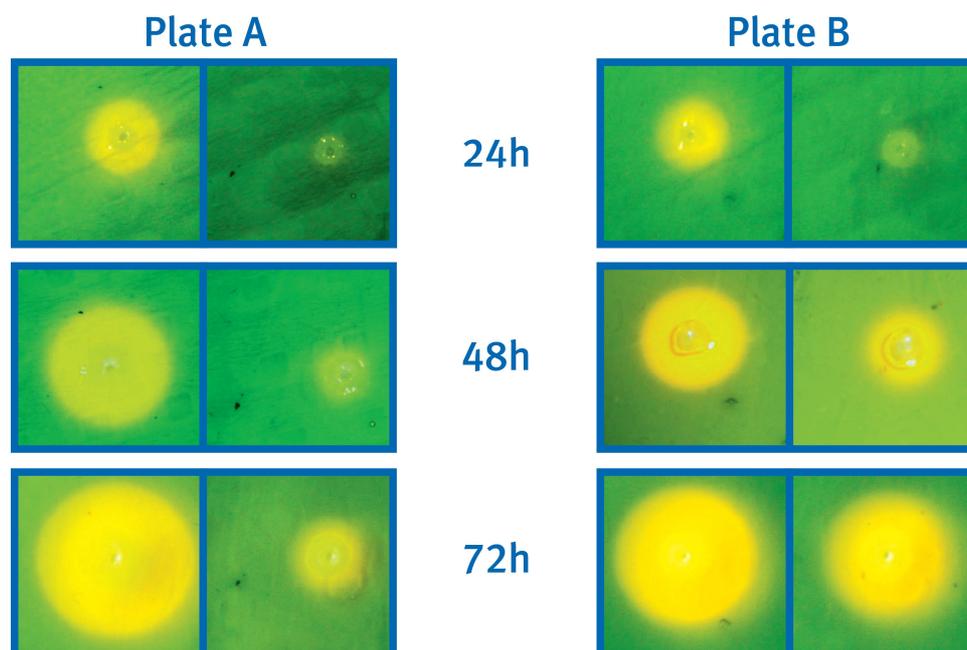


Figure 3.6 – Temperature-dependent production of a secondary siderophore by pyoverdine null Ps1448a

Wild type and pyoverdine null Ps1448a colonies were inoculated into identical Kings B plates containing CAS dye. Both plates were incubated at 28 °C for 24 h, following which plate B was removed to 22 °C for the remainder of the experiment while plate A was maintained at 28 °C. For each plate, wild type is on the left, and the pyoverdine null strain is on the right.

3.5.1 – *In silico* characterisation of the Ps1448a achromobactin biosynthetic locus

To identify candidate genes governing synthesis of this secondary siderophore, some known siderophore synthetase sequences from other phytopathogenic bacteria were aligned by BLASTP against the Ps1448a genome [138, 145]. This search revealed that Ps1448a contains orthologs of the *D. dadantii* achromobactin synthetase genes. Figure 3.7 compares the achromobactin gene locus of Ps1448a with that of *D. dadantii* [134], and presents a simplified scheme for achromobactin synthesis based on previous studies. In both *D. dadantii* and Ps1448a achromobactin appears to be synthesised by six proteins, AcsA-F. *D. dadantii* AcsD, a type A NRPS-independent synthetase, has been characterised both biochemically and structurally and found to catalyse the condensation of citric acid with L-ser [131, 133]. Roles for *D. dadantii* AcsACEF have also been proposed on the basis of protein homology studies and the known structure of achromobactin, but no clear role for AcsB has yet been identified

[132, 134]. The predicted functions of the corresponding Ps1448a orthologs are summarised in Table 3.3.

In addition to the *acsA-F* genes, the Ps1448a locus also contains an ortholog of *yhcA*, proposed to be involved in export of achromobactin [134]; *acr*, a TonB dependent siderophore receptor; and the four genes *cbrABCD*, the products of which form a permease complex for ATP-dependent import of ferric achromobactin [146]. Immediately adjacent to the Ps1448a achromobactin gene cluster are five genes (*Pspph2742-2746*) predicted (COG prediction) to encode proteins comprising an ABC-type dipeptide/oligopeptide/nickel transport system (not shown).

After the completion of the work described in this chapter, a paper was published in which the locus for achromobactin biosynthesis in *P. syringae p.v syringae B728a* (PssB728a) was characterised [104]. It was found that PssB728a contains an achromobactin locus identical to that of Ps1448a. The function of the biosynthetic enzymes of this locus was determined and agrees with the *in silico* characterisation described in this chapter.

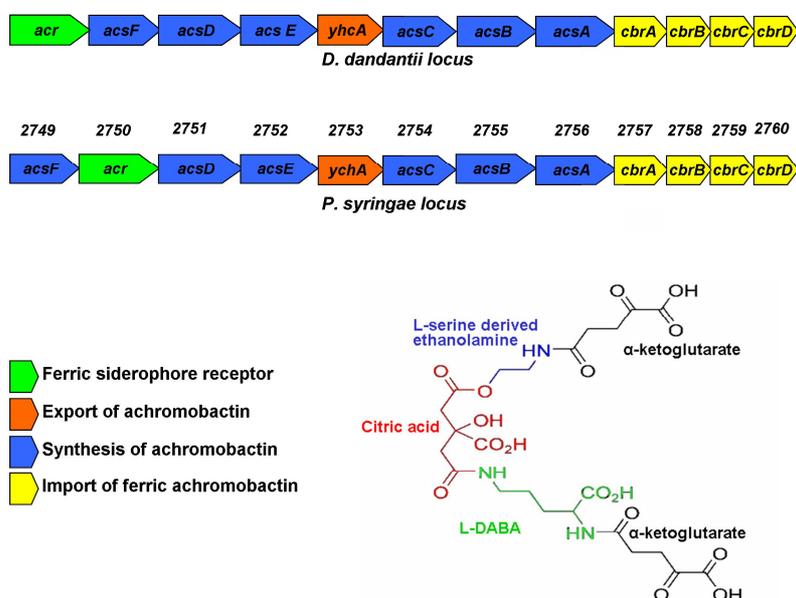


Figure 3.7 – Comparison of the achromobactin gene loci of *D. dadantii* and Ps1448a

Genes are colour coded according to function, as described in the key. The achromobactin structure is also colour coded to indicate the distinct constituent groups of this molecule that are hypothesised to be joined by the various achromobactin synthetase enzymes as summarised in Table 3.

Table 3.3 – Summary of *D. dadantii* and Ps1448a achromobactin gene alignment results

<i>D. dadantii</i> gene name	Ps1448a gene number	Putative function of Ps1448a ortholog [†]
<i>acr</i>	2749	TonB-dependent outer membrane ferric siderophore receptor (30%)
<i>acsF</i>	2750	Diaminobutyrate--2-oxoglutarate transaminase (74%). Makes L-DABA and α -ketoglutarates from aspartate semialdehyde and glutamate
<i>acsD</i>	2751	Type A NRPS-independent synthetase [131] (65%). Condenses L-serine with citric acid
<i>acsE</i>	2752	Pyridoxal-dependent decarboxylase (71%). Decarboxylates serine group of achromobactin to yield ethanolamine
<i>yhcA</i>	2753	Major Facilitator Family efflux pump YhcA (70%). Proposed to be involved in achromobactin export from cell
<i>acsC</i>	2754	Type C NRPS-independent synthetase (55%). Condenses L-DABA with product of AcsD reaction
<i>acsB</i>	2755	Aldolase/citrate lyase family protein (72%)
<i>acsA</i>	2756	Type B NRPS-independent synthetase (64%). Condenses two α -ketoglutarates with product of AcsC reaction
<i>cbrA</i>	2757	Ferric achromobactin-binding periplasmic protein precursor (70%)
<i>cbrB</i>	2758	Achromobactin inner membrane permease component (72%)
<i>cbrC</i>	2759	Achromobactin inner membrane permease component (70%)
<i>cbrD</i>	2760	ATP binding protein from the ferric achromobactin permease (80%)

[†]Based on *D. dadantii* alignment; numbers in brackets indicate aa identity with corresponding *D. dadantii* ortholog.

3.5.2 – Generation and characterisation of achromobactin deficient Ps1448a strains

In order to confirm the role of the proposed achromobactin biosynthetic locus the Ps1448a *acsA* homolog was deleted from both WT and *pvd*⁻ strains of Ps1448a using the same procedure described for generation of pyoverdine NRPS mutants. On solid media the achromobactin deficient (*acr*⁻) single mutant was indistinguishable in phenotype from wild type, growing effectively in the presence of 200 μ g/mL EDDHA and rapidly taking up iron on CAS agar. In contrast, not only could the *pvd*⁻/*acr*⁻ double mutant not grow in the presence of 200 μ g/mL EDDHA, it was also unable to take up any discernible amounts of iron on CAS agar irrespective of the duration or temperature of incubation (after 72 h at either 22 or 28 °C *pvd*⁻/*acr*⁻ colonies on CAS agar appeared identical to the 24 h *pvd*⁻ mutant pictured in Figure 3.5b.

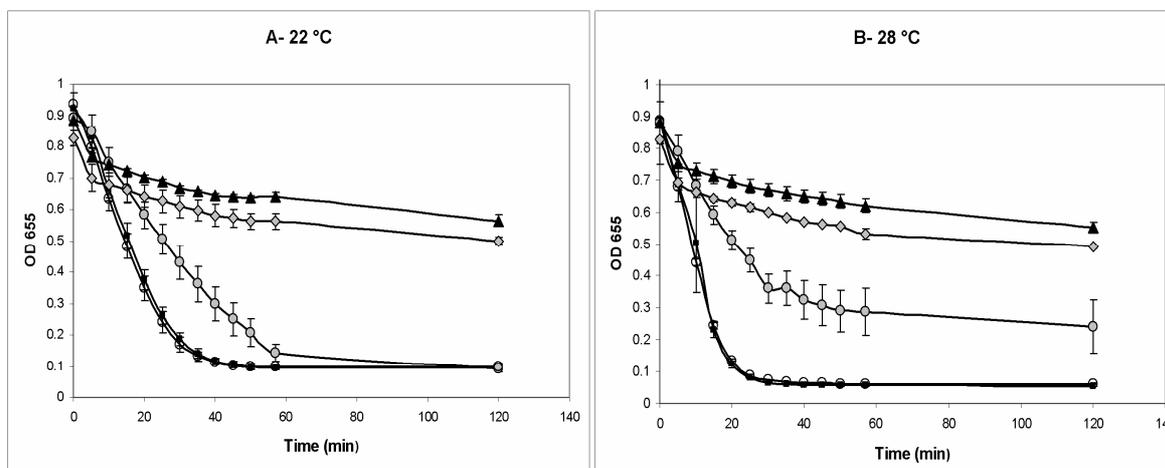


Figure 3.8 – Liquid CAS assay

96 well plate wells containing 200 μ l unamended King's B liquid media were inoculated in triplicate from synchronised overnight cultures of the following strains: WT (■), Acr⁻ (○), Pvd⁻ (●), and Pvd⁻/Acr⁻ (◆). A triplicate media-only control (▲) was also included. Plates were incubated with shaking at either 22 °C (A) or 28 °C (B) for 48 h. Cells were then pelleted and 150 μ l supernatant removed to fresh wells. CAS dye (30 μ l) was added to each well and the rate at which iron was removed from the dye by secreted factors in the supernatant was followed at OD 655 (monitoring loss of blue coloration). Error bars are presented as \pm 1 standard deviation.

Similar phenotypes were observed for each of the WT, pvd⁻, acr⁻, and pvd⁻/acr⁻ strains grown in liquid media, using a modified CAS assay that we developed to measure iron acquisition by factors secreted into the culture supernatant (Figure 3.8). In this assay strains were grown in quadruplicate in a 96 well plate for 48 h at either 22 or 28 °C. Following growth, cells were pelleted by centrifugation and supernatant removed to fresh wells. CAS dye was then added to the media and the rate at which iron was removed from the dye by secreted factors in the supernatant measured by tracking decrease in OD 650 (Absorbance maximum for CAS in the presence of iron). This assay showed that WT and acr⁻ strains were indistinguishable with regard to siderophore activity in the culture supernatant. Pvd⁻ strains had reduced siderophore activity compared to wild type, but this difference was less marked at 22 °C, indicating that Ps1448a is able to employ achromobactin as a temperature-regulated secondary siderophore that is secreted into the extracellular environment for active uptake of iron. The supernatant from the pvd⁻/acr⁻ strain appeared identical to a media-only control indicating that no

siderophores were being secreted by this strain, a result consistent with those obtained from the agar based assay (Figure 3.6). It is interesting to note that Ps1448a appears to have the genetic capacity to produce a third siderophore, yersiniabactin (Ybt) but we were unable to detect any evidence of Ybt production in any of the phenotypic assays we employed.

3.6 – Assessment of relative fitness of mutant stains under iron starvation conditions

In order to more precisely quantify the contribution of each siderophore under varying degrees of iron starvation, a serial dilution experiment was performed, employing EDDHA concentrations diluted 1:2 from 800 $\mu\text{g}/\text{mL}$ down to 0.2 $\mu\text{g}/\text{mL}$ in KB media in a 96 well plate. The WT, pvd^- , acr^- , and $\text{pvd}^-/\text{acr}^-$ strains were replica-inoculated into each well and incubated with shaking at 22 °C for 24 h, following which culture turbidity was measured. IC_{50} values (indicating the concentration of EDDHA that yielded only 50% turbidity relative to the unchallenged control) were calculated for each of the strains using Sigma Plot. The IC_{50} for the WT ($260 \pm 50 \mu\text{g}/\text{mL}$) and Acr^- ($220 \pm 70 \mu\text{g}/\text{mL}$) strains were approximately equal, confirming that pyoverdine is able to compensate for achromobactin deficiency. In contrast, the Pvd^- strain was sensitive to almost three orders of magnitude less EDDHA, with an IC_{50} of only $0.57 \pm 0.02 \mu\text{g}/\text{mL}$, demonstrating that achromobactin cannot completely compensate for the absence of pyoverdine. However, the IC_{50} for the $\text{Pvd}^-/\text{Acr}^-$ double mutant strain ($0.31 \pm 0.01 \mu\text{g}/\text{mL}$) was reproducibly lower yet, verifying that in the absence of pyoverdine achromobactin still makes a small contribution to fitness during iron starvation. At 28 °C the IC_{50} for WT and Acr^- strains were essentially unchanged, but the difference between the pvd^- mutant (0.38 ± 0.01) and $\text{pvd}^-/\text{acr}^-$ double mutant (0.26 ± 0.01) was less marked. Figure 3.89 gives a graphical representation of the results of this assay.

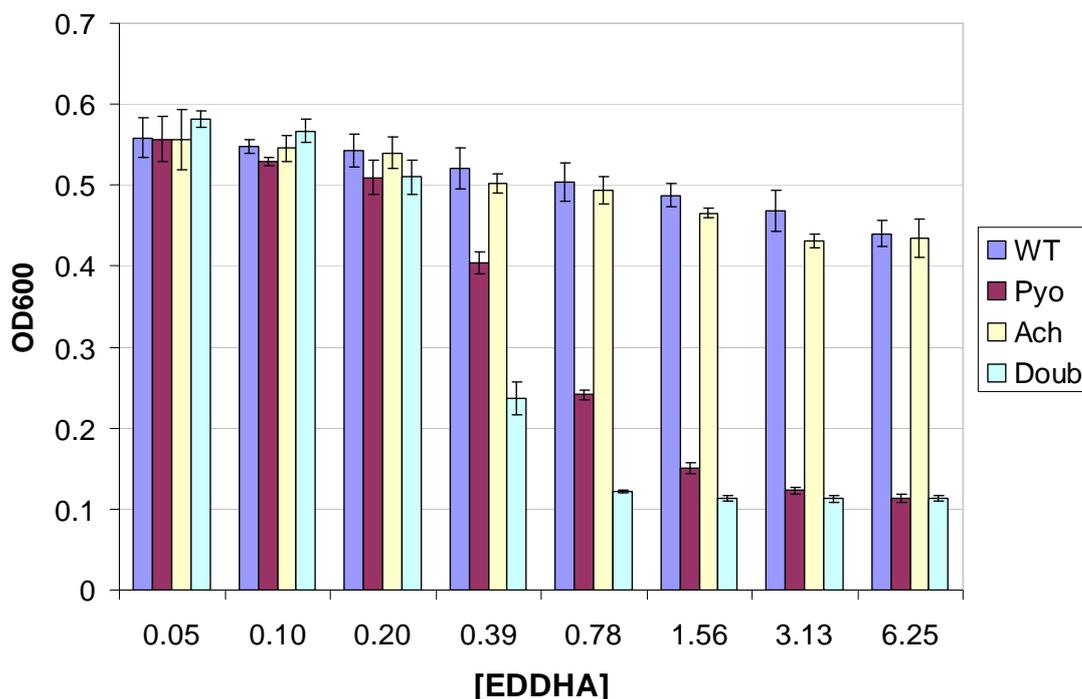


Figure 3.9 – Growth of WT and siderophore deficient strains during iron starvation

Quadruplicate cultures were set up for each strain at each concentration of EDDHA in a 96 well plate. Following 24 h growth at 22 °C, growth was assessed by measuring turbidity at 600 nm using a microplate reader. Increasing concentration of EDDHA results in greater restriction of iron availability. WT = Wild type Ps1448a, Pyo = pyoverdine null strain, Ach = achromobactin null strain, Doub = pyoverdine/achromobactin null strain. Data from two independent experiments are shown; error bars are presented as +/- one standard deviation

3.7 – Assessment of pathogenicity of mutant strains in *Phaseolus vulgaris*

In order to assess the pathogenicity in the natural host of Ps1448a, each of the mutant strains was subjected to the standard ‘bean prick’ pathogenicity test using bean pods [147]. All single and double mutant strains were still able to cause characteristic water soaked lesions after inoculation and incubation in bean pods, irrespective of temperature and whether or not the beans were picked or still attached to the parental plant. This indicates that neither pyoverdine nor achromobactin is essential in enabling Ps1448a to cause halo blight in the bean plant *Phaseolus vulgaris*. Although no evidence for Ybt production was observed in any of our

phenotypic assays, we wanted to investigate the possibility that this siderophore was still contributing in some manner to virulence in *Phaseolus vulgaris*. A *pvd*⁻/*acr*⁻/*ybt*⁻ strain was created by deletion of the proposed *ybt* biosynthetic NRPS *irp1* from a *pvd*⁻/*acr*⁻ background, this triple mutant was also assessed by the standard pathogenicity test and was also able to cause characteristic water-soaked lesions. Figure 3.10 shows a bean pod with water soaked lesions caused by each of the mutant strains and a wild type control.

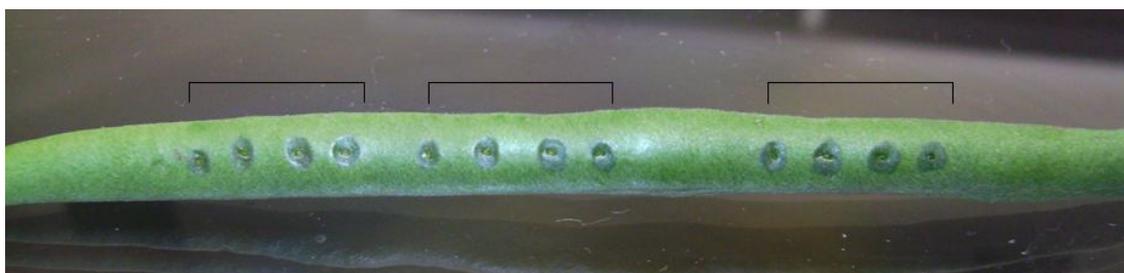
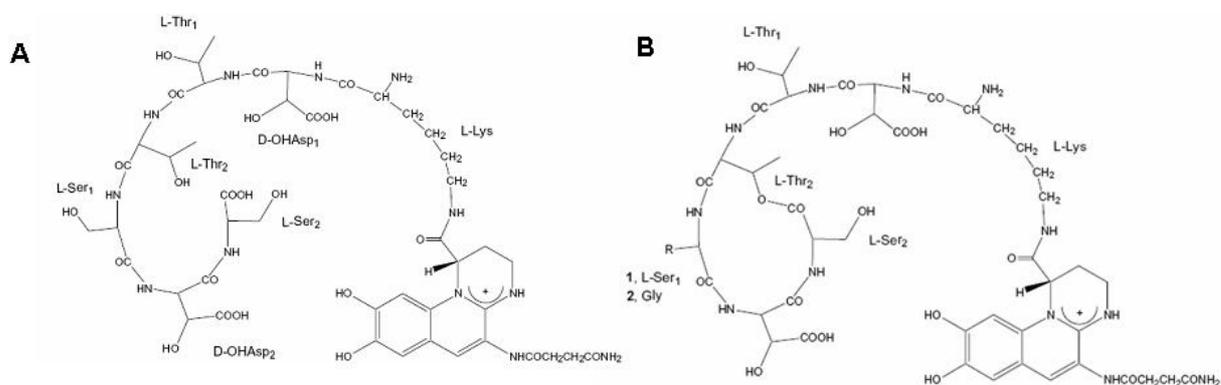


Figure 3.10 – Assessment of WT and mutant strains using standard pathogenicity test

Each of the indicated replicates contains, in order, WT, *pvd*⁻, *acr*/*pvd*⁻ and *acr*-/*pvd*⁻/*ybt*⁻ strains. Each strain was inoculated into a bean pod from a single colony, using a hypodermic needle. The pod was then incubated in a humid chamber at room temperature for 48 hr. All strains display characteristic water-soaked lesions indicating pathogenicity in *Phaseolus vulgaris*. Assay was repeated in triplicate, the photograph is representative of results from all repeats.

3.8 – Mass spectrometric analysis of pyoverdine and achromobactin purified from Ps1448a

In order to validate the *in silico* predictions of the structures of pyoverdine and achromobactin from Ps1448a, each siderophore was purified (by silica chromatography and amberlite bead purification methods for achromobactin and pyoverdine, respectively, as described in Sections 2.8.4.1 and 2.8.4.2) and subjected to analysis by mass spectrometry. Fractions were collected and analysed for siderophore activity using CAS dye, and those with the highest activity were sent to the Centre for Protein Research at the University of Otago for analysis. Each of the samples was subjected to MALDI TOF analysis to identify the mass of the primary constituents. The major peaks of the pyoverdine sample were also fragmented by MS/MS in order to discern the aa composition of the peptide side chain.



Pyoverdine, MS positive ion mode, zoom

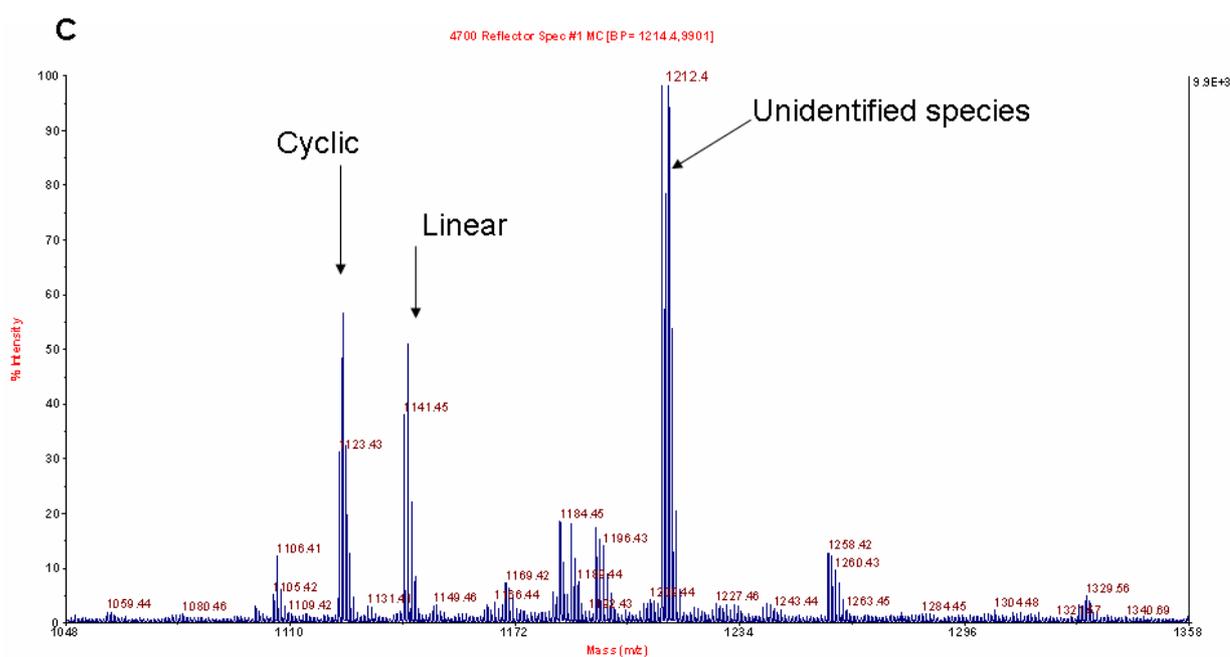


Figure 3.11 – MS analysis of pyoverdine purified from Ps1448a

Structures previously elucidated for pyoverdine from *P. syringae* have shown the molecule can either be in linear (**A**) or cyclic (**B**) form. The cyclic form results from the formation of an ester bond between the C-terminal carboxyl group and the sidechain of the second internal L-threonine residue (Figure 3.2b). Both of these forms of pyoverdine were identified in the MALDI TOF spectra (**C**) of pyoverdine purified from Ps1448a. A third peak not consistent with previously solved structures was also observed.

As shown in Figure 3.11, analysis of purified pyoverdine by MALDI TOF revealed the presence of three major positive ion peaks. Two of these peaks agree with the published masses of pyoverdines isolated from other *P. syringae* strains [136], the third peak does not. The peak at 1123 corresponds to cyclic pyoverdine, in which an ester bond between the C-terminal carboxyl and the side chain of an internal threonine residue results in a lactonic structure. The peak 18 mass units higher than this, at 1141, corresponds with linear pyoverdine in which this ester bond has either been hydrolysed or was never formed; pyoverdine in this form has not undergone a water loss due to ester bond formation and as a result is 18 mass units heavier than the cyclic form. The peak at 1212, 71 mass units higher than that of linear pyoverdine, can not be accounted for based on the structures previously elucidated for other *P. syringae* strains. In order to investigate the identity and order of the aas present, the peaks at 1141 and 1212 were subjected to MS/MS analysis. Analysis of peptides in this fashion is facilitated by fragmentation at amide bonds, the result of this is two characteristic ions. These ions are known as Y ions and B ions. As outlined in Figure 3.12 and Table 3.4, fragmentation of the peak at $m/z = 1141$ resulted in the formation of a set of B ions that corresponded exactly to the order and identity of aas that been predicted by *in silico* analysis of the pyoverdine NRPS of Ps1448a and that have previously been observed for pyoverdine from other *P. syringae* strains [136]. Fragmentation of the peak at $m/z = 1212$, results in an identical spacing and intensity, except all peaks are shifted 71 mass units higher. Possible explanations for this result are discussed in Section 3.9.

For achromobactin a major peak was present at 590.2, consistent with the published mass for achromobactin of 590.15 [104]; however the largest peak observed was at m/z 572.2, corresponding to a water loss (-18). For achromobactin, which had been evaporated to dryness then resuspended in solvent prior to analysis, the relative intensity of the peak at 572.2 was increased and the peak at 590.2 decreased, indicating that water was being lost as a result of heating during the purification procedure.

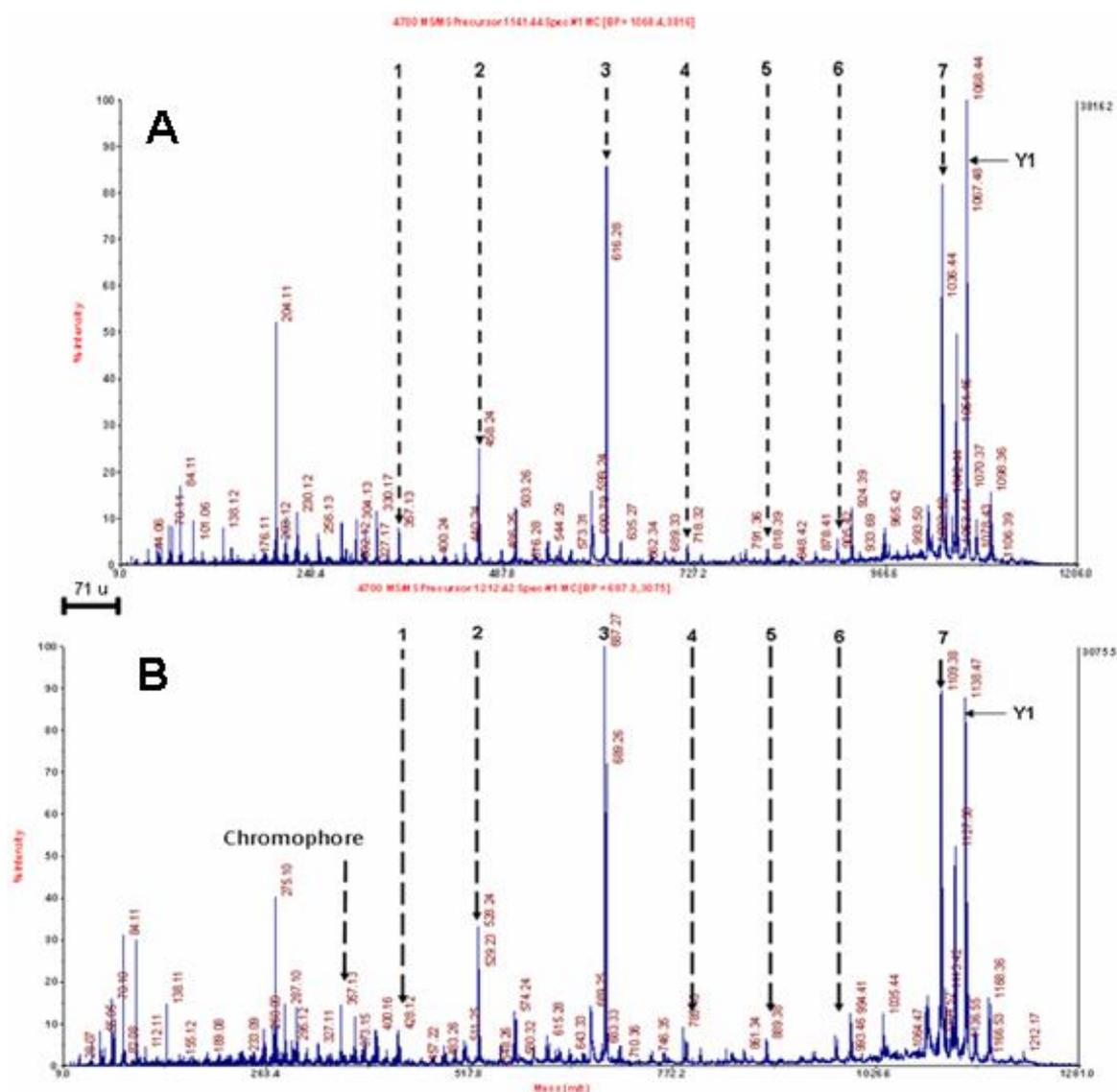


Figure 3.12 – Ions arising from MS/MS analysis of pyoverdines isolated from Ps1448a

A – Fragmentation of the peak found at 1141 mass units produces a set of molecular ions consistent with the predicted structure of Ps1448a pyoverdine. The masses of the labelled peaks and the ions they represent are given in Table 3.4.

B – Fragmentation of the peak at 1212 mass units produces a set of ions with identical spacing and intensity, shifted right by 71 mass units.

Peak number	Mass	Composition of ion
1	357.13	B ion: CH
2	458.24	A ion: CH_K
3	616.28	B ion: CH_K_OH-D
4	718.32	B ion: CH_K_OH-D_T
5	818.39	B ion: CH_K_OH-D_T_T
6	905.42	B ion: CH_K_OH-D_T_T_S
7	1036.41	B ion: CH_K_OH-D_T_T_S_OH-D
Y1	1067.48	Y ion resulting from loss of chromophore acyl group

Table 3.4 – Diagnostic ions arising from MS/MS analysis of the m/z = 1141 pyoverdine species

Fragmentation of the m/z = 1141 pyoverdine species resulted in identification of the following negative ions as shown in Figure 6.11. Peaks 1-7 match the expected pattern of A and B-ions previously reported for fragmentation of other *P. syringae* linear pyoverdine molecules [136]. Y1 has the correct mass for the Y ion resulting from loss of the acyl group of the chromophore. CH = chromophore, OH-D = hydroxyaspartate, all other aas indicated by standard one letter code.

3.9 – Discussion

This study constitutes the first experimental characterisation of pyoverdine NRPS genes in *Pseudomonas syringae*. The results of the bioinformatic, mutational and mass spectrometric analyses are consistent with the structure that has previously been proposed for the *P. syringae* pyoverdine molecule [136, 137], and support the mechanisms of synthesis that are outlined in Figure 3.2. As per previous pyoverdine NRPS gene knockouts in fluorescent *pseudomonads* [85], in-frame deletion of any of the chromophore or side chain NRPS genes in Ps1448a resulted in complete abolition of pyoverdine synthesis. Analysis of these mutants under iron-limiting conditions revealed the presence of a secondary siderophore, production of which was temperature sensitive. Temperature regulation of siderophore production has been observed for other bacterial species [148-150] and has been known to govern expression of other *P. syringae* genes, especially those implicated in causing disease [151]. Additional bioinformatic and mutational analyses identified achromobactin as the secondary siderophore. Achromobactin is known to contribute to virulence in *D. dadantii* [134], and these observations prompted us to test whether it is a virulence factor in Ps1448a also.

The contribution of both achromobactin and pyoverdine to virulence of Ps1448a during infection of *Phaseolus vulgaris* was assessed by inoculation of mutant strains and wild type controls into the bean pods. All single and double mutants were still able to cause lesions in this standardised pathogenicity test, indicating that neither siderophore is required for Ps1448a to cause halo blight in *Phaseolus vulgaris*. These results were surprising given that: iron is essential for core metabolic processes; is believed to be severely restricted in the plant extracellular environment [152]; and that siderophores are generally regarded as important for microbial pathogenesis of both plant and animal hosts [127, 134]. However, although the assumption is frequently made that pyoverdines are able to act as virulence factors in both animal and plant hosts, there is little experimental evidence for the latter. Indeed, pyoverdine from *P. syringae pv. syringae* has likewise been shown not to have a determinative role in pathogenesis of sweet cherry fruit [153]. It may be that phytotoxins render siderophores obsolete during the disease process by releasing iron from damaged plant cells into the extracellular environment. It should also be noted that the standard bean inoculation assay for Ps1448a virulence monitors only the ability to cause lesions, which is dependent primarily on toxin release and may not accurately report on the full progression of disease. Irrespective, it must be considered that any plant protection strategy which aims to target pyoverdine and/or achromobactin in *P. syringae pv. Phaseolicola* will not prevent the appearance of economically-damaging halo blight lesions in bean crops.

Despite no evidence for an active role in lesion formation, our phenotypic analyses of iron uptake and growth under iron limiting conditions confirmed that siderophores are indeed important for fitness of Ps1448a during iron starvation. Although *P. syringae* has traditionally been defined as a phytopathogen, it is unclear how important pathogenicity really is to the survival of this bacterium in the wild [154], and it may be that the Ps1448a siderophores are more important for epiphytic survival on leaf surfaces, in soil or water than during infection. However, given the clear superiority of pyoverdine as a siderophore, it is unclear why Ps1448a makes achromobactin also. All of the fluorescent *Pseudomonas* species known synthesise at least one secondary siderophore and there is presumably some fitness benefit to be derived from this investment. There is evidence that secondary siderophores can have affinity for

metals other than iron [155], and the presence of orthologs of known nickel-transport genes immediately adjacent to the Ps1448a achromobactin cluster may be indicative of a similar role in this bacterium (although there was no discernable phenotypic effect of nickel addition or exclusion on achromobactin synthesis in the *pvd* mutant; not shown). It has also been suggested that secondary siderophores can be involved in signalling pathways, or can have antimicrobial activities that are distinct from their iron scavenging properties [114, 156].

Alternatively, Dominique Expert and co-workers have demonstrated that achromobactin in the phytopathogen *D. dadantii* is synthesised temporally before the primary NRPS-derived siderophore chrysobactin [134]; and have proposed that achromobactin in this bacterium may function as a provisional measure, enabling cells to respond more rapidly to fluctuations in iron availability while the slower chrysobactin system is established [134, 152]. We suggest that a likely explanation for this scenario lies with the high energy investment required for activating NRPS mechanisms of siderophore synthesis. NRPS enzymes are amongst the largest known, with single proteins routinely exceeding 200 kDa [157]. The energy requirements for a cell to synthesise such large proteins are significant, and when already stressed this may represent an impassable barrier. However, once the NRPS enzymatic template is in place it is an extremely efficient method for synthesising short peptides, consuming significantly less ATP per peptide bond formed than ribosomal mechanisms [29]. It might therefore be useful to have a backup siderophore in place which is constitutively produced at low levels and can be rapidly up regulated in response to iron starvation. Such a system would provide the cell with small amounts of iron while the NRPS template for the more efficient primary siderophore is established. As the phenotypes of Ps1448a mutant strains indicate that achromobactin is only important when pyoverdine is not available, it is possible that achromobactin likewise serves as a ‘first response’ siderophore to cope with a sudden onset of iron starvation in Ps1448a.

The results of *in silico* and mass spectral analyses indicate that *Ps1448a* synthesises achromobactin which has the same structure as that of *PssB728a* [104]. MS analysis of pyoverdine purified from *Ps1448a* unexpectedly revealed the presence of three distinct pyoverdine molecules. Fragmentation by ms/ms suggested that the lighter two of these were

identical to the structures already elucidated for other *P. syringae* strains [136, 137]. The heavier species isolated was 71 Da larger. Three possibilities were considered in the determination of the identity of this molecule.

It was initially thought that this cryptic peak might be due to a pyoverdine molecule that had the same structure as previously published, with the exception of a larger acyl substituent attached to the chromophore or that it was a contaminant which had co-purified with pyoverdine. Fragmentation analysis conclusively refuted both of these possibilities. As shown in Figure 3.12, fragmentation of the 1212 Da species resulted in a set of peaks identical in spacing and relative intensity, to those seen for fragmentation of the 1141 Da species. The only difference between the two sets of peaks is that each peak arising from fragmentation of the heavier species is 71 Da larger. In both fragmentation sets, there is a peak at $m/z = 357$ which is the predicted mass of the chromophore of pyoverdine in which the attached acyl group is derived from succinic acid. In both spectra there are also intense peaks that correspond to a Y-ion formed as a result of loss of the acyl group from the chromophore; these peaks also differ by 71 Da. An alternative acyl group therefore seems an unlikely explanation for the cryptic pyoverdine species isolated. An alternative explanation for the spectrum obtained upon fragmentation of the larger pyoverdine species is that an extra monomer is incorporated into the side chain of pyoverdine. The B-ion pattern suggests that this monomer is the first residue of the sidechain, falling between the chromophore and L-lysine, increasing the mass by 71 Da. The only aa which could give this mass increase is alanine with a molecular weight of 89 which, following loss of water as a result of the condensation reaction would increase mass by 71. However, *in silico* analysis of the pyoverdine locus of Ps1448a did not reveal any NRPS modules specifying alanine, and likewise analysis of all annotated NRPS genes domains in the genome of Ps1448a also failed to reveal any A-domains predicted to specify alanine. Thus, the exact structure and mechanism of synthesis of the cryptic 1212 Da pyoverdine species remain unknown.

Chapter 4 – Techniques for purification and characterisation of NRPS A-domains

4.1 Introduction

The initial aim of this work was to biochemically determine the substrate specificity of each of the seven modules that synthesise the aa side-chain of pyoverdine in *Pseudomonas syringae* pv. *phaseolicola* 1448a. Modules characterised in this chapter were to be subsequently used in domain swapping experiments to generate novel pyoverdine derivatives and probe C-domain selectivity at the donor site (Chapter 5 of this thesis). As a first step toward achieving this goal, bioinformatic analyses were conducted to predict the preferred substrate for each synthetase module *in silico* (described in Chapter 3) and plasmid based constructs for the expression and purification of an AT bi-domain from each module were designed and constructed. Following optimisation of expression conditions in *E. coli*, soluble recombinant AT bi-domain was obtained for each module and subjected to a substrate binding assay. However, each of the proteins was inactive with all substrates tested. A number of approaches to generate active protein were subsequently undertaken, but despite exhaustive investigation, active protein could not be obtained for any of the seven modules.

This chapter describes techniques for production and purification of recombinant A-domains as well as new methods for assessment of their substrate specificity that were developed as a result of this work. While determination of substrate specificity for the seven *P. syringae* NRPS was ultimately unsuccessful, the techniques developed will be a valuable reference for future characterisation attempts.

4.2 General strategy for determination of substrate specificity

In order to biochemically determine the substrate specificity on a given NRPS module, it is necessary to first purify the A-domain of that module, either alone or in combination with other domains. One of the most convenient methods for doing this is to clone the nucleotide sequence encoding the desired domain or domains into a plasmid vector, allowing production of recombinant proteins in *E. coli* with a C or N terminal tag of six consecutive histidine residues. his₆-tagged proteins produced in this fashion can be easily isolated from crude cell lysate by virtue of the affinity of histidine residues for nickel ions immobilised on an agarose matrix [158, 159]. To allow for the production of his₆-tagged derivatives of *Pspph* pyoverdine synthetase modules in *E. coli*, the commercially available pET 28a+ (Figure 4.1) was chosen as an expression vector. pET 28a+ contains a multiple cloning site into which desired open reading frames (ORFs) can be introduced to yield either a C or N terminal his₆-tag. A search of the literature revealed only one published example where the substrate specificity of a *P. syringae* NRPS enzyme had successfully been determined [160], this paper assayed an AT bi-domain rather than an A-domain in isolation. In light of this, it was decided that AT bi-domains rather than A-domains would be purified for assessment of substrate specificity of each pyoverdine side-chain NRPS module.

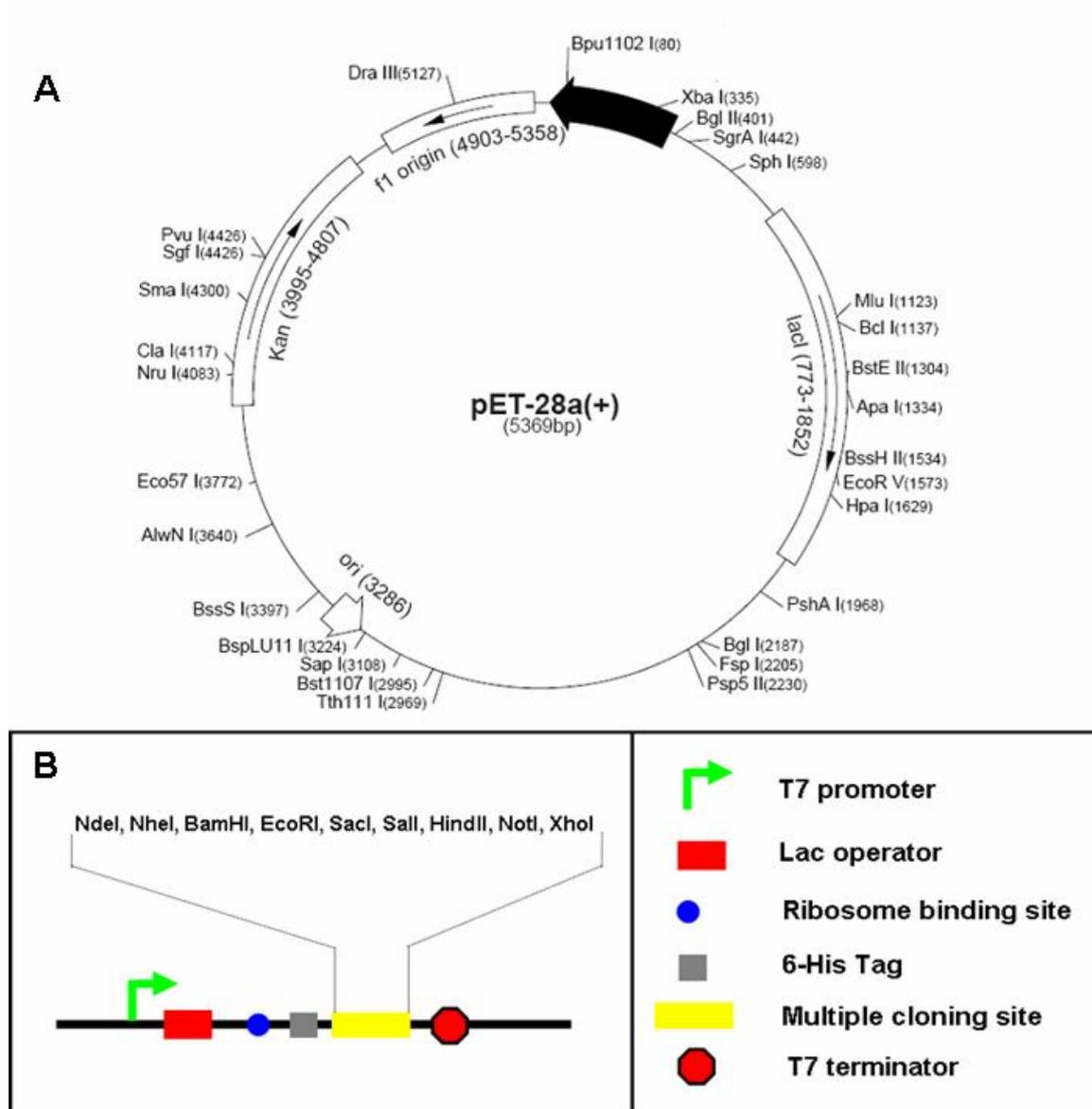


Figure 4.1 – pET28a+ map and features

A – A map of the vector showing important features: of particular note, the pBR 322 origin of replication (ori) allows plasmid replication in *E. coli*, Kan encodes a kanamycin resistance selectable marker, and lacI encodes the lac repressor protein.

B – Detail of the expression region and multiple cloning site: Expression is driven by the T7 promoter, in the absence of IPTG the lac repressor protein binds to the lac operator preventing transcription initiation. Addition of IPTG to the culture media causes dissociation of the lac repressor and initiation of transcription and subsequent translation of plasmid encoded ORFs. Location of the N-terminal his6-tag region is also indicated. The T7 promoter is only recognised by T7 polymerase, hence expression from this plasmid is only possible in *E. coli* strains that have a chromosomal or plasmid encoded copy of T7 polymerase. BL21(DE3) is the best known example of such a strain. Panel A is taken from the Novagen pET28a+ manual.

4.2.1 Design and construction of AT bi-domain expression vectors

The genes for production of the *Pspph* pyoverdine side-chain synthetases were identified *in silico*, as described in Section 3.3. The AT domain boundaries were approximated using the TIGR NRPS server (<http://nrps.igs.umaryland.edu/nrps/>) and primers were designed to amplify the AT bi-domain from each module. Henceforth, each AT bi-domain will be referred to by a number designation, according to the last two digits of its gene number and the number of the module from which it was derived. For example the AT bi-domain from the second module of the gene *Pspph1925* is designated 252. Primer pairs were picked using the “amplify selection” function in Vector NTI and restriction sites for directional cloning of AT bi-domains into pET 28a+ were introduced at the 5’ end of the forward and reverse primer for each construct.

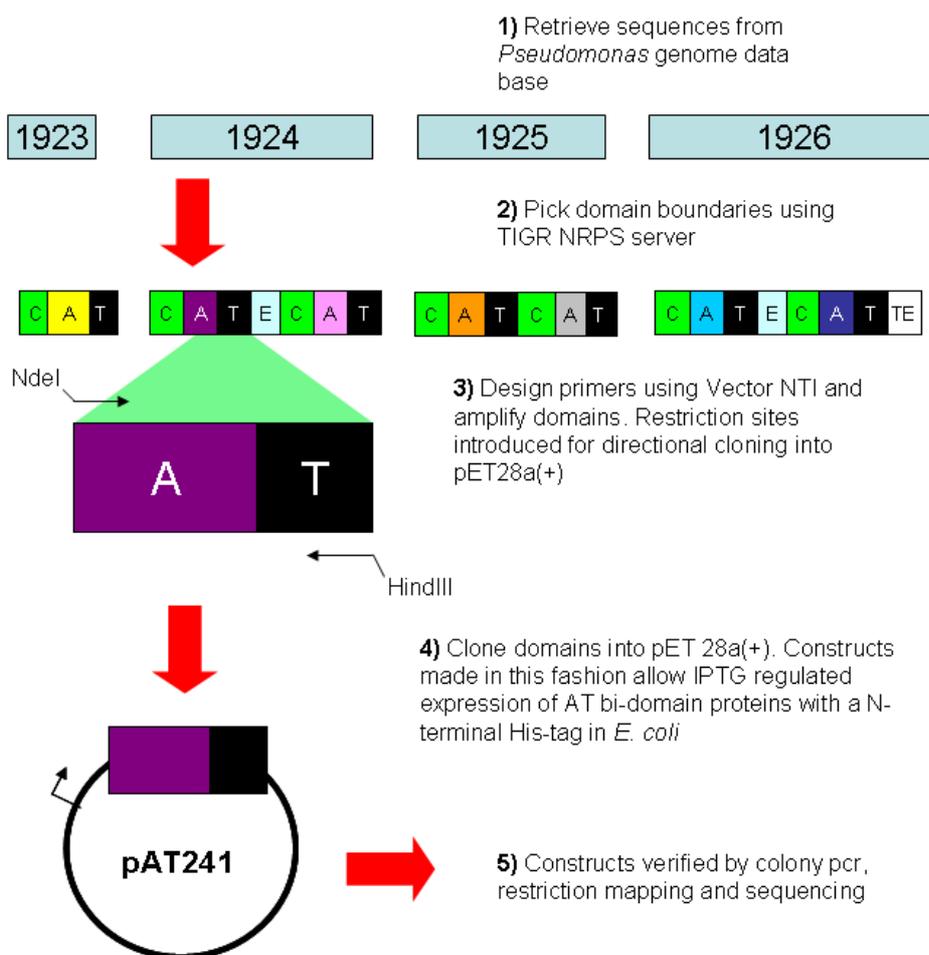


Figure 4.2 – Construction of AT bi-domain expression vectors

The upstream restriction site was NdeI and the downstream restriction site was Hind III for all constructs except 231 and 251, which had an XhoI and an EcoRI site respectively on the downstream primer (Refer to Table 2.5 for sequences). Amplification of domains, cloning and screening for recombinants were carried out following standard molecular biology protocols as described in Section 2.7; Figure 4.2 gives an overview of design and construction of the expression vectors. Following confirmation by colony PCR and restriction digest, the identity of the AT bi-domain insert in recombinant plasmids was verified by sequencing.

4.2.2 – Preliminary expression testing

After verification at the nucleotide level, each construct was transformed into the *E. coli* expression host strain BL21(DE3), and tested to determine whether or not the correct sized protein was expressed. Typically, expression was tested by growing a small-scale culture to an appropriate density, taking an uninduced sample, adding IPTG to induce protein expression and then taking aliquots of the culture 1, 2 and 3 h after induction. Cells from the aliquots were then pelleted and lysed and total cellular protein analysed by SDS PAGE. In each case it was possible to see clear induction of expression of the AT bi-domain.

4.2.3 – Optimisation of expression conditions to achieve soluble recombinant proteins

In order to purify an active his6-tagged derivative of a protein, it is necessary that the protein is not only expressed to high levels but is also present in a soluble form [161]. In order to determine whether each recombinant AT bi-domain was soluble under standard expression conditions (37 °C, 0.5 mM final concentration IPTG), cells from a 50 mL culture were harvested 3-6 h after induction, lysed by French-press and the resulting lysate separated into soluble and insoluble fractions by centrifugation. The insoluble fraction was then re-solubilised in a denaturant, and samples of each fraction analysed by SDS PAGE. For each of the constructs, analysis of this type revealed no discernable target protein in the soluble fraction (Figure 4.3a), it was therefore necessary to optimise expression conditions so that recombinant protein was present in a soluble form.

It is well known that over-expression of proteins in *E. coli* can lead to the formation of insoluble aggregates of misfolded proteins known as inclusion bodies; in part, this is thought to be due to excessive production overwhelming the molecular chaperones that guide folding of nascent proteins [161-163]. NRPS proteins are particularly difficult to work with in this regard and expression conditions generally have to be optimised to obtain soluble protein [37, 80, 160, 164, 165]. Two established methods for increasing the solubility of recombinant proteins in *E. coli* are expression at low temperatures and reduction of induction levels; the former is believed to increase protein stability, the latter to prevent protein folding machinery being overwhelmed [161]. Both of these factors were investigated in small-scale (50 mL) cultures, however expression at 25 °C – with final concentrations of IPTG as low as 0.0625 mM – failed to yield any soluble protein as assessed by SDS PAGE (not shown).

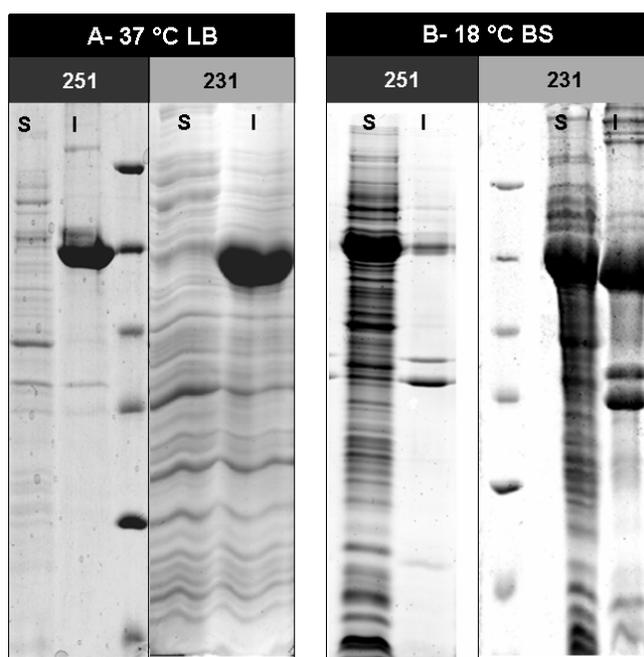


Figure 4.3 – Obtaining soluble protein by expression at low temperatures with 1M D-sorbitol and 2.5 mM glycine betaine

A – Soluble (S) and insoluble (I) fractions for 251 and 231 expression strains following expression for 3 h under standard conditions (37 °C, LB medium). The target protein is located exclusively in the insoluble fraction in both cases.

B – Soluble and insoluble fractions from the same strains following expression at 18 °C for 24 h in LB medium amended with 1 M D-sorbitol and 2.5 mM glycine betaine. There is now a significant amount of each target protein in the soluble fraction.

Molecular weight marker used for this and all subsequent gels are described in Section 2.9.3.1

Following the failure of low temperature and reduced expression rate to yield any soluble recombinant protein, it was decided to carry out expression at low temperatures in the presence of 1 M D-sorbitol and 2.5 mM glycine betaine [166]. Under these conditions Ackerley *et al.* were previously successful in purifying a functional his6-tagged AT bi-domain

from a *P. aeruginosa* pyoverdine NRPS [37]. The rationale behind the media composition is first, that the high concentration of D-sorbitol induces osmotic shock and expression of associated chaperone proteins; and second, that uptake of glycine betaine is promoted by osmotic shock, and once internalised is believed to assist in the folding of nascent proteins [10, 14]. Initially a small-scale (50 mL) trial of this technique was run using the 231 and 251 expression strains and found in both instances to yield large amount of soluble recombinant AT bi-domain (Figure 4.3b).

The betaine/sorbitol expression procedure was subsequently scaled up to 500 mL and then used to purify soluble his6-tagged AT bi-domain from each of the seven modules. In each case cell lysis was carried out by French press and removal of insoluble debris by centrifugation. Target protein was purified from the resulting soluble fraction by Ni-NTA affinity chromatography using a Novagen® Hisbind kit, following the manufacturer's directions. Figure 4.4a shows SDS PAGE analysis of such purification, Figure 4.4b shows SDS PAGE analysis of all purified AT bi-domains. After purification desalting and quantification, proteins were ready for use in a substrate specificity assay.

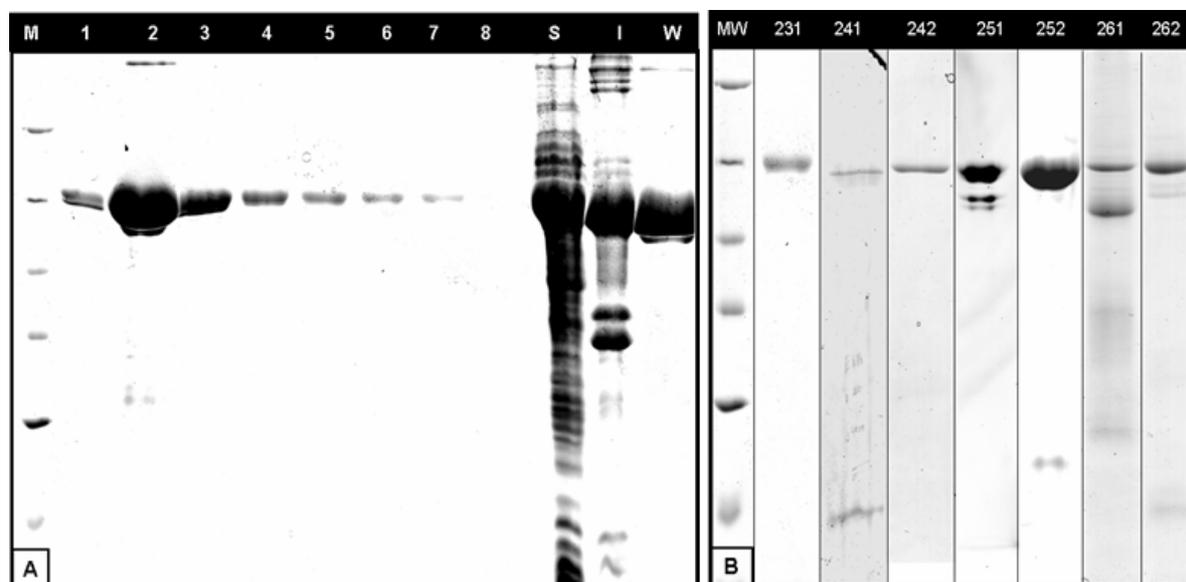


Figure 4.4 – Ni-NTA purification of AT bi-domains

Soluble his6-tagged AT bi-domain was purified from each of the seven over-expression strains using a Novagen® Hisbind® column containing 1.5 mL of settled resin. Protein binding, wash steps and elution were carried out according to the manufacturer's directions. Eluted protein was captured in 1 mL fractions for analysis by SDS PAGE, a sample of column flow through from washing steps was also retained for analysis.

A – SDS PAGE analysis of one such purification (that of 231); **1-8** = 1 mL fractions of eluted protein in the order they came of the column; **S** = soluble fraction; **I** = insoluble fraction; **W** = flow through from washing steps.

B – SDS PAGE analysis of each AT bi-domain following purification and buffer exchange. This panel is a compilation of various gels and shows each purified protein immediately after buffer exchange.

4.2.4 – Assessment of substrate specificity by ATP/³²PPi exchange assay

After soluble recombinant AT bi-domain had been purified for each module, it was assessed for substrate specificity using the ATP/³²PPi exchange assay. This assay was the first to be developed for assessment of A-domain specificity and has since become the accepted standard for substrate specificity determination [1, 4, 167]. Many variants of this assay protocol have been published [1, 3, 10, 12, 18, 19], all of which operate on the same basic principle. The assay relies on the fact that the production of aa adenylates by an A-domain is an equilibrium reaction. Activation of a cognate aa substrate (**S**) by the enzyme (**E**) consumes one molecule of **ATP** and generates enzyme-bound aa adenylate (**E-S-AMP**) and free pyrophosphate (**PPi**); whereas the reverse process yields ATP and free substrate. The equilibrium situation can thus be represented by the equation $E+S+ATP \leftrightarrow [E-S-AMP] + PPi$ [143]. If radiolabeled ³²PPi is

present in a reaction it will be incorporated into AMP to yield ^{32}P -ATP, but only if a substrate that is recognised and activated by the enzyme is present. Two versions of this assay were employed for assessment of substrate specificity, both of which were validated using positive and negative control reactions prior to use with the purified AT bi-domains described above.

4.2.4.1 – Validation of ATP/ ^{32}P PPi exchange assay version 1

The first version of the ATP/ ^{32}P PPi exchange assay used (EA1) followed the protocol of Ackerley and Lamont [37, 143]. In this assay exchange reactions containing enzyme, buffer, essential cofactors, ^{32}P PPi and individual aa substrate candidates were set up in triplicate in 1.5 mL microfuge tubes and incubated for 30 min at 28 °C. Following incubation, a suspension of activated charcoal in perchloric acid solution was added to terminate the reaction and bind ATP. The charcoal was then applied to a 0.22 μM filter and washing steps carried out to remove free ^{32}P PPi. After washing, ATP was eluted from the filter and transferred into scintillation vials containing distilled water. Cerenkov counts were then taken to assess the amount of ^{32}P -ATP that had been formed in the presence of each substrate.

To test this protocol a control experiment was run, using soluble fractions from *E. coli* or *P. syringae* crude cell lysate in the place of purified enzyme. Cell lysate contains a mixture of all soluble enzymes, including those that carry out reversible aa adenylation reactions such as NRPS enzymes and tRNA synthetases. It was found that inclusion of *E. coli* lysate resulted in a 31.6-fold increase in radioactivity recovered from activated charcoal when compared to reactions without any lysate. Inclusion of *P. syringae* lysate resulted in 6.4-fold increase in radioactivity. The results of this experiment are shown graphically in Figure 4.5. These results indicated that the assay was effective for recovering and quantifying ^{32}P -ATP formed as a result of an enzymatic reaction and therefore suitable for the assessment NRPS A-domain substrate specificity.

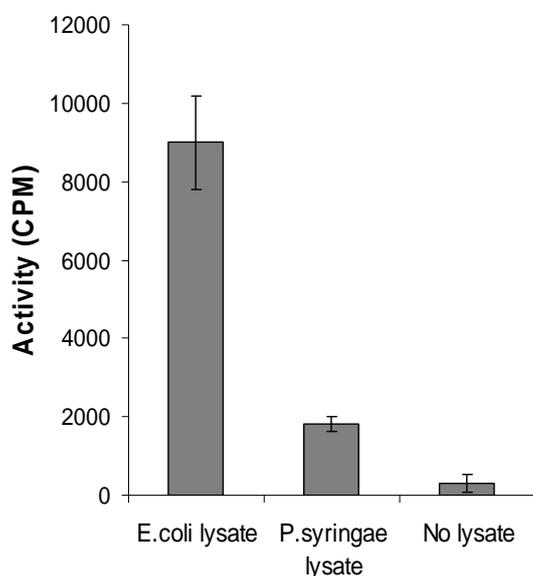


Figure 4.5 – Validation of exchange assay EA1 using cell lysate positive controls

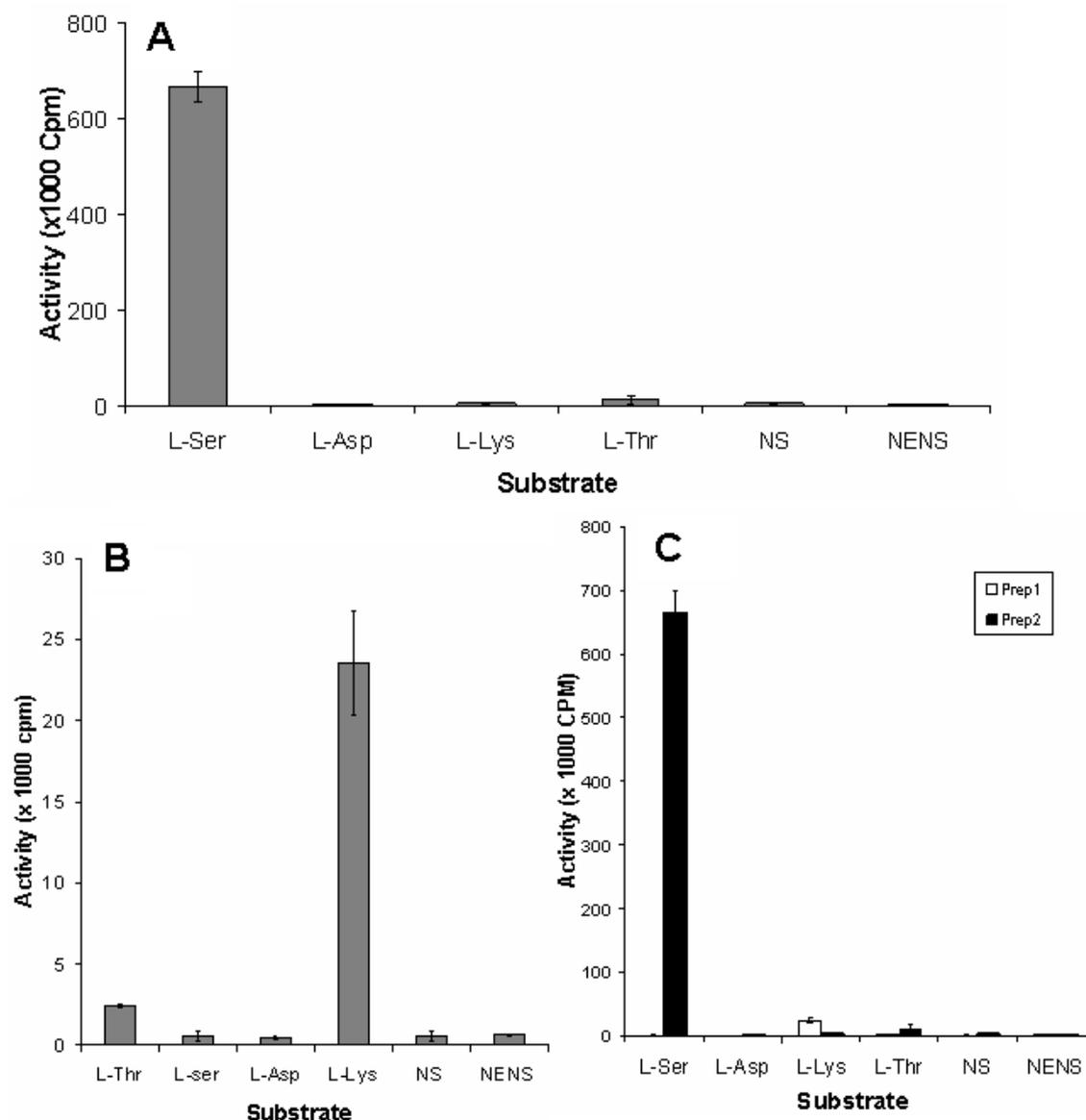
Assays were performed in triplicate as described in Section 2.10.1, error bars represent +/- one standard deviation. Lysate positive controls contained 10 μ L of filter sterilised cell lysate soluble fraction from either *E. coli* or *P. syringae*. In the case of the no-lysate reactions, 10 μ L lysis solution was included in the place of cell lysate.

4.2.4.2 – Validation of exchange assay version 2

Although shown to be effective for recovering and quantifying 32 P-ATP, EA1 proved time-consuming and cumbersome, owing mainly to the need for immobilisation of charcoal from each reaction on a separate filter followed by washing and elution steps. During the course of this study an alternative version of the assay (EA2) was tested, employing a micro-centrifuge-based protocol that has been used in a number of published characterisation reports [10, 16-19]. Exchange reactions were set up, incubated and terminated by charcoal/perchloric acid addition as for EA1, except a 10-fold increase of 32 PPi was included to improve sensitivity. Following reaction termination, charcoal was pelleted by centrifugation and washing steps carried out in 1.5 mL microfuge tubes. The charcoal was then suspended in distilled water and transferred to a scintillation vial for direct measurement of Cerenkov counts. Total processing time from termination to measurement of Cerenkov counts for EA2 was about 1/3 of that for EA1.

A positive control experiment was also run for EA2. In this experiment a previously characterised serine activating A-domain, that of the *E. coli* siderophore synthetase EntF [20,

21], was employed. EntF A-domain protein was expressed from pET28a+ in a BL21 background at 18 °C in LB medium (without betaine and sorbitol addition being required) and subsequently purified. Exchange reactions were set up in triplicate, testing four different aas (L-ser, L-asp, L-lys and L-thr) and a no substrate control. Following washing and transfer to scintillation vials, the amount of ^{32}P -ATP formed in each reaction was assessed by Cerenkov counting. The first attempt of this assay showed weak but reproducible ATP/ ^{32}P PPi exchange activity for recombinant EntF in the presence of lysine only (Figure 4.6A) – an unexpected result given that EntF is a known activator of serine. A second batch of recombinant EntF was subsequently prepared, this time following the betaine/sorbitol expression procedure used with the *P. syringae* AT bi-domains. The results of this second experiment, shown in Figure 4.6B, are clearly consistent with EntF being an activator of L-serine. The EA2 assay procedure, as well as prior enzyme purification steps, was therefore deemed suitable for determination of NRPS A-domain specificity. The difference in activity between the two preparations of EntF indicate that NRPS A-domains are capable of adopting soluble but inactive conformations in solution and also demonstrates the utility of the betaine/sorbitol expression method (Section 4.2.3) for promoting correct folding of recombinant proteins.



4.6 – Validation of exchange assay EA2 using EntF A-domain positive control

A – Exchange activity of his6-tagged EntF A-domain preparation 2 in the presence of 10 mM of the indicated aa substrate.

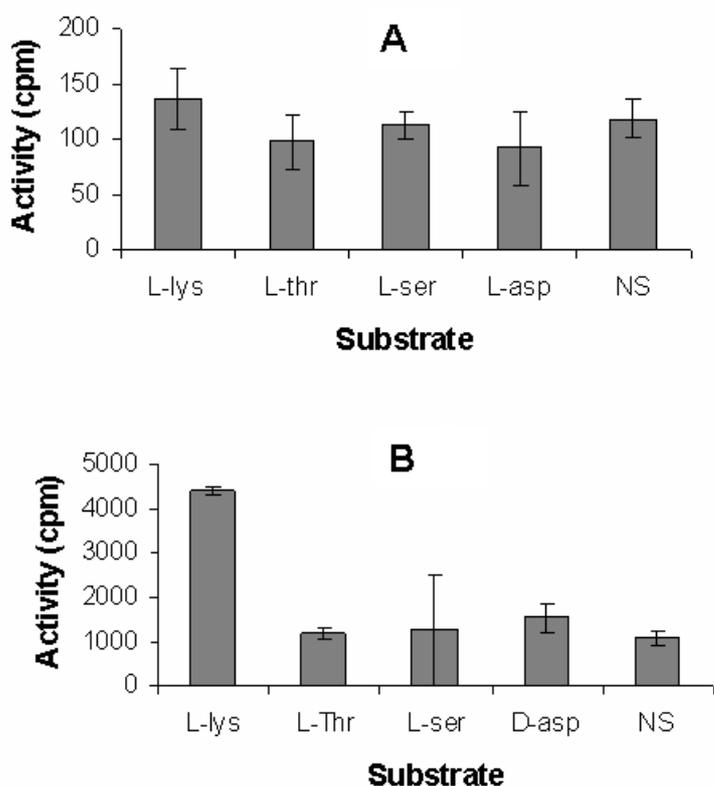
B – Exchange activity of his6-tagged EntF preparation 1 showing weak activation of L-lysine.

C – Comparison of unexpected activity of preparation 1 with that seen for the subsequent preparation highlighting the difference between weak activation of L-lysine and true exchange activity in the presence of L-serine. Assays were performed in triplicate as described in section 2.10.1, error bars represent +/- one standard deviation. NENS= no enzyme or aa substrate, NS= no aa substrate.

4.2.4.3 – Assessment of substrate specificity of *P. syringae* pyoverdine side-chain NRPS by ATP/³²PPi exchange assay

Each of the seven AT bi-domains purified, except 261, was subjected to either EA1 or EA2 in order to determine substrate specificity. In each assay reactions were set up in triplicate with either L-serine, L-lysine, L-aspartic acid or L-threonine as a substrate. Triplicate no-substrate controls and no-enzyme controls were also included. It was not deemed necessary to include any additional aa substrates, as each of the modules had already been confirmed as a pyoverdine synthetase by gene knockout (See section 3.10.2) and the side-chain of *P. syringae* pyoverdine has been shown to contain only the four aas that were tested in all substrate specificity assays [22, 23]. None of the AT bi-domains displayed the expected ATP/³²PPi exchange activity with their expected substrate, however in some instances unexpected low levels of activity were seen in the presence of L-lysine. Table 4.1 gives a summary of assays conducted and results for each bi-domain. Figure 4.7 shows representative assay results. A control assay was also run with L-lysine in the presence of no enzyme and no significant difference was seen between reactions containing L-lysine and those without (not shown).

One possible explanation for the lack of exchange activity seen was that the enzymes were co-



4.7 – Representative assay results for EA1 and EA2

A- Substrate specificity assessment of his-tagged 231 AT bi-domain using EA1. Assays were performed in triplicate as described in section 210.1, error bars represent +/- one standard deviation.

B- Substrate specificity assessment of his tagged 242 AT bi-domain using EA2 showing weak exchange activity in the presence of L-lysine only.

Assays were performed in triplicate as described in section 210.1, error bars represent +/- one standard deviation.

purifying with small amounts of the molecular chaperone DnaK; this phenomenon was previously observed during purification of the pyoverdine NRPS PvdD AT bi-domain from *P. aeruginosa*, and appeared to inhibit exchange activity [37]. To test this possibility, DnaK contamination was removed from two *Pspph* enzyme preparations (251 and 252) as per the previous PvdD study, by incubation at 37 °C in the presence of 2 mM ATP and 10 mM MgSO₄, followed by a second passage of Ni-NTA affinity chromatography [37, 143]. The re-purified proteins were then assayed for exchange activity as before, and were once again found to be inactive. Interestingly 252, which previously showed reproducible low levels of activity with lysine, no longer exhibited this activity following ATP treatment (Figure 4.8), indicating that such activity may be symptomatic of DnaK contamination. Since DnaK is known to associate with improperly folded or aggregated proteins [4, 24, 25] it follows that the unexpected lysine activation observed was a result of the protein in the assay being improperly folded.

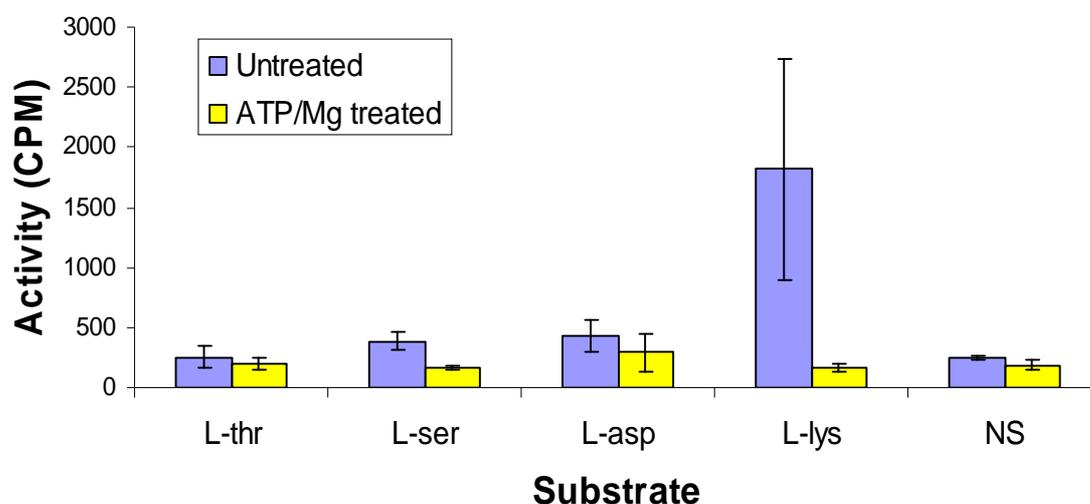


Figure 4.8 – Effect of ATP/Mg²⁺ treatment on ATP/³²PPi exchange activity of 252

his6-tagged 252 AT bi-domain (blue bars) was assessed for substrate specificity using EA1 and found to have weak exchange activity in the presence of L-lysine. The remainder of the purified protein was incubated at 37 °C with 2 mM ATP and 10 mM MgCl₂, then subjected to an additional passage of Ni-NTA affinity chromatography. Following ATP/Mg²⁺ treatment exchange activity was no longer observed in the presence of L-lysine (yellow bars).

4.3 – Exploring alternative protein expression strategies

Following the failure of *E. coli* based expression to yield active his6-tagged AT bi-domain, five additional strategies for expression and subsequent purification were investigated. The rationale behind these, as discussed at length in Section 4.4 of this chapter, was that the protein initially obtained, though soluble, may have been inactive due to partial or improper folding. Additional measures were therefore tested in an attempt to purify properly folded protein for subsequent assays.

Two new assay procedures, which did not require the use of radioactive isotopes, were also implemented. The impetus for developing new, non-radioactive assays was decreasing the time between production and assessment of proteins. Since EA1 and EA2 relied on having access to a short-lived stock of ^{32}PPi , it had previously been necessary to purify all proteins prior to substrate specificity assessment, in order to allow all proteins to be assessed without the need for ordering multiple, costly batches of radioisotope. Development of the new assays allowed assessment of proteins immediately following purification, decreasing the likelihood of stocks becoming denatured while in storage.

4.3.1- *Pseudomonas* species as an expression host

It stands to reason that the environment most conducive to correct folding of a protein is the protein's native environment, since this is where it has resided and evolved into its present state. Differences in osmolarity, translational machinery, pH and molecular chaperone function are all possible reasons that have been attributed for failure of a protein to fold correctly in a heterologous host [161, 163]. With this in mind it was decided to attempt overexpression of AT bi-domains in *Pseudomonas* hosts.

4.3.1.1 – Development of a plasmid based system for expression and purification of his6-tagged proteins from *Pseudomonas* species

Initial attempts at expression in *Pseudomonas* were carried out using the previously developed broad host range vector pMMB67EH [168]. For testing, the 231 bi-domain was sub-cloned

into this vector and initial expression and purification trials run using both *P. syringae* and *P. putida* as hosts. It was determined from these trials that pMMB67EH was not a good vector for expression and purification in these hosts for a number of reasons. The expression levels observed from this plasmid were very low, especially in *P. syringae*, making it difficult to obtain sufficient protein for characterisation assays. Furthermore, we have observed that the ampicillin resistance marker of pMMB67EH functions poorly in *Pseudomonas* species leading to difficulty in plasmid selection and maintenance. Possibly due to unusual supercoiling, this plasmid is also extremely resistant to cleavage with restriction endonucleases and difficult to purify in large amounts owing to its low copy number in *E. coli*. It was therefore decided to develop a new expression plasmid that allowed higher levels of protein expression in *Pseudomonas* species, had a higher copy number in *E. coli* and was more amenable to genetic manipulation.

4.3.1.1.1 Construction of pSX, a novel *Pseudomonas* expression vector

The plasmid constructed to address the shortcomings of pMMB67EH was based on the existing broad host range vector pUCP22 [98]. Primers were designed to amplify the backbone of pUCP22 such that the resulting PCR product lacked only the original multiple cloning site and contained terminal AgeI and SpeI restriction sites. A second set of primers was designed to amplify *lacIQ*, the *tac* promoter and MCS from the *E. coli* expression vector pUCX [99], flanked by the same restriction sites. The two PCR products were then digested, ligated together and transformed following standard molecular biological protocols. The resulting plasmid, designated pSX, was found to be suitable for IPTG-regulated expression of recombinant proteins in *Pseudomonas* species, had a high copy number in *E. coli* and was readily cleaved by restriction endonucleases. The gentamycin resistance marker of pSX allowed reliable selection and maintenance of the plasmid in *Pseudomonas* species.

4.3.1.1.2 – Optimisation of pSX-based expression of his6-tagged proteins in *Pseudomonas* species

For the purpose of initial expression trials, the 251 AT bi-domain was sub-cloned from pET28a+ into the newly developed pSX plasmid. The resulting construct was transformed into *P. syringae*, *P. putida* and *P. aeruginosa* by electroporation and protein induction profiles

were run for each of these hosts. Initial trials using a final concentration of 0.5 mM IPTG resulted in no visible expression of recombinant protein. For subsequent expression profiles in each host, 50 mL cultures were grown to an appropriate density (OD 600 = 0.5-1.0) and protein expression induced by addition of IPTG to a final concentration of 1, 2.5, 5 or 10 mM. Cultures were incubated for an additional 12-16 h after which analysis of total cellular protein by SDS PAGE was conducted. For each strain an uninduced control was also included. Results of these experiments (Figure 4.9) showed that higher levels of IPTG were required for induction of protein expression in *Pseudomonas* hosts than *E. coli* and that even when a final concentration of 10 mM IPTG was used, expression was still much lower than that seen in *E. coli* at a final concentration of 0.5 mM IPTG.

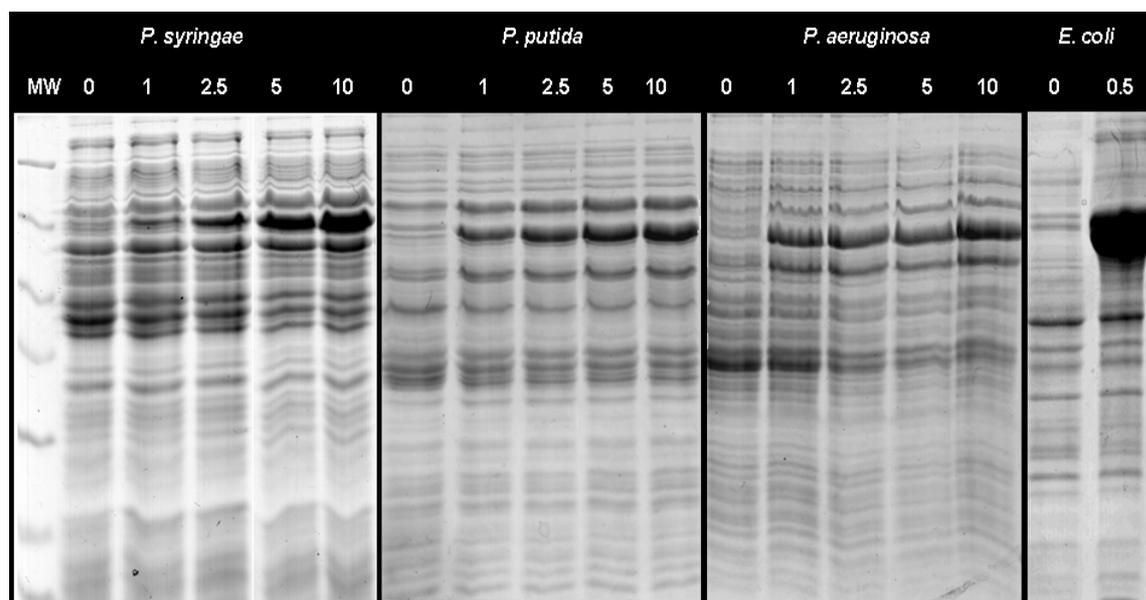


Figure 4.9 – Expression tests for pSX:251 in *P. syringae*, *P. putida*, *P. aeruginosa* and *E. coli*

50 mL cultures were grown to an OD 600 of 0.5-1.0 prior to induction of protein expression by addition of IPTG. Cultures were incubated for a further 12-16 h before analysis of total cellular protein by SDS PAGE. The final concentration of IPTG in mM is indicated above each lane. MW = Molecular weight marker.

4.3.1.2- Purification of his6-tagged proteins from *Pseudomonas* expression hosts

Following expression trials, AT bi-domains for 231 and 252 were sub-cloned from pET28a+ into pSX to give the plasmids pSX:231 and pSX:252. A new construct was also created for expression of the A-domain of Pspph1923 without any additional domains, this construct was designated pSX:23a. Plasmids were introduced into *P. syringae* and *P. putida* by electroporation and their expression tested using a final concentration of 5 mM IPTG, with results comparable to those described above (not shown). For purification of these domains from *P. syringae* and *P. putida*, 500 mL cultures were grown at 28 °C to an OD 600 of 0.5-1.0 mM. Protein expression was then induced by addition of IPTG to a final concentration of 5 mM and cultures were incubated for an additional 16-24 h. Cell harvest, lysis by French press, removal of insoluble debris and purification of his6-tagged protein was then conducted as previously described for *E. coli*. Purification of each of the proteins for which a pSX construct was available was attempted from both *P. syringae* and *P. putida*. Figure 4.10 shows SDS PAGE analysis of the successfully purified proteins. From *P. syringae* only 252 and 23a had sufficient yield for analysis by SDS PAGE. Greater yields of 231 and 251 were obtained from *P. putida*, but in each case contaminating protein species were present in the final preparation of desalted protein.

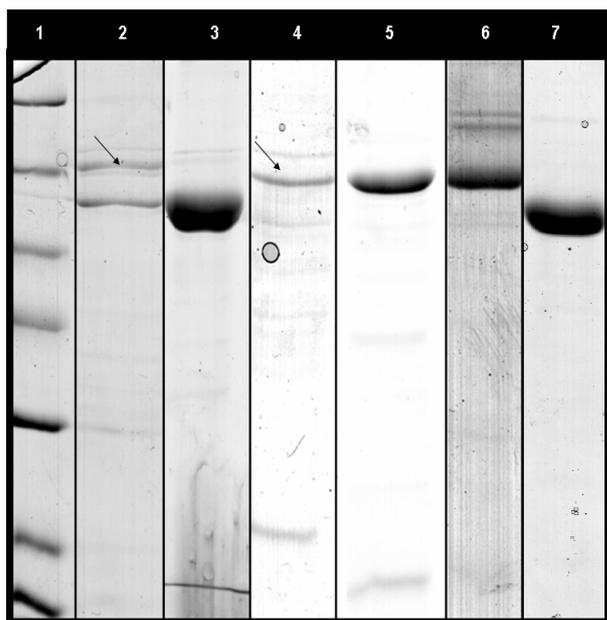


Figure 4.10 – His6-tagged A and AT proteins purified from *P. syringae* and *P. putida*

1) Molecular weight marker. 2) 231 purified from *P. putida*. 3) 23a purified from *P. putida*. 4) 251 purified from *P. putida*. 5) 252 purified from *P. putida*. 6) 252 purified from *P. syringae*. 7) 23a purified from *P. syringae*. Where contaminating species are present, the band corresponding to the expected size of the target protein is indicated with

4.3.1.3 – Development of a non radioactive assay for assessment of A-domain substrate specificity

In order that purified proteins might be rapidly assessed for substrate specificity, without the need for radioisotope, an absorbance based 96 well plate assay for A-domain specificity was developed. The assay was based on the Enzcheck[®] pyrophosphate assay kit, which is commercially available from Invitrogen. This assay, known as the PPi release assay, has previously been used to determine substrate-dependent release of PPi from ATP by A-domains in a 1 mL cuvette format [78, 160, 169, 170]. This assay relies on the fact that aa adenylate formation by an A-domain in the presence of a cognate aa results in the release of PPi, levels of which are indicative of substrate preference [111]. PPi levels cannot be measured directly however inclusion of two additional enzymatic steps and one additional substrate in the reaction results in the formation of a compound which can be assayed spectrophotometrically. The first enzyme is inorganic pyrophosphatase (IP), which cleaves PPi into two units of inorganic phosphate (Pi). These Pi intermediates are then covalently linked to the nucleotide analogue 2-amino-6-mercapto-7-methylpurine ribonucleoside by the action of a second enzyme, purine nucleoside phosphorylase, resulting in a shift of absorbance maximum from 330 to 360 nM proportional to the original levels of PPi in a sample.

To ascertain whether or not the Enzcheck[®] assay was suitable for quantification of PPi in a 96 well plate format, reactions were scaled down from 1 mL to 200 μ L and a PPi standard curve set up with final concentrations ranging from 0-40 μ M PPi (sodium pyrophosphate). It was found that in a 96 well plate format, the assay produced a linear relationship between [PPi] and A_{360} for concentrations ranging from 0 to 20 μ M PPi (Figure 4.11A). To determine whether the assay was suitable for determination of A-domain substrate specificity, a positive control experiment was run using the EntF A-domain. In this experiment A-domain reactions with one of four substrates, as well as a no-substrate negative control, were set up in triplicate in a final volume of 125 μ L. These reactions were allowed to proceed for 60 min before addition of Enzcheck[®] assay components. The results of this end-point assay were consistent with the previously characterised role of EntF as a serine activator (Figure 4.11B), thus validating the assay (henceforth referred to as RA1) for assessment of A-domains substrate specificity.

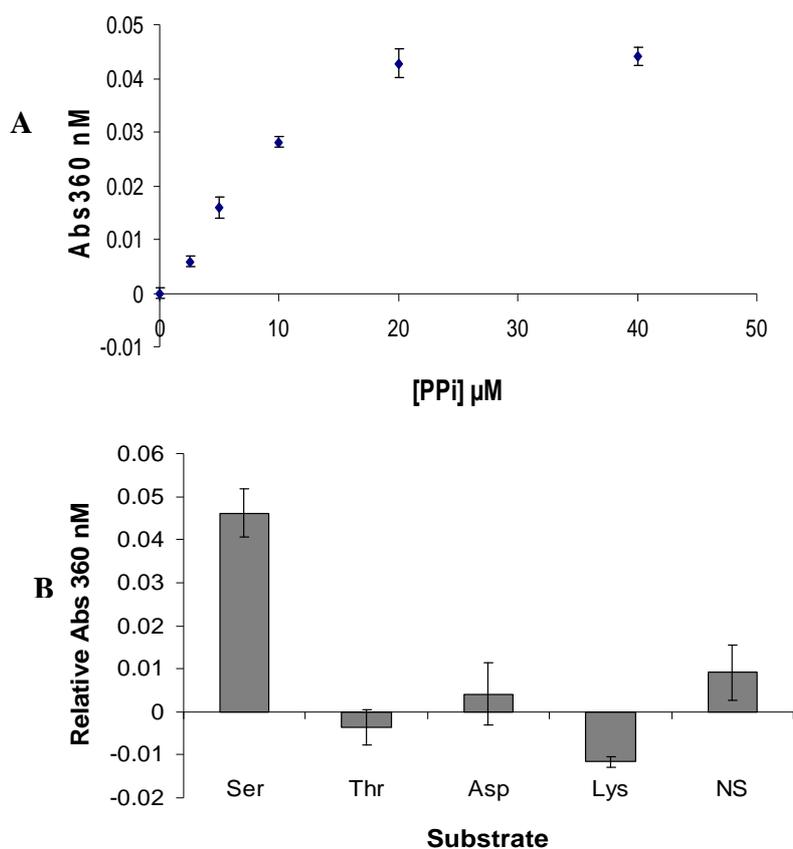


Figure 4.11 – Validation of absorbance-based PPI release assay RA1 for determination of A-domain substrate specificity.

A- RA1 standard curve for reactions incubated and quantified in a 96 well plate. Reactions for each PPI concentration were set up in triplicate, in a final volume of 200 μ l as described in Section 2.10.2. Following incubation at room temperature for 20 min, absorbance at 360 nM was measured using a microplate reader. Error bars represent +/- one standard deviation.

B- Assessment of EntF substrate specificity using RA1. RA1 measures substrate preference by determining levels of PPI released in the presence of a particular aa substrate candidate. A-domain reactions for each substrate were set up in triplicate in a 96 well plate as described in Section 2.10.2; NS= no substrate. Error bars are presented as +/- one standard deviation. Absorbance values are given relative to control reaction containing no enzyme or substrate control.

A second, continuous, version of this assay was tested by adding the components of the A-domain and Enzcheck[®] reactions to wells simultaneously. These reactions were initiated by addition of the EntF A-domain and the change in A_{360} was measured continuously for 60 min, followed by a final reading at 220 min. In this continuous assay, two substrates (L-ser and L-thr) as well as a no substrate control were assessed in parallel, with the preferred substrate serine included at two concentrations (10 mM and 100 mM). Reactions including 200 μ M PPI

were also included. The purpose of this experiment was to ascertain whether the reaction was reaching saturation before the endpoint of 60 min and also to determine whether increasing substrate concentration would increase reaction rate. As shown in Figure 4.12, reactions were not reaching saturation point (as defined by the 200 μ M PPi control reactions) and increasing the concentration of the aa substrate 10-fold had little effect on the overall reaction rate. Taken together, the results summarised in Figures 4.11 and 4.12 indicated that the absorbance-based PPi release assay adapted for 96 well plate use was suitable for determination of A-domain substrate specificity.

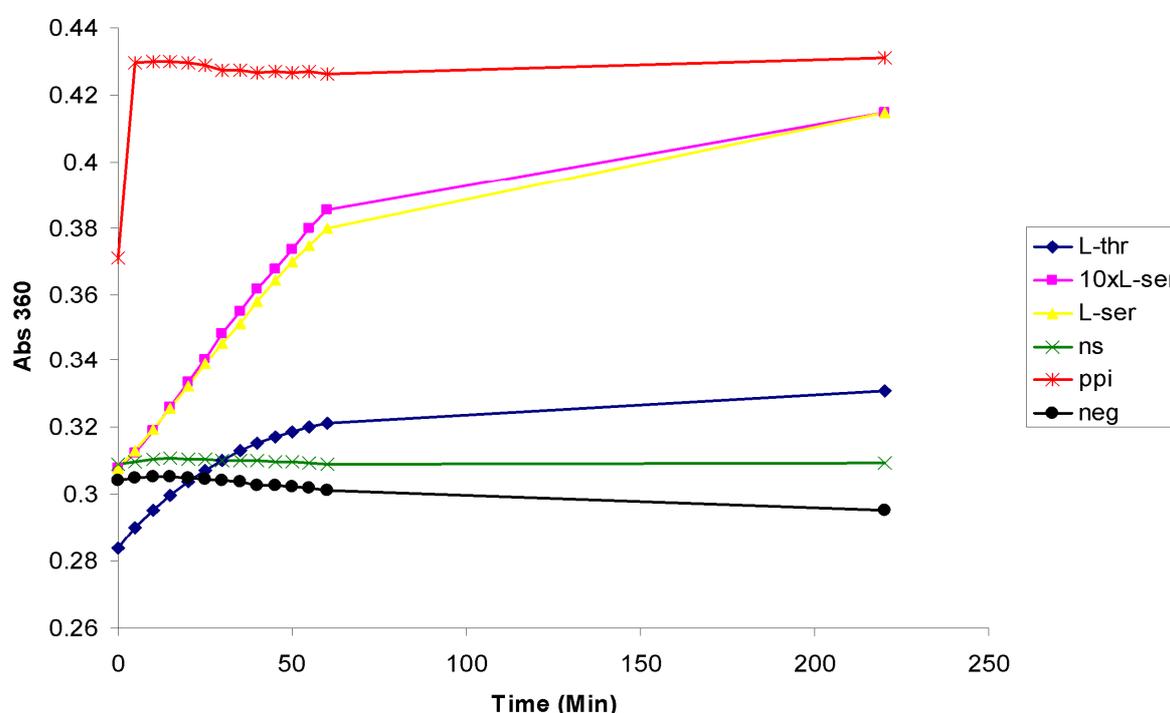


Figure 4.12 – Monitoring continuous release of PPi by EntF A-domain

Reactions were set up in a 96 well plate as for the PPi release RA1 (2.10.2), except Enzcheck[®] components were added prior to initiation of reactions. Immediately following addition of EntF A-domain the plate was transferred to a microplate reader set to measure Abs 360 every 60 s; for clarity only every fifth point is shown. Neg = no A-domain or aa substrate, NS = no substrate, PPi = 200 μ M PPi saturation control, 10 x L-ser = L-serine at 20 mM, L-ser/L-thr = aa substrates at 2 mM.

4.3.1.3.1 – Design and testing of an alternative PPi release assay not based on a commercially available kit

While RA1 was a convenient method for rapidly assessing substrate specificity of A-domains, it relied on reagents supplied in the Enzcheck[®] kit, making the assay relatively expensive to

run. To reduce the cost associated with running multiple assays, an alternative version, RA2 was developed. This assay was effectively the same as the malachite green reagent assay for A-domain assessment described by McQuade *et. al* [111], except with modifications to the malachite green reagent [171] as described in 2.10.2. For RA2, PPi release by an A-domain proceeds as for RA1 and, as before, PPi is cleaved into two equivalents of Pi by IP. RA2 differs from RA1 only in the means of detection of Pi levels. In RA2, reactions are terminated by addition of acidic ammonium molybdate solution to each well, followed by addition of malachite, or brilliant green dye. A green complex between the dye, molybdate ions and free Pi is then formed. The intensity of green colouration is dependent on the concentration of free Pi and can be measured via absorbance at 650 nM. The effective range of RA2 was determined to be 0.2-25 μ M free PPi (Figure 4.13A). The suitability of RA2 for assessment of A-domain specificity was assessed, using the previously characterised NRPS protein BpsA[80] as a positive control. Results of BPSA assessment with RA2 (Figure 4.13B) were consistent with previous experimental results showing BPSA to be an activator of L-gln [80]. A detailed description of BPSA is given in Section 6.1.

4.3.1.4 – Substrate specificity assessment of domains purified from *Pseudomonas* species

Each of the his6-tagged domains purified from *P. syringae* and *P. putida* was assessed for substrate specificity using RA1 or RA2. Each purified protein was assessed for PPi release activity in the presence of L-lysine, L-serine, L-aspartic, L-threonine and no aa substrate, as described in Sections 2.10.2 and 2.10.3. In no case was there any significant increase in activity in the presence of any aa substrate, suggesting that purification from *Pseudomonas* species had not produced active protein.

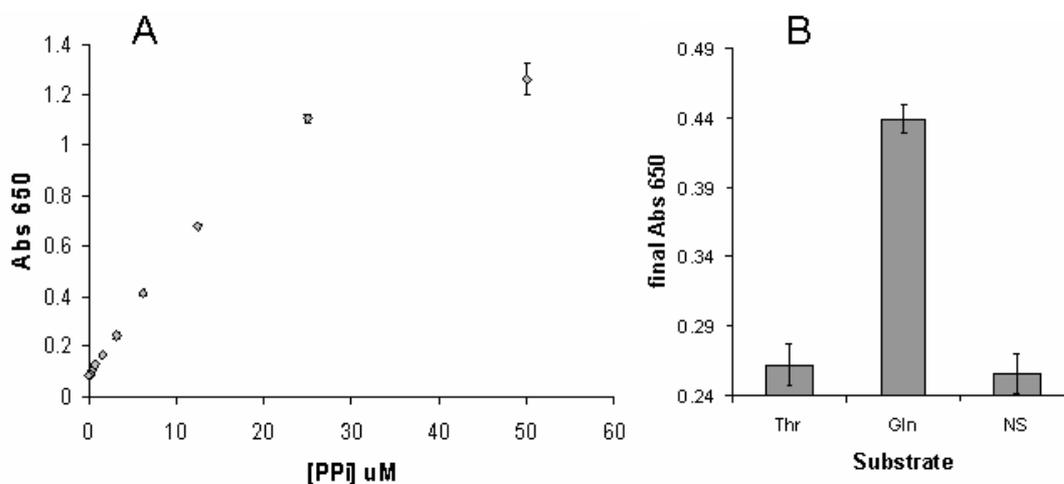


Figure 4.13 – Validation of A-domain specificity assay RA2

A – A two-fold dilution series from 0.2-50 μM PPi, in 1 x RA2 reaction buffer, was set up in triplicate in a 96 well plate. Following incubation at RT for 5 min to allow the pyrophosphatase reaction to proceed, RA2 detection reagents were added and absorbance at 650 nM measured using a microplate reader.

B – Purified BpsA, a characterised L-gln specific NRPS, was incubated with 10 mM of each of the indicated aa substrates in 1 x RA2 reaction buffer for 30 min at 28 °C. RA2 detection reagents were then added and absorbance at 650 nM measured using a microplate reader. Reactions were set up in triplicate for each substrate. For both A and B, error bars represent +/- one standard deviation. NS = no aa substrate.

4.3.2 – Purification and assessment of proteins produced in *E. coli* co-expressing molecular chaperone proteins

The most common cause of protein inactivity following over-expression and purification is incorrect folding [5, 6, 28], a process dependent on not only the primary sequence of the over-expressed protein but also on the action of various chaperone proteins that guide the folding of a nascent protein into its correct, active conformation [29, 34-36]. One recently developed strategy for improving the conformational quality of proteins over-expressed in *E. coli* is the co-expression of the molecular chaperone proteins DnaK, DnaJ, GroESL and GrpE; this process has been shown to be effective in a number of instances for improving both the solubility and activity of recombinant proteins [25, 29-32]. In order to investigate the possibility that co-expression of molecular chaperone proteins might promote correct folding of recombinant *Pspph* A- and AT-domains in *E. coli*, four chaperone co-expression strains (CH1-4) were used as expression hosts. As outlined in Table 2.7, the chaperone co-expression strains harbour plasmids, allowing IPTG regulated co-expression of different combinations of the chaperone proteins DnaK, DnaJ, GroESL, ClpB and GrpE in BL21(DE3) cells. The use of

these strains for improving solubility of recombinant proteins over-expressed in *E. coli* has been well studied [106, 172, 173], although not previously for NRPS protein expression.

4.3.2.1-Assessment of chaperone strains for improvement of NRPS solubility in *E. coli*

An initial trial using the A-domain of Pspph1923 (23a) was conducted to assess the suitability of CH1-4 for improving yield of soluble recombinant NRPS proteins. Expression of 23a was carried out in each of the four chaperone expression strains as described in Section 2.9.1.4. Following expression soluble fractions were separated (Section 2.9.2.1) then analysed by SDS PAGE and the amount of soluble 23a, relative to total 23a protein, was determined by volumetric analysis using the programme Imagequant[®]. The results of the solubility trial, shown graphically in Figure 4.14, indicated that expression using CH4 as a host resulted in the highest levels of soluble 23a, with CH2 a distant second. For Pspph1923 it was necessary to use the A-domain without its cognate T-domain in solubility trials, as the recombinant AT bi-domain protein was found to run at the same location as DnaK on SDS PAGE gels (DnaK being expressed at high levels in chaperone strains).

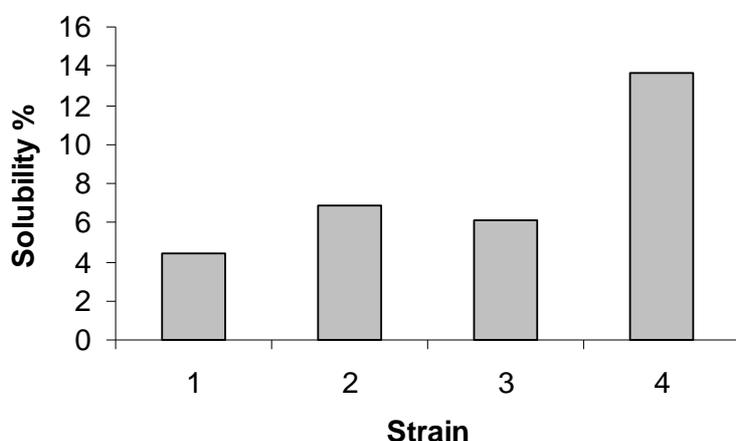


Figure 4.14 – Solubility trial using Chaperone strains 1-4

23a was expressed in each chaperone strain under identical conditions. Soluble and insoluble fractions were then analysed by SDS PAGE and % solubility of 23a determined by volumetric analysis.

4.3.2.2- Substrate specificity assessment of 23a and 251 purified from chaperone strains two and four

Following initial trials to assess the suitability of each chaperone co-expression host, 23a and 251 were purified from both CH2 and CH4. Expression and purification was conducted as described in Section 2.9 and purified domains were assessed for substrate specificity using RA1. No activity was seen in the presence of any substrate for either 23a or 251 purified from CH2. For both 23a and 251 purified from CH4, weak activity in the presence of L-lys was observed. Although lysine is the predicted substrate of 23a, the activity seen was extremely weak compared to that of recombinant EntF with serine (included as a positive control) so is not likely to represent true adenylation activity, rather the non-specific activation that had previously been observed with putative inactive A-domains in the presence of L-lys.

4.3.3- Refolding of a denatured AT bi-domain isolated from inclusion bodies

When isolating recombinant proteins prone to inclusion body formation, an alternative to purification of soluble protein is to denature insoluble proteins trapped in inclusion bodies and subject them to refolding treatment [38, 39]. There are a number of methods for refolding of denatured proteins isolated from inclusion bodies, all of which operate on the same basic principle[174]. Inclusion bodies are first solubilised in a harsh denaturant, such as SDS, guanidium HCl or urea, and the concentration of this denaturant is then gradually decreased. As the denaturant concentration is reduced, denatured proteins are forced to adopt a soluble conformation in order to remain in solution. For any globular cytoplasmic protein, refolding has (at least theoretically) the potential to regenerate the native conformation of a protein, which should be the most energetically favourable state in solution [172, 175]. A particularly convenient protocol for refolding of affinity tagged recombinant proteins purified by column chromatography is on column refolding [176, 177]. In this process denatured, affinity-tagged proteins are bound to a column in the presence of denaturant and a decreasing concentration gradient of denaturant is applied to the column. Protein bound to the column is refolded as the concentration of denaturant is slowly decreased to zero. The process is thought to be assisted by immobilisation on the column matrix, which prevents the formation of aggregates [174].

4.3.3.1 – Isolation, denaturation and on-column refolding of recombinant NRPS domains sequestered in inclusion bodies

A detailed description of the refolding process is given in Section 2.9.2.3, briefly; a 500 mL expression culture was set up for 252 under standard conditions (induction at 37 °C with 0.5 mM IPTG) in order to bring about inclusion body formation. Following lysis, inclusion bodies containing the target proteins were collected by centrifugation, dissolved in binding buffer containing 8 M urea and applied to a Ni-NTA column. Washing steps were then carried out using standard buffers with urea added to a final concentration of 8 M. Following washing, a linear gradient of urea from 8 to 0 M in binding buffer was applied to the protein bound to the column over a period of 24 h. Any soluble protein remaining on the column was eluted with standard elution buffer that did not contain urea. A schematic representation of the refolding process and apparatus is given in Figure 4.15. Using this process it was possible to regenerate soluble recombinant protein for 252 (Figure 4.16). Refolded 252 was assessed using both RA1 and EA2, but did not show any adenylation activity, indicating that it had not been refolded into an active conformation.

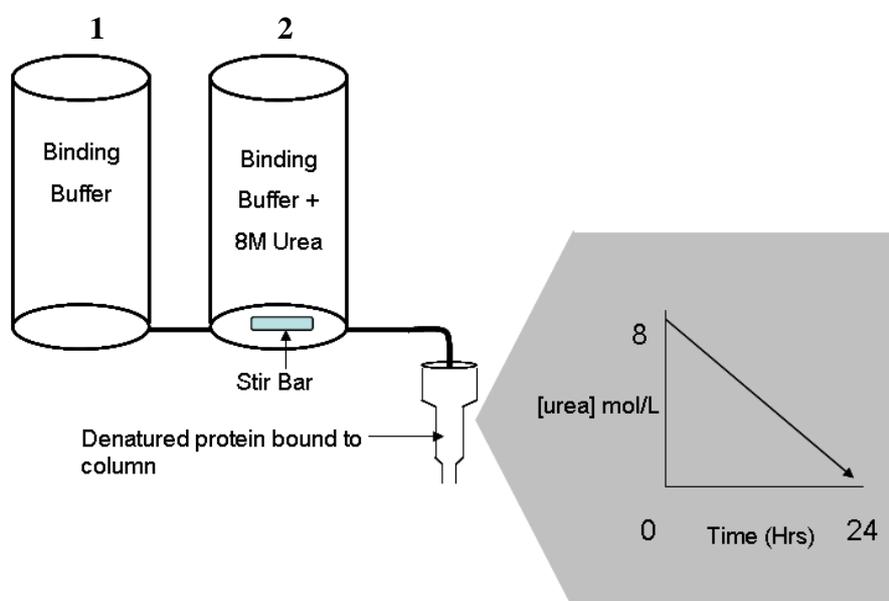


Figure 4.15 – Apparatus for on column refolding of denatured proteins

The column was connected to a peristaltic pump, so that as buffer flowed from chamber 1 to chamber 2 the urea concentration decreased in a linear fashion. Flow rate was adjusted such that concentration of urea on the column decreased from 8 to 0 M over 24 h. Any refolded protein remaining on the column was then eluted as before.

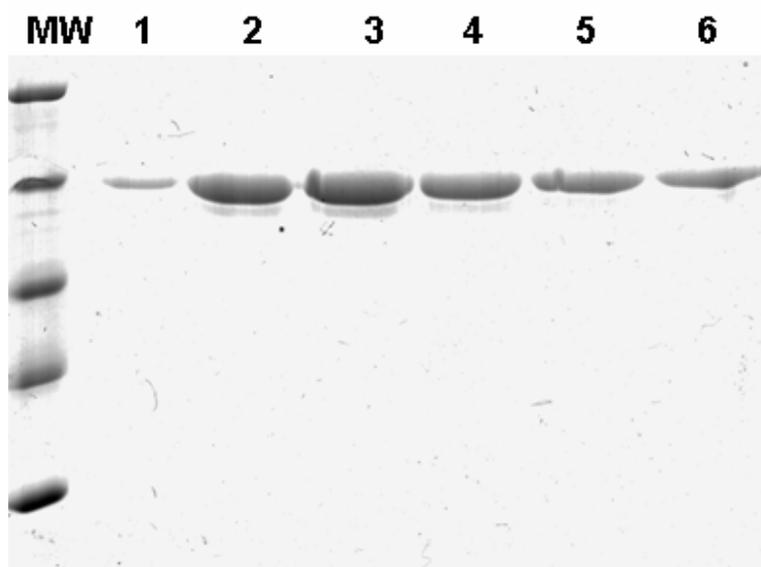


Figure 4.16 – SDS Page analysis of soluble 252 recovered from denatured inclusion bodies

MW = Molecular weight marker. 1-6 = 10 μ l from 1.5 mL fractions of refolded 252 in the order they were eluted from the column. Fractions were subsequently pooled and, following buffer exchange, assessed for substrate specificity using RA1 and EA2.

4.3.4 – Exploration of alternative domain boundaries and constructs for substrate specificity assessment

As outlined in the previous sections, active AT bi-domain protein was not recovered for any of the seven *P. syringae* pyoverdine side-chain synthetase modules. A construct expressing only the A-domain of *Pspph1923* (23a) also failed to produce active protein. In order to investigate the possibility that this lack of activity was due to inappropriate domain boundary definition in the original constructs, new constructs were produced in which the domain boundaries were altered. Two different domain expression platforms were constructed and tested as outlined in the following sections.

4.3.4.1 – Expression and purification of entire NRPS side-chain synthetase proteins for substrate specificity assessment

The entire gene for *Pspph1925* was cloned into both pET28a+ and pSX. Initially the pSX construct was transformed into the corresponding side-chain synthetase knockout strain of *P. syringae* to determine whether functional protein was being produced. As described in Section 3.3.3, pSX:1925 was found to restore pyoverdine synthesis in *PspphΔ1925*, establishing that functional *Pspph1925* protein was being produced. A particularly interesting feature of complementation of *PspphΔ1925* by pSX:1925 (Section 3.3.3), was that restoration of pyoverdine synthesis was only achieved on agar plates grown at 18 °C when IPTG was absent from the growth medium, suggesting that when expressed at high levels from pSX, *Pspph1925* was not being produced in an active conformation.

Purification of the entire side-chain synthetase for both *Pspph1923* and *Pspph1925* was attempted from both *E. coli* and *P. syringae*. For *E. coli* purification, expression at 18 °C in betaine/sorbitol media was used, as previously described. For purification from *P. syringae*, in addition to the expression protocol described for this organism (Section 2.9.1.3), a plate based purification protocol was also attempted. For the plate-based protocol, four iron-limited media plates were streaked to confluency with *PspphΔ1925::pSX:1925* and grown at 18 °C for five days. After five day's growth there was clear evidence of pyoverdine production (See Figure 3.4c for representation), indicating that functional his6-tagged *Pspph1925* protein was being produced. For purification, cells were scraped from the plates, resuspended in binding buffer and subjected to the usual lysis and purification steps. However, purification attempts from both *P. syringae* and *E. coli* did not yield discernable amounts of recombinant NRPS protein, as assessed by SDS PAGE and Biorad DC assay, so substrate specificity assessment was not possible.

4.3.4.2- Alternative constructs for production and purification of AT bi-domains

After it was determined that purification of entire NRPS proteins from *E. coli* or *P. syringae* could not be achieved, alternative strategies were considered for production of AT bi-domains. These were guided by the recent publication of a paper in which the substrate specificity of an AT bi-domain from the *P. syringae* NRPS protein *syrE* was determined [78]. The new

constructs were modelled on the AT bi-domain overexpression construct described in this paper. These constructs, and the proteins they expressed are henceforth indicated using the prefix xAT. For example, the construct pXAT:252 contains the first module from the gene *pspph1925* with new domain boundaries defined as described below. The protein expressed from this construct is referred to as xAT251.

4.3.4.2.1- Design and construction of pXAT expression vectors

In order to define the new domain boundaries for pXAT constructs, the protein sequence of the SyrE AT bi-domain successfully characterised in [78] was aligned against each of the *P. syringae* pyoverdine side-chain NRPS sequences using the alignment function of Vector NTI[®]. Primers were then designed to amplify the corresponding nucleotide sequence and domains were cloned into pET28a+ as previously described. In all cases the downstream restriction site used for cloning was NdeI and the upstream site was HindIII. Confirmation of constructs by sequence and testing protein expression testing was performed as previously described.

4.3.4.2.2- Purification and substrate specificity assessment of XAT bi-domains

An alternative expression protocol to those previously described in this chapter was used for the purification of xAT bidomains. This expression protocol, which is described in detail in Section 2.9.1.5, made use of auto-induction media [178], with growth conducted at 16 °C for 36-48 h. Using this auto-induction protocol, good yields of soluble xAT251 and 241 were obtained with the resulting protein preparations being of better purity than those for the original AT bi-domain constructs using betaine/sorbitol expression. The substrate specificities for xAT-domains purified for 251 and 241 were assessed using RA2 and in each case no specificity for any of the aas tested was observed. For x251, weak activation of lysine was again observed. It has been noted previously by structural biology groups that inclusion of low concentrations of a mild denaturant in protein assays can increase activity, possibly by prevention of protein aggregate formation (Dr TT Caradoc-Davies, *pers. comm.*). Given that weak activity in the presence of L-lysine was repeatedly observed for otherwise inactive NRPS modules, it is possible that lysine was promoting enzyme activity by acting as a weak denaturant. The possibility that L-lysine was promoting activity of purified recombinant A-

domains was investigated by repeating Assays for x251 in the presence of various concentrations of L-lys, with either L-ser or L-thr as a co-substrate. As illustrated in Figure 4.17, addition of L-lys to reactions did not promote activation of either L-ser or L-thr by x251 (the latter being the predicted substrate for this construct), although it did result in elevated PPI release by xAT 251 independent of any co-incubated aa substrate. Table 4.1 gives a summary of all purification procedures described in this chapter and subsequent assay results.

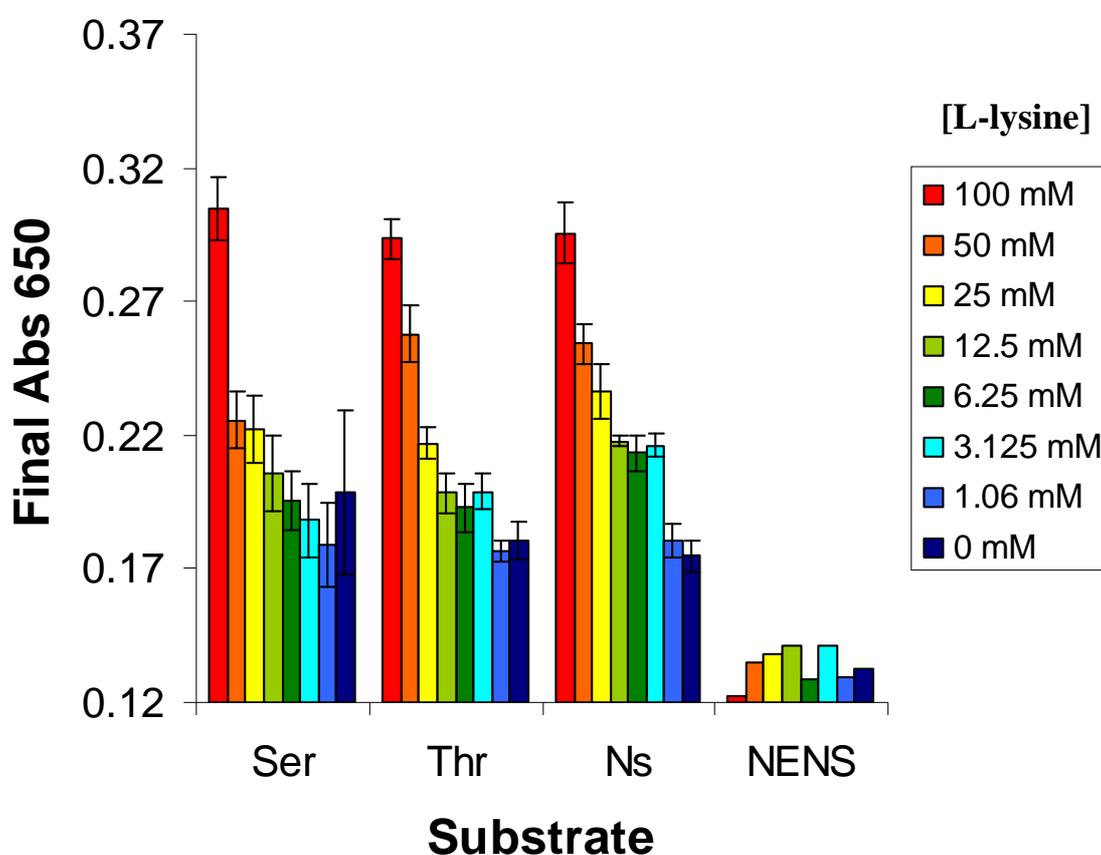


Figure 4.17 – Assessment of L- lysine concentration on PPI release activity of x251

Purified x251 was incubated with 10 mM of indicated aa substrates in 1 x RA2 reaction buffer for 30 min at 28 °C. RA2 detection reagents were then added, with absorbance at 650 nM measured using a microplate reader. Reactions were set up in triplicate for each substrate, for no-enzyme controls one reaction was set up for each L-lys concentration. Error bars represent +/- one standard deviation. NS = no aa substrate, NENS= no x251 or substrate. Key indicates the final L-lys concentration present in a given reaction.

Protein and expected substrate	Purification procedure(s)	Assay procedure(s) used	Results
231: L-lysine	BS, PP, PS	EA1	No activity with any substrate
23a: L- lysine	PS, PP, CH	RA1(PS, PP) RA1 (CH)	No activity with any substrate Weak activity with L-lysine
241: L-aspartic acid	BS, CH, xAT	EA2 (BS, CH) RA2(xAT)	No activity with any substrate No activity with any substrate
242: L-threonine	BS	EA2	Weak activity with L-lysine
251: L-threonine	BS, BS-ATP PS, PP, CH xAT	EA1 (BS) EA1 (BS-ATP) RA1 (PS, PP, CH) RA2(xAT)	No activity with any substrate No activity with any substrate No activity with any substrate Weak activity with L-lysine
252:L-serine	BS, PS, PP, CH, RF	EA1(BS) EA1(BS-ATP) RA1 (CH) EA2 (RF)	Weak activity with L-lysine No activity with any substrate Weak activity with L-lysine No activity with any substrate
261: L-aspartic acid	BS	Not assessed	-
262 L-serine	BS	EA1	Weak activity with L-lysine

Table 4.1 – Summary of purification procedures and assay results

For purification procedures: **BS** = betaine sorbitol, **BS-ATP** = Betaine sorbitol followed by ATP/Mg treatment and a second passage of Ni-NTA affinity chromatography; **CH** = purified from chaperone strains 2 and 4; **PP** = purified from *P. putida*; **PS** = purified from *P. syringae*; **RF** = Refolded from denatured inclusion bodies; **xAT** = Alternative domain boundaries investigated, autoinduction protocol used for expression. For Assay procedures: **EA** = ATP/³²PPi exchange assay; **RA** = PPi release assay. Bracketed letters after assay procedure indicate which purification/ assay combination gave a particular result.

4.4 – Discussion

This chapter describes intensive efforts to biochemically determine the substrate specificities of seven A-domains derived from the five NRPS enzymes proposed to synthesise the aa side-chain of pyoverdine in the bacterium *Pseudomonas syringae* pv *phaseolicola* 1448a. In spite of rigorous and systematic investigation, it was not possible to derive active protein for substrate specificity assessment for any of the seven domains investigated. There are a number of possible reasons for the failure to derive active protein, the feasibility of which is now discussed.

Before any conclusions can be drawn regarding the activity or inactivity of the proteins assessed in this chapter, it is necessary to ensure that the assays employed were suitable for determination of substrate specificity. There were four assay procedures used to determine

substrate specificity: two ATP/³²PPi exchange assays (EA1 and EA2) and two PPi release assays (RA1 and RA2). Variant forms of each of these assays have been previously described for the determination of A-domain specificities [78, 80, 111, 160], and validation experiments were carried out to identify potential problems that may have resulted from reagents, equipment and procedures specific to our laboratory. EA1 was validated using a cell lysate positive control experiment, the results of which demonstrated that this assay was effectively recovering and measuring ³²P-ATP formed as a result of enzymatic reactions. For EA2, RA1 and RA2, positive control experiments went a step further and employed a previously characterised NRPS protein as a positive control. For each assay, the results of the positive control experiments were consistent with the previously described results. These data suggest that the failure to detect any activity stemmed from inactivity of the purified proteins, and not from deficiencies in the various assay procedures.

A possible reason for the inactivity of purified proteins is the presence of a small molecule inhibitor, either in the protein preparation itself or in one of the components of the subsequent assay. All examples of NRPS A-domain inhibitors thus far characterised are aminoacyl-AMP analogs that are thought to possess inhibitory activity by virtue of their similarity to the native adenylated substrate of an A-domain. Such inhibitors are generally specific to a particular A-domain or set of A-domains with similar specificity [179, 180]. The likelihood that inhibitors specific to all six modules tested – but not to either the recombinant BpsA or EntF positive controls – seems remote, especially given that expression and purification were performed in three different host organisms (*E. coli*, *P. putida* and *P. syringae*). The activity of the EntF positive control is especially relevant to the putative serine activating domains 252 and 262, as any inhibitors specific to serine activating A-domains would be expected to inhibit the EntF A-domain also.

Another possible factor to consider is the definition of domains used to derive over-expression constructs. Initially, domain boundaries were defined using the TIGR online NRPS analysis tool. AT bi-domains and A-domains defined in this fashion were used for construction of *P. syringae* and EntF over-expression constructs. In order to verify that all structural elements were present, aa sequence alignment of A-domains, as defined for use in over-expression

constructs, was performed against the A-domain of GrsA (pdb:1AMU) [25]. Visualisation of the resulting sequence alignments using DeepView [181] revealed that in each case all of the structural elements believed to be necessary for function of the A-domain were present in the final over-expression construct. Although the recombinant EntF construct was active, the domain boundaries selected might not have been optimal for the Pspph enzymes; and so an alternative set of AT bi-domain over-expression constructs was also designed and tested. Domain boundaries for these constructs were defined by protein sequence alignment against an AT bi-domain from *P. syringae*, which has successfully been used in substrate specificity determination experiments [78]. The AT-domains used for this second set of constructs were still inactive, despite good solubility.

While there are no immediately obvious reasons why the domain boundaries chosen should have precluded production of active recombinant A- or AT-domain proteins, the possibility can not be completely discounted. Even if an exhaustive set of domain boundaries were tested, the possibility that the A-domains in question are incapable of functioning outside of their native setting would remain. While the domains of NRPS are certainly autonomous in some instances [1, 30, 169], there is no way of knowing if this dogma holds true for all, or even the majority of NRPS A-domains. The number of published examples in which A-domain specificity has been successfully determined biochemically is very small, and the specificity of the majority of A-domains assigned in public data bases is based on *in silico* prediction and/or determination of product structure. As failed attempts at A-domain characterisation are seldom published ([182] being a rare example of this), it is difficult to know how prevalent a phenomenon A-domain inactivity *in vitro* actually is.

What is particularly surprising about the characterisation attempts reported here is that purified recombinant domains were inactive, in spite of having good solubility. Up until recently, the accepted paradigm for purification of recombinant proteins has been that solubility is a good indicator of activity, and that – especially for cytoplasmic proteins such as NRPS – proteins present in the soluble fraction of cell lysate are expected to be in their native, active conformation [161, 172, 183, 184]. However, increasing evidence suggests that solubility is not always a good indicator of conformational quality, and that for a given protein many

misfolded but soluble conformations may exist [185-187]. Since the native conformation of a protein should represent the lowest energy arrangement of all constituent atoms in a particular solution [175], inactive soluble conformations can be thought of as local energy minima that a polypeptide can become trapped in, if a cell's protein folding machinery is overwhelmed [188]. A graphical representation of this principle is given in Figure 4.22.

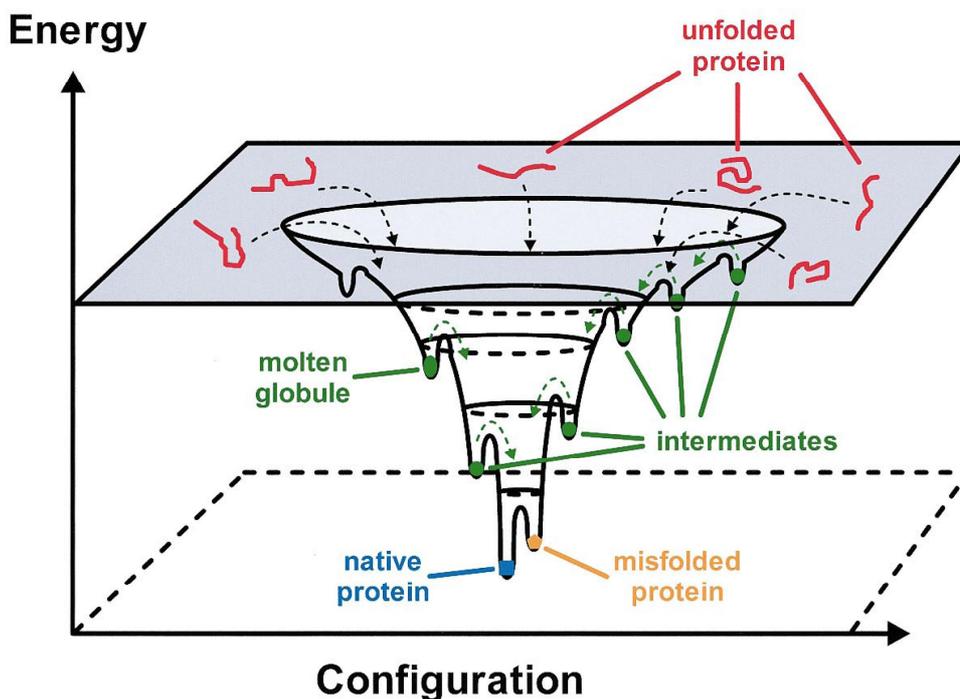


Figure 4.18 – The relationship between energy and conformation of proteins in the cytoplasm

The native, active conformation of a protein is presumed to be at its lowest energy state in solution; however, there are thought to be a number of alternative, inactive soluble conformations which a protein can adopt. These conformations are represented as local energy minima in which a protein can become trapped without proper guidance from molecular chaperone proteins. Figure reproduced from [188].

It is possible that the *P. syringae* NRPS proteins described in this chapter are only able to adopt their active conformation when expressed at low levels in a cell. Evidence for this can be found in the complementation experiments described in Sections 4.3.4.1 and 3.3.3. In these experiments the entire *Pspph1925* gene was cloned into pSX and the resulting construct used to successfully complement the corresponding *P. syringae* NRPS knockout strain, as

evidenced by restoration of fluorescence and ability to survive severe iron starvation. However, when IPTG was present in the medium – even under severe iron starvation, when survival of the bacterium depended on production of active 1925 protein – no enzyme activity was apparent. This suggests that elevation in expression levels as a consequence of IPTG addition may have resulted in excessive levels of nascent protein exceeding the folding capacity of molecular chaperone proteins. Even low levels of functional 1925 would presumably have allowed the bacterium to survive iron starvation, and the evidence suggests that once the cell's capacity for folding has been exceeded, all of the protein produced ended up in an inactive conformation. If this hypothesis is correct then it is likely that the Psp^{ph} NRPS domains are particularly prone to misfolding when over-expressed at high levels.

In light of this evidence, the most feasible explanation for the consistent inactivity of proteins observed in this chapter is that the expression levels exceeded folding capacity, even though numerous steps were taken to avoid these phenomena. Consequently, any soluble proteins that were expressed and purified were likely in an inactive conformation, and therefore unusable for substrate specificity determination.

Chapter 5 – Probing C-domain specificity using the pyoverdine biosynthetic NRPS of *Pseudomonas aeruginosa* as a model system.

5.1 – Introduction

Pyoverdine biosynthesis in *pseudomonads* provides a unique model system for conducting module swapping experiments with the aim of developing successful strategies for NRPS manipulation. This chapter describes the use of this model to investigate the specificity of NRPS C-domains. The majority of hands-on bench work was conducted by Honours student Mark Calcott following my research design and under my supervision, based on an original concept provided by Dr David Ackerley. However, with the exception of acknowledged Figures, the presentation and analysis of data in this chapter is my own.

5.1.1 – Pyoverdine synthesis as a model system for module swapping experiments

As described in Chapter 3, pyoverdine is a siderophore synthesised by NRPS enzymes in fluorescent *Pseudomonas* species. There are three features of pyoverdine synthesis that are particularly well suited for studying NRPS manipulation. First, pyoverdine is fluorescent and coloured. This fluorescence and colour is a property of the chromophore group of pyoverdine, and, as evidenced by the diversity of pyoverdine structures in nature, is independent of the composition of the aa side chain [119, 121, 127]. Thus, the NRPS responsible for synthesis of the side chain can be manipulated without destroying the fluorescence of the final product, providing a rapid and unambiguous means for assessment of the function of modified NRPS genes. Second, the diverse array of side chain structures in nature is evidence that the number, order and specificity of modules in the synthesising NRPS enzymes can be varied without destroying function [124]. Nature has evidently already conducted a wide range of module swapping experiments, with a great deal of success. Third, the fluorescent *pseudomonads* are genetically tractable, allowing them to be readily transformed with plasmid encoded copies of modified NRPS genes. This study uses *Pseudomonas aeruginosa* as a host for module

swapping experiments. A previously established system in our lab allows complementation of *pvdD* mutant *P. aeruginosa* strain with plasmid encoded genes, either modified or wild type [85]. This system has already been used to assess C-domain specificity, as described below.

5.1.2 – A description of previous module swapping experiments conducted using the *P. aeruginosa* pyoverdine biosynthetic system.

The experiments described in this chapter build on earlier work conducted by Dr Ackerley in which module swapping was carried out using the pyoverdine NRPS of *P. aeruginosa* in an attempt to generate modified pyoverdine molecules (Figure 5.1) [85, 143]. In this previous work, the penultimate A-domain of the pyoverdine NRPS system was substituted by foreign A-domains. This was achieved by manipulation of the first module of the *pvdD* gene, which in its native state contains two L-threonine specifying modules [37]. It was found that substitution of the first *pvdD* A-domain by an A-domain specifying an aa other than L-thr completely abolished pyoverdine biosynthesis. This result was consistent with other experimental data showing that C-domains exhibit selectivity for both the chirality and sidechain of downstream (acceptor) aa residues in a condensation reaction [48, 49].

In order to test the hypothesis that C-domain specificity for the acceptor residue during the condensation reaction was preventing incorporation of novel aas, additional module swapping experiments were conducted. In these experiments A-domains were swapped along with their cognate C-domain partners. However, this strategy of swapping CA bi-domains also failed to generate pyoverdine containing a new aa at the penultimate position. Furthermore, the foreign CA domains present in one of the non-functional recombinant genes were taken from a module in which L-thr was the native acceptor substrate. One possible explanation put forward to explain these data was that C-domains exhibit specificity not only for the aa acceptor in a condensation reaction, but also for the side chain of residues present in the peptide donor.

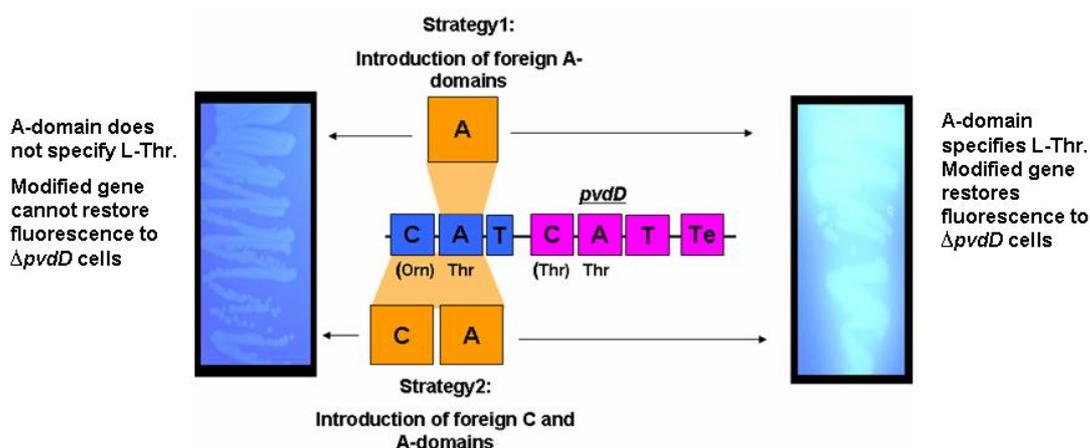


Figure 5.1 – Summary of previous domain swapping strategies using *pvdD*

In previous experiments [85] the A-domain from the penultimate module of *pvdD* was substituted by foreign domains, either alone or in combination with a cognate C-domain. Pyoverdine synthesis was only restored if the substitute A-domain specified L-thr. In the case of non-thr swaps generated using strategy 1, loss of function could be attributed to the acceptor specificity of the native C-domain (Blue). Non-thr swaps created using strategy 2 are not restricted by their acceptor specificity, but were still non functional suggesting that donor specificity of the upstream (orange) or downstream (Pink) C-domain may also have been a barrier to manipulation.

5.1.3 – Aims and overview

The aim of this work was to more rigorously assess the specificity of C-domains for the peptide donor in a condensation reaction using the pyoverdine NRPS model system. To this end a set of modified derivatives of the *P. aeruginosa pvdD* gene was created in which the CA domains of the final module were substituted for alternative CA domains from both *P. aeruginosa* and *P. syringae* pyoverdine NRPS genes. These modified genes were designed specifically to test the ability of substituted C-domains to utilise L-threonine at the C-terminus of a peptide donor chain during a condensation reaction. Based on the previous work described it was expected that only C-domains excised from a setting in which they originally utilised an L-threonine at this position would be able to fulfil this role. Assessment of one of these genes revealed that a C-domain, whose native role was to use L-N5-formyl-N5-hydroxyornithine (L-FHorn) at the C-terminal of the peptide donor, was able to efficiently use L-thr and restore pyoverdine synthesis to a *pvdD* knockout ($\Delta pvdD$) strain of *P. aeruginosa*. This provided direct evidence against the initial theory of C-domain donor specificity, suggesting this was not the cause of inactivity in previous pyoverdine module swapping experiments.

5.2 – Design and construction of recombinant *pvdD* genes

PvdD is the final NRPS protein involved in the biosynthesis of pyoverdine and contains two modules, both of which specify L-thr. The final module, which contains a thioesterase domain, was chosen for CA-domain swapping. In contrast to previous experiments, which had used the first module, the experiments described in this chapter replaced the CA-domains in the second module. Because this is the terminal pyoverdine module, there are no additional C-domains downstream that will be affected if the aa substrate for this module is changed. This new strategy therefore isolates donor site selectivity to a single module. Acceptor site specificity is also avoided because all swaps were conducted using a C-domain along with its cognate A-domain. The novel C-domain introduced would therefore always be utilising its native aa substrate as an acceptor and the ability of said C-domain to use L-thr as a donor could be unambiguously assessed via the ability or inability of a recombinant gene to restore pyoverdine synthesis in a $\Delta pvdD$ *P. aeruginosa*.

In order to facilitate expression of recombinant genes in $\Delta pvdD$ *P. aeruginosa*, the plasmid pSW196 was employed. This vector was chosen for two reasons. First, it is an integrative vector, which can be inserted into the chromosome of any bacteria containing an *attB* site in its chromosome [189]. This feature was important as it had previously been determined that complementation of the *pvdD* mutation using a high copy number plasmid resulted in a reduction in pyoverdine synthesis efficiency, most likely due to expression levels of plasmid encoded *pvdD* being too high [37, 85, 143]. Insertion into the chromosome would ensure that only a single copy of a recombinant gene was present in each cell, leading to more appropriate expression levels. Second, pSW196 contains an arabinose inducible promoter that drives expression of cloned genes. This promoter system is known to allow very fine regulation of expression levels by altering arabinose concentration in culture medium or agar plates [97]. The combination of these features means that pSW196 allows fine regulation of expression levels, allowing recombinant genes to be expressed at levels optimal for function.

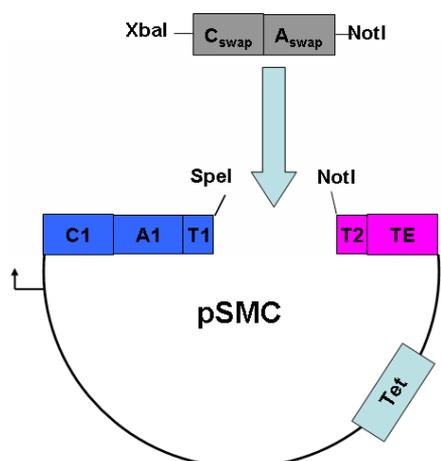


Figure 5.2 – Outline of pSMC staging vector

The Staging vector contains the *pvdD* gene, lacking the C and A domain of the second module. Restriction sites allow introduction of desired CA domains into pSMC in order to create recombinant *pvdD* genes in which the native CA-domains of module 2 have been replaced with foreign CA-domains. The recombinant genes can then be integrated into the chromosome of *P. aeruginosa* and their expression controlled by adjusting arabinose concentration. The plasmid contains a tetracycline resistance marker (Tet) allowing selection of integrants on tetracycline containing medium.

In order to allow introduction of foreign CA-domains, in place of the final CA-domain of *pvdD*, the regions up and downstream of the swap region were cloned into pSW 196 with restriction sites introduced on primers to allow introduction of foreign domains. As illustrated in Figure 5.2 the result was a plasmid with the domain order **CAT**_{-SpeI NotI}-**TTE**. SpeI and NotI are unique restriction sites that allow introduction of foreign domains with flanked by appropriate restriction sites as follows: _{XbaI}-**CA**-_{NotI}. Cutting DNA with XbaI produces sticky ends compatible to those of DNA cut with SpeI. It was necessary to use XbaI as the upstream site on CA domains, as some of the domains to be swapped contained internal SpeI sites. The identity and location of restriction sites was chosen based on restriction sites previously used for construction of recombinant *pvdD* genes, which were known to preserve function [85, 143]. The staging vector created for substitution of CA-domains into the last module of *pvdD* is henceforth referred to as pSMC. Before creation of genes harbouring non-native CA domains, a positive control was created to assess the impact of introduced restriction sites on the function of PvdD. In this control the native CA domains of *pvdD* were introduced into pSMC. The resulting plasmid was integrated into the *attB* site of *P. aeruginosa*:: Δ *pvdD* and was found to restore pyoverdine synthesis to this strain to levels comparable to those observed in wild type. This was an improvement over previous work in which the *pvdD* mutation was complemented by the wild type gene on a multi-copy plasmid, where *pvdD* overexpression

was observed to reduce overall levels of pyoverdine production by 2-3 fold relative to wild type.

Having demonstrated that the introduction of restriction sites had not destroyed function, five recombinant genes were created. Table 1 outlines the genes created, and gives the specificity of the substitute A-domain as well as the donor residue that the C-domain processes in its native setting. A graphical representation of recombinant genes is given in Figure 5.3.

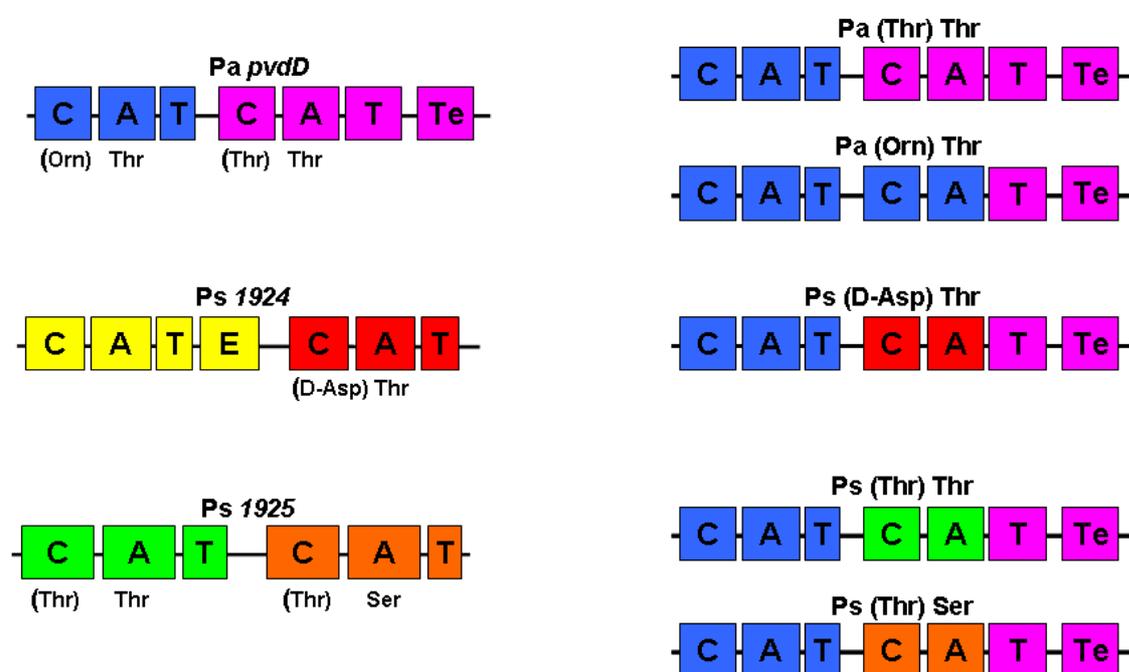


Figure 5.3 – Recombinant genes *pvdD* created in this study

The left panel shows the origin of the domains introduced into pSMC, the right panel shows the domain composition of the resulting recombinant *pvdD* gene. The donor substrate utilised by a C-domain in its native setting is shown in parentheses followed by A-domain specificity. Ps = *P. syringae*, Pa = *P. aeruginosa*. Figure produced by Mark Calcott.

5.2.1 – Nomenclature for describing modified genes and strains created in this study

Modified genes will henceforth be described as outlined in Figure 5.3. For each modified PvdD gene harbouring introduced CA-domains in the last module, a descriptor of the

organism from which these genes were derived is first given (Pa for *P. aeruginosa*, and Ps for *P. syringae*). This is followed in parentheses by the donor aa that is recognised by the introduced C-domain in its native setting. Finally, the substrate specificity of the introduced A-domain is given. For example, the modified *pvdD* gene in which the native CA-domains of the last module have been replaced with the CA-domains from the second module of the *P. syringae* gene *pspph1924* is referred to as Ps(Thr)Ser, as the introduced C-domain receives a thr residue at its donor site in its native setting, and the substrate specificity of the introduced A-domain is serine. Unless otherwise noted, all strains harbouring modified genes lack a chromosomal copy of the native *pvdD* gene and are named according to the recombinant *pvdD* gene integrated into their chromosomal *attB* site.

5.3 – Assessment of function of recombinant genes

Each of the plasmids harbouring the recombinant *pvdD* derivatives indicated in Figure 5.3 was delivered to *P. aeruginosa*:: $\Delta pvdD$ by electroporation, with plasmid integrants selected on LB agar containing tetracycline and their identity confirmed by colony PCR. These strains were then subjected to a series of phenotypic analyses to characterise the activity of the recombinant genes. To maximise pyoverdine production, a plasmid encoding the sigma factor PvdS was transformed into two of the recombinant gene expression strains, Pa(Thr)Thr and Ps(Thr)Ser, together with a wild type *P. aeruginosa* control, to generate *pvdS* over-expression derivatives of these strains. As PvdS is a sigma factor known to stimulate production of the pyoverdine biosynthetic genes in *P. aeruginosa* [121], it was expected that increased levels of this protein would promote pyoverdine production even if the final product was not able to transport ferric iron into the recombinant strains.

5.3.1 – Pyoverdine production on solid media

The first phenotypic characterisation to which these strains were subjected was assessment of pyoverdine production on Kings B (KB) agar (2.8.6). KB is an iron limited medium known to stimulate production of pyoverdine in fluorescent *Pseudomonas* species. Duplicate plates were streaked with strains harbouring each of the recombinant genes, as well as the WT and $\Delta pvdD$ controls and the *pvdS* overexpression strains. Arabinose was present at a final concentration of

0.4% in one of the plates and absent in the other. After incubation at 37 °C for 16 h the strains were assessed for pyoverdine production by visualisation of green pigmentation and UV-fluorescence. As shown in Figure 5.4, there was no evidence of pyoverdine production for $\Delta pvdD$, Ps(Asp)Thr, Ps(Thr)Thr or Ps (Thr)Ser in either the presence or absence of arabinose. In contrast, strong colouration and fluorescence were evident for WT regardless of arabinose inclusion. Pa(Thr)Thr and Pa(Orn)Thr were also strongly coloured and fluorescent, but only when arabinose was included in the medium. By eye, inclusion of the *pvdS* over-expression plasmid made no apparent difference to pyoverdine production in any of the strains tested.

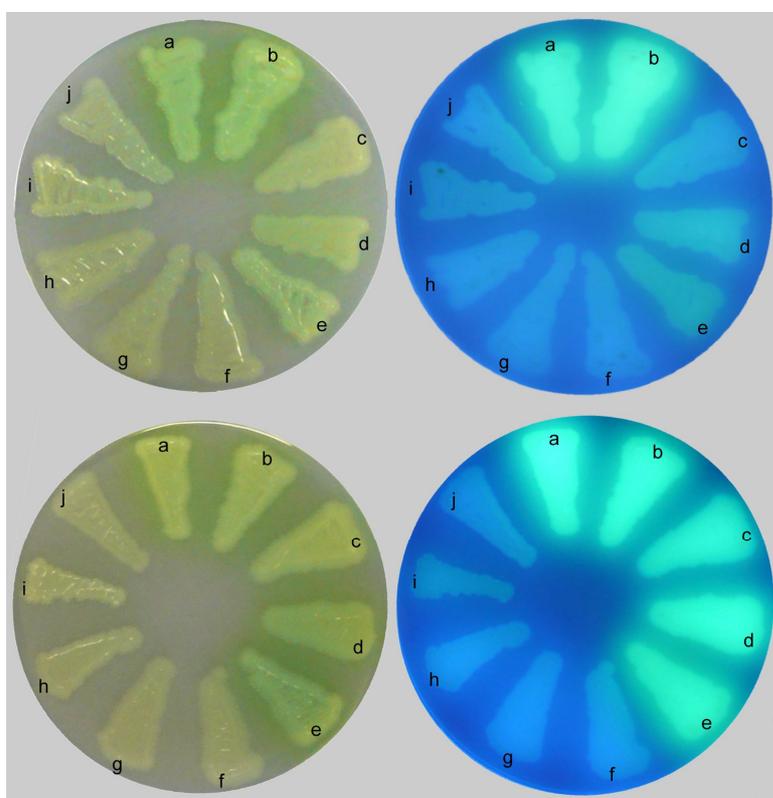


Figure 5.4 – Assessment of pyoverdine production strains on KB agar

Two separate plates are shown, the plate in the top panel is unamended KB agar, the plate in the bottom panel contains 0.4 % arabinose to induce expression of recombinant genes under the control of the arabinose regulated pSMC promoter. Plates were photographed under natural lighting (left) and UV light (right) after 24 h growth at 37 °C. Strain identity is as follows: a) WT, b) WT *pvdS* overexpression, c) PvdD Pa(fOHOrn)Thr, d) PvdD Pa(Thr)Thr, e) PvdD Pa (Thr)Thr *pvdS* overexpression, f) PvdD Ps(Asp)Thr, g) PvdD Ps(Thr)Thr, h) PvdD Ps(Thr)Ser, i) PvdD Ps(Thr)Ser *pvdS* overexpression, and j) *pvdD* deletion mutant. Strains c-i contain modified *pvdD* genes integrated into the chromosomal *attB* site and lack the native *pvdD* gene. Figure produced by Mark Calcott. *PvdS* overexpression refers to strains harbouring a plasmid borne copy of the *pvdS* gene of *P. aeruginosa* (Plasmid pucP22::*pvdS*, table 2.4)

5.3.2 – Pyoverdine production in liquid media

In order to more quantitatively assess the level of pyoverdine production in strains bearing recombinant *pvdD* genes, growth in liquid media, followed by measurement of pyoverdine levels in culture supernatant was carried out as described in Section 2.8.6. The chromophore pyoverdine absorbs maximally at a wavelength of 400 nm, so absorbance at this wavelength is proportional to the concentration of pyoverdine in a particular solution [190, 191]. Pyoverdine levels in culture supernatant were measured following 24 h growth in KB broth in a 96 well plate. Three cultures were assessed for each strain in both the presence and absence of arabinose. As shown in Figure 5.5, pyoverdine was not detected in the culture supernatant of Ps(Thr)Thr or Ps(Thr)Asp, consistent with the results of assessment on KB agar. Of particular interest, pyoverdine was detectable in the culture supernatant of Ps(Thr)Ser, suggesting that this strain was successfully producing a modified pyoverdine. Furthermore, Pa(Orn)Thr was functional, but Ps(Thr)Thr was not, challenging the donor site hypothesis. The highest levels of pyoverdine were observed for WT followed by Pa(Thr)Thr and then Pa(Orn)Thr. For WT, Pa(Thr)Thr, Pa(Orn)Thr and Ps(Thr)Ser, *pvdS* over-expression strains were also assessed and in all cases found to produce higher levels of pyoverdine than the corresponding strain not harbouring a *pvdS* over-expression plasmid.

An additional experiment was run in which $\Delta pvdD$, Pa(Thr)Thr and Ps(Thr)Ser were assessed for pyoverdine production following growth at various concentrations of arabinose. In this experiment, pyoverdine levels were assessed by measuring emission at 450 nm following excitation at 403 nm – the characteristic excitation and emission wavelengths of the chromophore of pyoverdine [190, 191]. The results of this experiment, represented graphically in Figure 5.6, showed that up until a final concentration of 0.4 % w/v, higher arabinose concentration resulted in more pyoverdine being produced for both Ps(Thr)Ser and Pa(Thr)Thr; with concentrations above 0.4 % reducing pyoverdine levels.

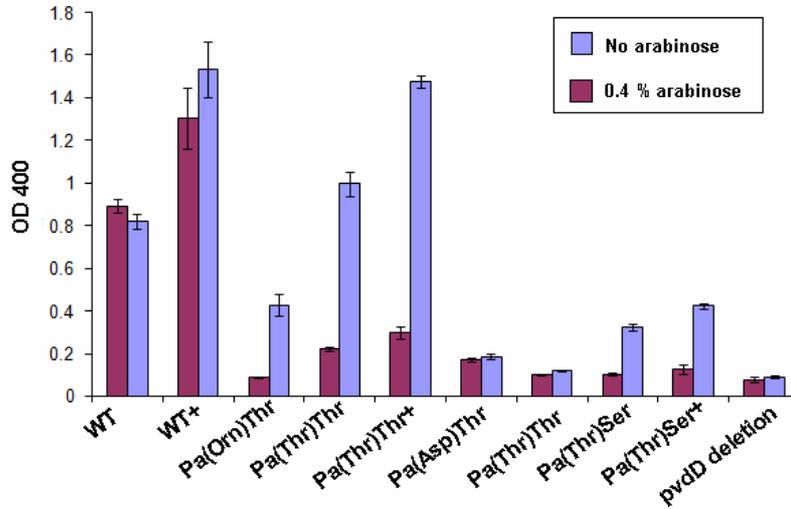


Figure 5.5 – Quantitative assessment of pyoverdine production in KB broth

The indicated strains were grown in KB for 24 hr with or without 0.4% arabinose. Cultures were set up in triplicate for each condition/strain in a 96 well plate. Following growth, cells were pelleted and pyoverdine levels in culture supernatant determined by measuring absorbance at 400 nm using a microplate reader. The suffix + indicates a strain harbouring a *pvdS* overexpression plasmid. Error bars are presented as +/- one standard deviation. Figure produced by Mark Calcott.

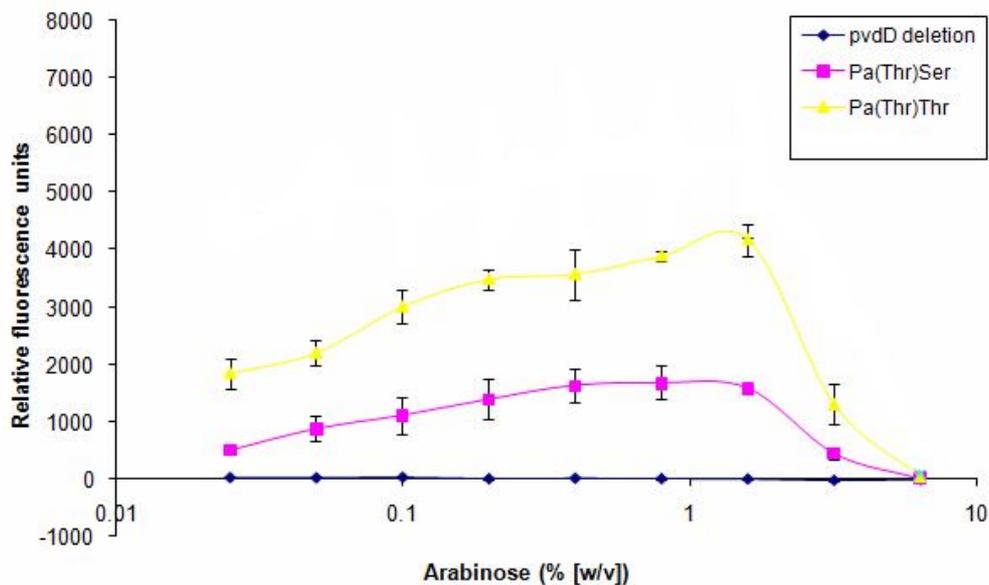


Figure 5.6 – Effect of arabinose concentration on pyoverdine production

Triplicate cultures for the indicated strains were inoculated for each arabinose concentration in a 96 well plate. The base medium used was KB broth. Following 24 hr growth at 37 °C, pyoverdine levels in culture supernatant were determined by measuring emission at 450 nm following excitation at 403 nm. Error bars are presented as +/- one standard deviation. Figure produced by Mark Calcott.

5.3.3 – Assessment of pyoverdine function for strain harbouring modified *pvdD* genes

Having established that a subset of strains harbouring modified *pvdD* genes were capable of producing pyoverdine, the ability of these strains to utilise the pyoverdine they produced for iron uptake was assessed. This was achieved by streaking on to KB agar containing 200 µg/mL EDDHA. EDDHA is an iron chelator that prevents passive uptake of iron, and it has previously been demonstrated that *P. aeruginosa* cannot grow in the presence of 200 µg/mL EDDHA unless functional pyoverdine is being produced [37, 85, 143]. As illustrated in Figure 5.7, it was found that only WT, Pa(Thr)Thr and Pa(Orn)Thr were producing a pyoverdine species that enabled them to grow in the presence of 200 µg/mL EDDHA. The fact that the pyoverdine species produced by Ps(Thr)Ser was not functional was consistent with this molecule having an alternative aa installed at the final position, resulting in a loss of function.

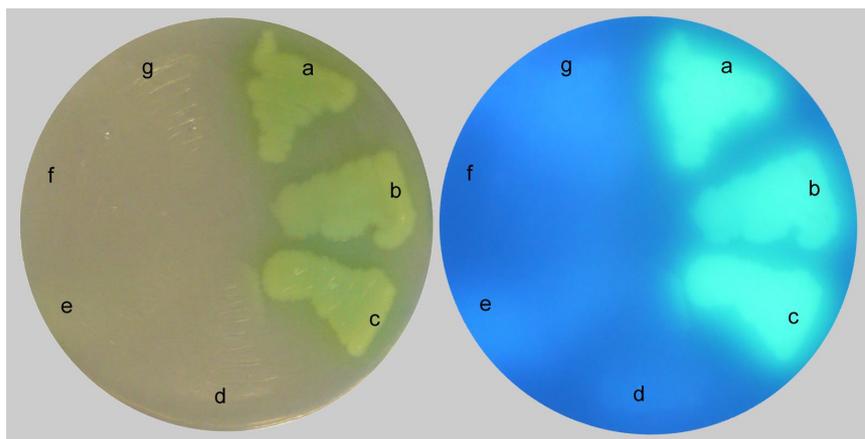


Figure 5.7 Growth of mutant strains under iron starvation.

Strains were generously streaked onto KB agar containing 200 µg/mL EDDHA. Under these conditions passive uptake of iron is completely restricted. After 18 h incubation at 37 °C the plate was photographed under natural (left) and UV (right) light. Strain identity is as follows: **a)** WT, **b)** Pa(fOHOrn)Thr, **c)** Pa(Thr)Thr, **d)** Ps(Asp)Thr, **e)** Ps(Thr)Thr, **f)** Ps(Thr)Ser, and **g)** *pvdD* deletion mutant. Figure produced by Mark Calcott.

5.3.4 – Analysis of intracellular pyoverdine levels in modified strains using flow cytometry

The results outlined thus far suggested that the Ps(Thr)Ser was producing a modified pyoverdine derivative harbouring ser instead of thr as its terminal aa. This modified species was produced at levels much below those of WT and was not able to effectively function as a

siderophore. These results presented the possibility that the modified pyoverdine produced by Ps(Thr)Ser was not being effectively exported from the cell. In order to assess the levels of intracellular pyoverdine present in Ps(Thr)Ser, WT, $\Delta pvdD$ and Pa(Thr)Thr strains, washed cells from iron starved cultures were subjected to flow cytometric analysis using the Pacific Blue™ channel (ex 410/em455). The results of this analysis, in which 5,000 events were collected for each strain, are shown in Figure 5.8. As expected, $\Delta pvdD$ had lower mean fluorescence levels than WT. Unexpectedly, both Ps(Thr)Ser and Pa(Thr)Thr had higher mean fluorescence than wild type. This result suggests that these strains were accumulating pyoverdine in their periplasm. In the case of Ps(Thr)Ser, this is supporting evidence for the proposition that a modified pyoverdine was being produced, but was unable to be effectively exported from the cell. The increased intracellular fluorescence of Pa(Thr)Thr may be a symptom of a disruption of the correct molar ratio of NRPS modules involved in pyoverdine biosynthesis. The reduced export levels seen in Ps(Thr)Ser prevented purification of the putative modified pyoverdine for MS analysis, however this goal is being pursued by Mark Calcott as part of his doctoral research.

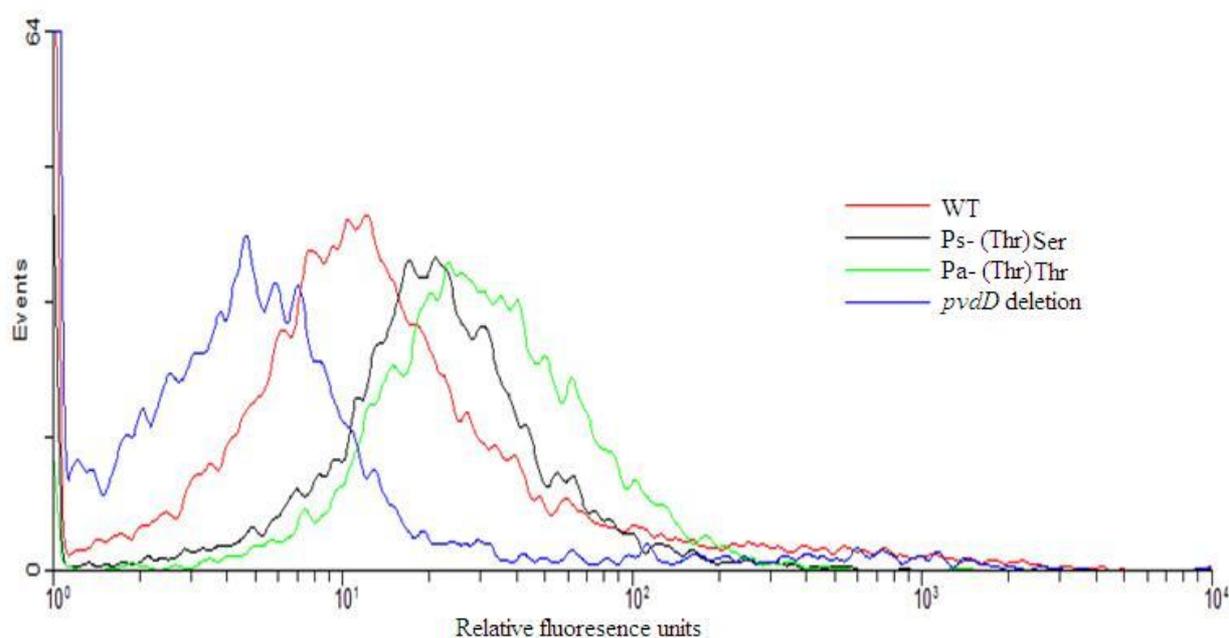


Figure 5.8 – Determination of intracellular pyoverdine levels by flow cytometry

The fluorescence of individual cells was assessed by measuring emission at 455 nm following excitation at 410 nm. 5000 events were recorded for each strain; the frequency of events giving a particular fluorescence level is presented as a histogram. Figure produced by Mark Calcott.

5.4 – Discussion

The aim of the experiments described in this chapter was to assess C-domain specificity for the upstream (donor) aa during a condensation reaction. Based on previous experiments [85, 143], it was hypothesised that donor site specificity would prevent modified *pvdD* genes restoring fluorescence to *P. aeruginosa::ΔpvdD* cells unless the terminal C-domain in such genes was able to recognise an L-thr donor. However, the results of the experiments described are not consistent with this hypothesis.

Of the five modified *pvdD* genes created and tested, two contained terminal C-domains that do not receive L-thr as a donor substrate in their native setting. These were Ps(Asp)Thr and Pa(Orn)Thr, both of which contained foreign A-domains specifying thr. Whereas Ps(Asp)Thr was not able to restore fluorescence to *P. aeruginosa::ΔpvdD* under any of the conditions tested, Pa(Orn)Thr was not only functional, but actually restored fluorescence to levels comparable to those seen for WT *P. aeruginosa*. The pyoverdine produced by Pa(Orn)Thr cells was able to scavenge iron from the environment under conditions of severe iron limitation, indicating that it was identical to the native pyoverdine. The substituted C-domain in Pa(Orn)Thr must therefore be able to receive both L-fhorn and L-thr residues at the donor site. As shown in Figure 5.9, the side chain of these two substrates is vastly different in terms of both size and chemical composition, providing evidence that the terminal C-domain in Pa(Orn)Thr does not discriminate between donor substrates based on the sidechain of the C-terminal residue. Investigation of the generality of this result is one of the aims of the doctoral research of Mark Calcott and will require synthesis of additional recombinant genes, coupled with purification and MS analysis of products.

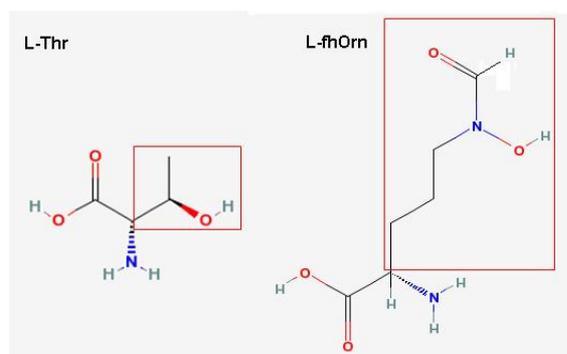


Figure 5.9 – Comparison of L-thr and L-fhorn structure

It is more difficult to draw definitive conclusions from negative results, as there may have been multiple reasons for the failure of recombinant *pvdD* constructs to yield a pyoverdine product. However, that the Ps(Asp)Thr construct did not yield a pyoverdine product is consistent with the C-domain not being able to receive L-thr as a donor substrate from module 1 of *pvdD*. In its native setting, this C-domain utilises D-asp as a donor substrate, and previous *in vitro* characterisation experiments have indicated that C-domains do possess specificity for the chirality of donor substrates in a condensation reaction [48, 49].

More surprisingly, the gene Ps(Thr)Thr was not able to function in place of *pvdD*. This result cannot be explained by C-domain specificity constraints since the substitute C-domain in Ps(Thr)Thr uses the same donor and acceptor substrates in its native setting as those required in its new setting. This demonstrates that there can be additional barriers, independent of C-domain specificity, which prevent optimal function of recombinant NRPS proteins. One possible explanation for the lack of function of Ps(Thr)Thr is that the introduced C and A-domains are unable to correctly interact with the native T-domains of *pvdD*. As discussed at length in Chapter 6 of this thesis, there is good evidence that T-domains enter into specific interactions with partner domains in NRPS enzymes and that disruption of optimal T-domain interactions can contribute to loss of function in modified NRPS enzymes [15, 53, 92-94, 192, 193]. There are three possible interactions that could break down as a result of substituting foreign CA-domains into the last module of *pvdD*. As illustrated in Figure 5.10, the substituted C-domain is required to interact with both up and downstream T-domains during the condensation whereas the substituted A-domain is required to interact only with its downstream T-domain during substrate transfer.

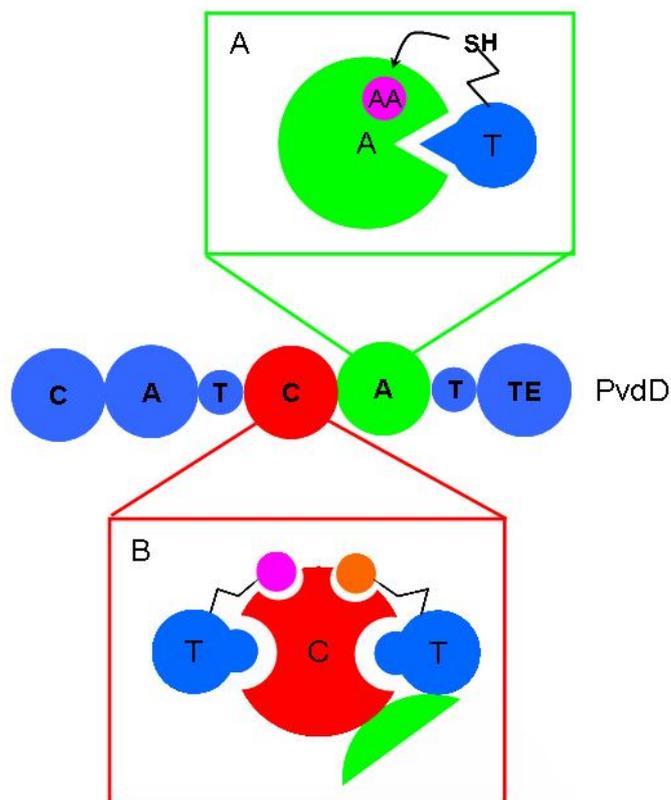


Figure 5.10 – Interaction of substituted CA domains with T-domains in PvdD

In order to function correctly in PvdD, substituted A-domains (Green) and C-domains (Red) must interact correctly with the native T-domains of PvdD. In the case of the substituted A-domain, this interaction occurs exclusively with the downstream T-domain during substrate transfer (box A). In the case of the substituted C-domain, interaction with both the upstream and downstream T-domains is required during peptide bond formation (box B).

Evidence for the lack of function of Ps(Thr)Thr being a consequence of breakdown in T-domain interactions was found by *in silico* analysis. The bioinformatic and directed evolution studies described in Chapter 6, as well as existing structural [53, 119, 192] and biochemical data [92-94], identify specific sequence elements in T-domains that may be important for correct interaction with downstream domains. One such residue is found at position +4 relative to the conserved ser 4'-PP attachment site (S+4) (Figure 5.11). In the bioinformatic analysis described in Appendix 1 it was found that, in T-domains that are immediately adjacent to a downstream C-domain, this residue can be either thr, ser, val, ala or met; whereas in T-domains that are adjacent to a downstream E or TE-domain this position is always occupied by a hydrophobic residue. The first T-domain of PvdD has a thr at position S+4; and for each

of the functional recombinant *pvdD* constructs the native upstream T-domain partner of the introduced C-domain also has a thr in this position. Conversely, neither of the inactive recombinant C-domains is partnered with a T-domain that has a thr at position S+4 in its native setting (Figure 5.11). However, the hypothesis that the position S+4 threonine in the T-domain of the first *pvdD* module may have been a primary cause of inactivity for the Ps(Thr)Thr and Ps(Asp)Thr constructs is yet to be tested experimentally.

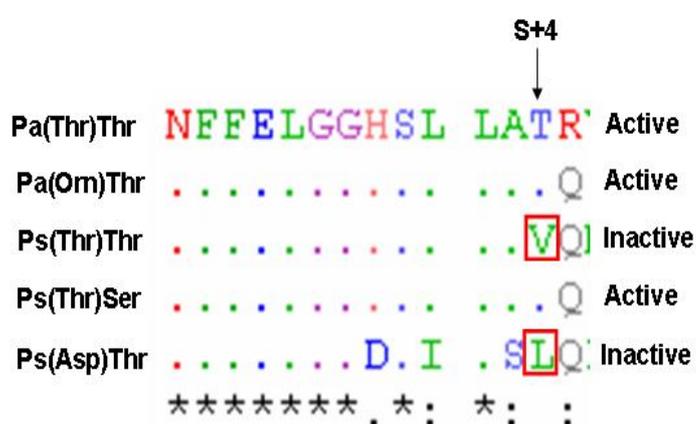


Figure 5.11 – Alignment of T-domains found upstream of swapped C-domains in nature

The top sequence is the first T-domain of PvdD, which is the upstream interaction partner for all C-domains in the recombinant *pvdD* constructs. Below this are the sequences of the T-domains that occur immediately upstream of the introduced C-domains in their native setting. The activity of each recombinant construct is summarised on the right. For clarity only part of the sequence alignment is shown here, the full alignment can be found in Appendix 1.

One possible avenue for such investigation is to perform site-directed mutagenesis and/or directed evolution of the T-domain immediately upstream of the substituted C-domain in Ps(Thr)Thr. Directed evolution could be achieved using a system similar to that described in Chapter 6, in which T-domains are randomly mutated by error prone PCR and variations resulting in improved interaction with partner domains are selected based on improvements in product synthesis. In the case of Ps(Thr)Thr, the natural function of pyoverdine provides the basis for a screen even more powerful than that described in Chapter 6. Functional pyoverdine is essential for survival on plates containing 200 µg/mL EDDHA, and consequently T-domain mutations resulting in functional Ps(Thr)Thr could be categorically selected in a single step by plating error prone libraries on this medium. For a more quantitative assessment of function *in vivo*, surviving clones could then be assessed for level of fluorescence. A system similar to this has been successfully implemented for the directed evolution of the aryl carrier protein of

the enzyme EntB, involved in production of the siderophore enterobactin in *E. coli* [92]. However, the selection step in this system relied on assessment of relative growth rate rather than absolute selection of viable clones, and subsequent quantitative analysis was arduous as enterobactin does not possess any inherently quantifiable properties akin to the fluorescence of pyoverdine.

The modified *pvdD* derivative Ps(Thr)Ser created in this study was able to restore fluorescence to cells lacking the native *pvdD* gene. This result is particularly exciting as based on the specificity of the new A-domain in Ps(Thr)Ser, the pyoverdine species produced is likely to possess ser instead of thr at the last position. The fact that the pyoverdine produced by Ps(Thr)Ser is not functional is strong evidence that it differs in structure from the native pyoverdine species produced by *P. aeruginosa*. Additional evidence that the pyoverdine produced by Ps(Thr)Ser contains an alternative aa is found in the results of the flow cytometric analysis described in Section 5.3.4. In this analysis it was found that Ps(Thr)Ser cells accumulate higher levels of pyoverdine intracellularly than WT *P. aeruginosa*. This result is consistent with production of a modified pyoverdine that is less efficiently recognised and exported from the cell. Confirmation of the identity of the pyoverdine species produced by Ps(Thr)Ser cells could be confirmed by purification followed with MS analysis as described in Section 3.7. Or if purification is not possible owing to the impaired export of this compound, whole cell mass spectrometry (WSMS) could be used. There are a number of WSMS techniques, including whole cell MALDI TOF that allow identification of the products [194, 195] and intermediates [196, 197] of NRPS pathways in bacteria. These techniques have proven effective for detecting molecules of similar size to pyoverdine with good resolution, and could potentially be used to accurately determine the mass difference between the pyoverdine species produced by Ps(Thr)Ser and wild type PAO1, without the need for prior purification steps.

Chapter 6 – Probing T-domain interactions using BpsA as a reporter system

6.1 – Introduction

This chapter describes development and implementation of a directed evolution platform, the purpose of which was to identify key residues involved in the interaction between T-domains and other domains in an NRPS enzyme. The system described was based on the recently characterised Blue Pigment Synthetase A (BpsA), from *Streptomyces lavendulae*. The native T-domain of BpsA was substituted by foreign T-domains using three different schemes and the resulting hybrid proteins assessed for activity *in vivo*. In all cases, substitution of the native T-domain caused severe reduction or complete ablation of pigment synthesis, presumably due to inability of the substituted T-domain to correctly interact with other domains in BpsA. A high throughput screening system was developed which allowed hybrid synthetases with reduced function to be improved by random mutagenesis of the substituted T-domain, followed by selection of improved variants using a multi-tiered screening process. By sequencing improved variants and determining changes that consistently led to improved function it was possible to identify key residues likely to be involved in interaction between the T-domain of BpsA and other domains in the NRPS. Based on these results, a model for T-domain interaction with the thioesterase domain of BpsA is presented.

Blue Pigment Synthetase A (*bpsA*) is a single module NRPS gene from *Streptomyces lavendulae* that was recently genetically and functionally characterised [80]. BpsA catalyses the formation of the blue pigment indigoidine from two molecules of L-gln. A proposed mechanism for this reaction [80, 198] is outlined in Figure 6.1B; it should be noted that this mechanism has not been experimentally verified and there is some uncertainty over the order of the condensation and oxidation steps. Indigoidine is synthesised by an array of phylogenetically diverse bacteria [80], is believed to play a role in quorum sensing, and is necessary for virulence in *Erwinia chrysanthemi* [198]. BpsA has unusual domain organisation, containing an oxidation domain integrated into its single A-domain between conserved motifs 8 and 9. In addition to its single Ox and A-domain, the protein contains a T-

domain and a TE-domain. The modular architecture of BpsA is outlined in Figure 6.1A. BpsA is the only known example of a single NRPS module that is capable of synthesising an easily detectable, coloured product, without the requirement of additional enzymatic steps [80]. This factor, coupled with its ability to function in *E. coli*, make BpsA a particularly attractive model for conducting domain swapping experiment to examine domain interactions in NRPS.

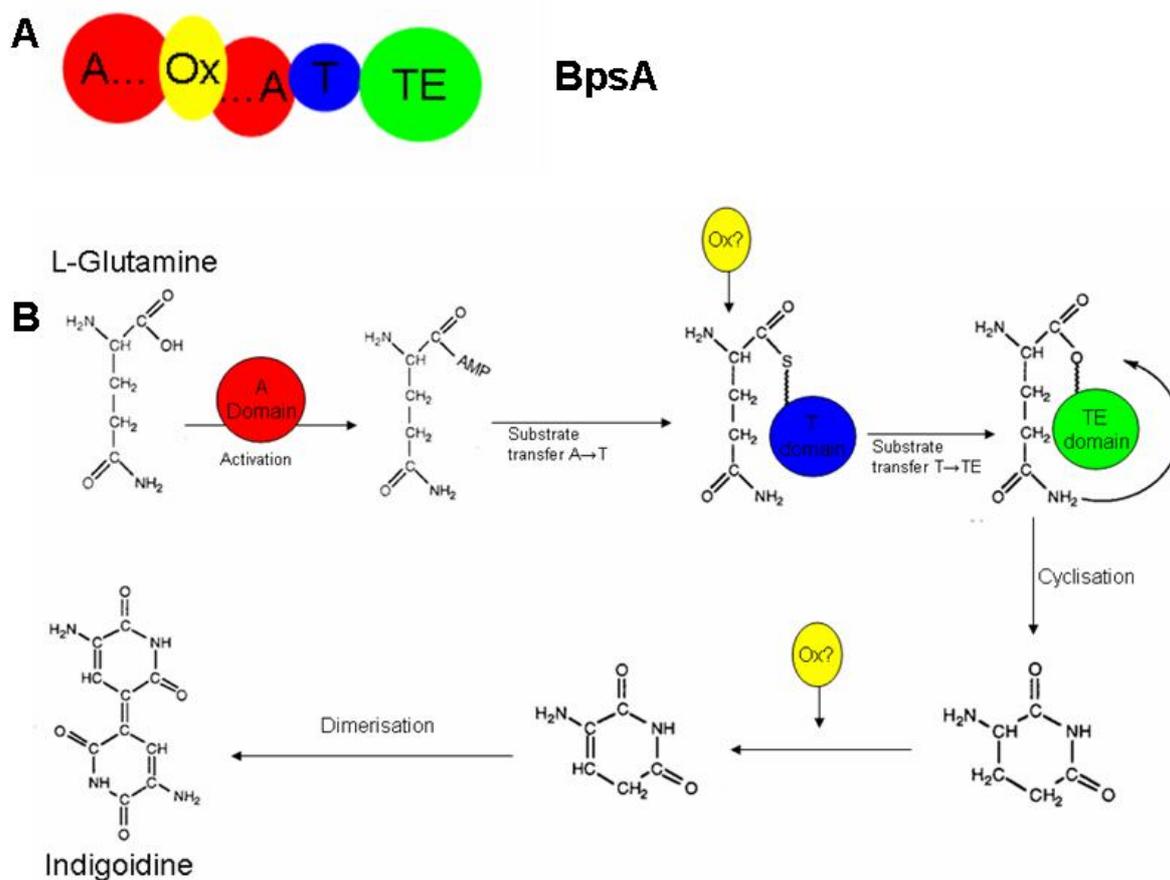


Figure 6.1 – BpsA

A – BpsA is an NRPS enzyme with unusual domain organisation, it contains a single A, T and TE-domain with an Ox-domain integrated into the A-domain.

B – BpsA synthesises the blue pigment indigoidine from two molecules of L-gln. The function of each of the domains is shown. It is unknown if the oxidation reaction occurs before or after cyclisation. The mechanism of dimerisation is also unknown. The proposed synthesis mechanism is as described in [198], structures for substrate intermediates and product were reproduced from this reference.

Recent biochemical and structural data indicate that NRPS T-domains play an essential role in the coordination of enzymatic steps during product synthesis by NRPS via specific temporal and spatial interactions with other synthetase domains [15, 44, 47, 53, 93, 94, 192, 193]. It is not presently clear exactly how T-domains interact with other domains, however what is clear is that the T-domain can no longer be thought of simply as a passive tether, to which substrates and intermediates are attached. As outlined in Section 1.4, T-domains are known to adopt at least three distinct conformations in solution, one of which is exclusive to apo-domains, another exclusive to holo-domains and the third an intermediate between the two [44]. There is evidence that switching between these conformations alters the affinity of the T-domain for other proteins and domains, suggesting that conformational switching modulates T-domain interactions [3, 44]. One possible interpretation of the available data is a cycle of T-domain conformational switches. In this model conformational changes modulate interaction with the appropriate partner domain, as illustrated in Figure 6.2. Elucidation of the crystal [47, 53] and solution [192] structures of multi-domain NRPS subunits in which a T-domain is present, has supported the hypothesis that T-domains interact specifically with partner domains, and that this interaction is involved in correct positioning of substrates and intermediates of the biosynthetic process. However, one limitation of structural data in the context of proteins such as NRPS that exhibit considerable conformational plasticity, is that a particular structure may represent only one of many conformational states. Interestingly, in the case of multi-domain crystal structures, generation of diffraction quality crystals has required that the T-domain be exclusively in its inactive apo-state (without attachment of 4'-PP), a situation achieved either by mutation of the active site serine [53] or repression of PPTase activity in culture medium [47]. Having the T-domain present in its apo-state presumably allows crystal formation by locking the protein in one conformation.

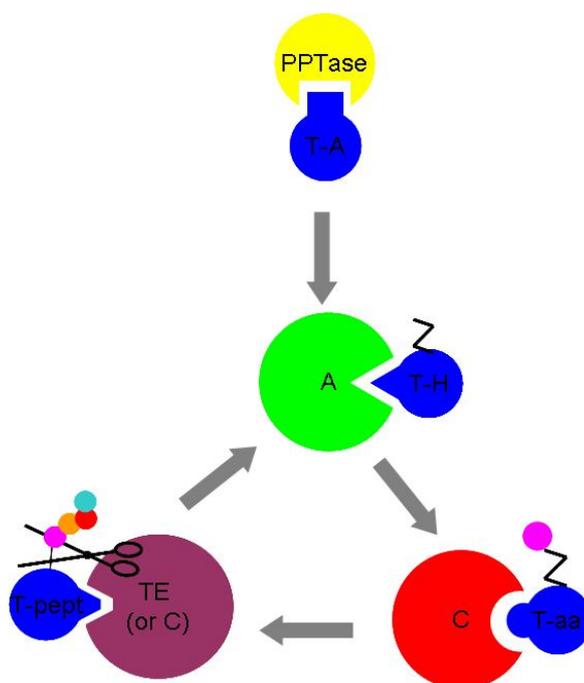


Figure 6.2 – A model for T-domain conformational switching and domain interactions. Statements in bold have been experimentally verified; others are hypothetical

The Apo (T-A) state is exclusive to T-domains prior to 4'-PP attachment and might promote interaction with a PPTase. **4'-PP attachment allows access to the Holo (T-H) state. Switching to the H-state alters affinity for other domains** and might promote interaction with an A-domain leading to attachment of an amino acid substrate. Once amino acylated (T-aa), the T-domain may again undergo a conformational switch, allowing it to interact with the upstream C-domain. Following transfer of the growing peptide chain to the T-domain (T-pept) additional conformational switching might facilitate interaction with a downstream C or TE-domain. Interaction with this downstream domain would result in peptide removal, resetting the T-domain to its H state. The T-A and T-H states have been experimentally detected along with an intermediate T-A/H state. T-aa and T-pept structures have not been investigated, so the existence of distinct conformations is hypothetical. An alternative possibility is that there is not a distinct conformation in these states and that interaction with the correct partner domain occurs out of necessity, since no other domain is capable of correctly processing the T-linked intermediate.

6.1.1 – T-domain structure and function

All NRPS T-domains perform essentially the same function, carrying aa substrates and peptide intermediates during synthesis. Aside from the difference in substrates and intermediates, this function can be extended to other carrier proteins found in PKS and fatty acid (FAS) biosynthetic pathways [1, 3, 11]. This common function is evident in structural data which has revealed that different NRPS T-domains as well as the carrier proteins of FAS and PKS have

the same basic 3-4 helical (depending on conformation) structure, in spite of having low sequence identity [43, 53, 192]. Secondly, it should be noted that there is a high level of diversity in the aa sequence of T-domains, even within a single species. If this sequence diversity reflects purpose, e.g. specialisation to interact with a particular domain or set of domains in a synthetase enzyme, T-domains from NRPS enzymes may be viewed not only as carriers of aas and peptides, but also of information [92-94, 193, 199-201].

6.1.2 – Aims

The aim of the work described in this chapter was to identify key residues involved in interaction between the T-domain and other NRPS domains in the protein BpsA by using directed evolution to improve the function of BpsA enzymes harbouring foreign T-domains.

6.2 – Examination of T-domain interactions *in silico*

As a guide for the design and interpretation of the experiments described in this chapter two *in silico* analyses of T-domains were carried out. The first of these was an alignment based analysis, the second an examination of existing T-domain structural data.

6.2.1 – Alignment based analysis of T-domain interactions

The vast majority of NRPS T-domains are required to interact with both an A-domain and a C-domain located upstream in a synthetase complex. The interaction with downstream partners is more varied, with three distinct possibilities existing for the downstream interaction partners of a T-domain. These possibilities are a C-domain only, an E-domain and a C-domain or a TE-domain only. It was theorised that if T-domains are specialised for interaction with particular partner domains, this would be reflected by the presence of specific sequence elements common to T-domains interacting with a particular downstream domain. In order to test this theory a simple bioinformatic analysis was carried out. This analysis is described in detail in Appendix 1 of this thesis and the key findings are summarised here:

1) T-domains exhibit very high sequence diversity compared to carrier proteins that interact with a limited set of partner proteins such as the carrier protein of type II fatty acid synthases (refer to Figure A1.1 in appendix one).

2) The analysis of multiple T-domain sequences interacting with a C, E or TE-domain downstream revealed that specific sequence elements are common to T-domains that have to interact with the same downstream domain (refer to Figure A1.2 and table A1.1 in appendix one).

3) Analysis of sequences arising from a whole-module (CAT) duplication upstream of a TE-domain showed that the sequence of duplicated T-domain diverges much more quickly than that of the A or C-domain. This accelerated divergence presumably reflects adaptation of the duplicated T-domain for optimal interaction with its new downstream partner.

4) T-domains that have TE-domains as downstream interaction partners have the highest sequence diversity; this suggests that there are multiple ways that a T-domain can interact with a downstream TE-domain. By contrast, T-domains that have an E-domain as a downstream interaction partner have the lowest sequence diversity suggesting a less variable mode of interaction.

The overall conclusion from the bioinformatic analysis was that T-domains have specific genetic features that provide a structural basis for their interactions with specific partner domains. Alignment of multiple T-domain sequences revealed key residues likely to play key roles in interaction with downstream domains.

6.2.2 – Structural insights into T-TE interaction

Existing biochemical data, as well as the alignment analysis described, suggests that the native interaction between a T-domain and a downstream domain is specialised and that interaction with a TE-domain is particularly sensitive to disruption in manipulation experiments [15, 92, 93, 192]. To provide a structural foundation for assessing which mutations may have improved function by restoring interaction between a substituted T-domain and the TE-domain of BpsA, the solution structure solved for the EntF T-TE bi-domain [192] was examined in detail. The PDB file for this structure, 2roqA, was downloaded and visualised using Deep View, this PDB file contains 20 conformers, the first of which is thought to be the best representation of the average structure. In this structure the T-domain is in close association with the TE-domain in a conformation that is believed to represent productive interaction between the two domains [192].

The structure of the T-domain in 2roqA, henceforth referred to as the TE-state, differs considerably from previously solved structures for T-domains in isolation. It is most similar to the H-state observed for the TycC3 T-domain however differs considerably even from this structure. The relative size and orientation of helices differs between the H and TE-states, and there is an additional single turn helix ($\alpha T1'$) present in the TE-state [192]. $\alpha T1'$ occurs at the site of two consecutive phe (F41 and F42) residues, which are absolutely conserved in all T-domains. The presence of $\alpha T1'$ results in a drastic alteration in the orientation of these phe residues, as shown in Figure 6.3. In the H-state the side chains of both phe residues are pointed toward the hydrophobic core of the T-domain, in the TE-state however, the F41 is embedded in the core of the T-domain while F42 is found in the core of the adjacent TE-domain. F42 lies at the heart of a predominantly hydrophobic interface between the T and TE-domains during their interaction in EntF which is now examined.

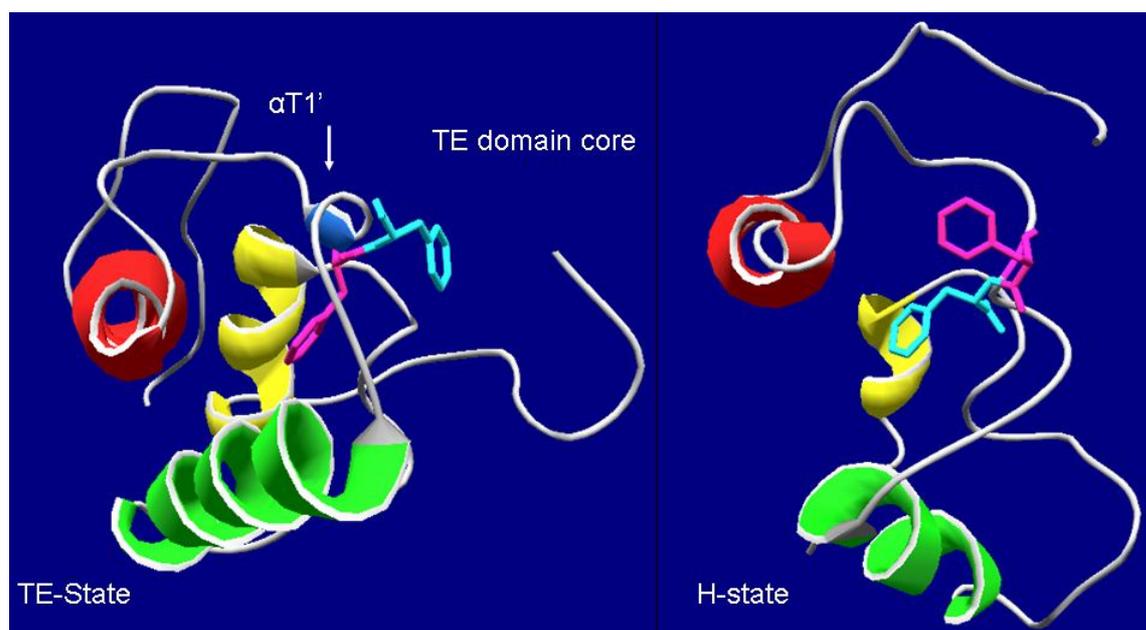


Figure 6.3 – Comparison of T-domain H and TE-states

The relative length and orientation of the helices in the two structures differs and there is an additional single turn helix ($\alpha T1'$) present in the TE state. Two consecutive F residues which are absolutely conserved in all T-domains are shown. In the H-state, the side chain of both of these residues points toward the hydrophobic core of the T-domain. In the TE-state, the side chain of the second F residue (blue) extends into the hydrophobic core of the adjacent TE-domain (not shown).

Visualisation of all residues within 5Å of the side chain of F42 (Figure 6.4) reveals a cluster of hydrophobic residues from both the T and TE-domains of EntF that form the interface between the two domains. Analysis of TycC3 in the H state shows the equivalent residues are all buried in the core of the T-domain. Transition from the H-state to productive interaction with a TE-domain therefore requires that the sidechains of residues previously buried in the core of the T-domain now extend into the hydrophobic core of the TE-domain.

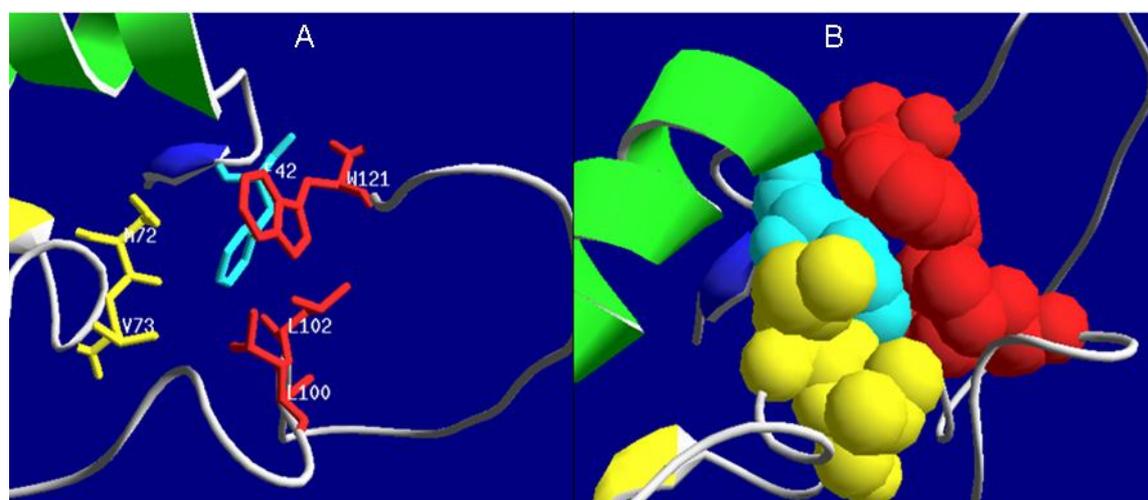


Figure 6.4 – Hydrophobic interface of the T and TE-domains of EntF

A – Detail of the interaction region of the T and TE-domains of EntF showing hydrophobic residues which are found at the interface of the two domains. Residues from the T domains are coloured yellow, residues from the TE-domain are red. F42 (Blue) from the T domain is found at the heart of this cluster.

B – Addition of van der Waals radii to the side chains forming the hydrophobic interface emphasises the formation of a tightly packed hydrophobic core.

The analysis of the structure 2roq suggests that the interface between the T and TE-domain of EntF is stabilised by hydrophobic interactions. In order to assess whether a similar situation is likely for the interaction of the T and TE-domains of BpsA, a model for the structure of these domains was generated using the Swissprot server. The T and TE-domains of BpsA were aligned against those of EntF and a set of preliminary structural models generated using Deep View. These models were then submitted to the online server for refinement and the quality of refined models assessed as described in 2.16.2. The best quality model was then visualised in Deep View and the interface between the T and TE-domains in this model was examined. This

structural model, which is presented and discussed in Sections 6.11.3-6.11.5, predicted a similar hydrophobic interface between the T and TE-domains of BpsA.

Two additional models for T-TE-domain interfaces were generated using 2roqA as a template. The T-TE bidomains for which models were generated were taken from the terminal modules of the *P. syringae* NRPS Pspph1926 and an uncharacterised NRPS from *Vibro angularium* which shared 29 and 46 % sequence identity with EntF respectively. For both of these structural models, a hydrophobic interface between the T and TE-domains was again predicted.

6.3 – Cloning of BpsA and determination of compatibility with Pseudomonas PPTases

BpsA is non-functional when expressed in otherwise unmodified *E. coli*, due to the inability of native PPTase enzymes to recognise and activate its T-domain [80]. Before moving on to manipulation experiments using BpsA, it was therefore necessary to develop protocols for co-expression of this protein with a cognate PPTase in *E. coli*. Since the sequence for the native PPTase of *S. lavendulae* is not available, it was decided to test the competency of PPTase enzymes from other bacterial strains in our laboratory for activation of BpsA *in vivo*. In their characterisation of BpsA, Takashi *et al.* used the PPTase of the genome sequenced organism *S. lividans* to activate BpsA; however gDNA from this organism was not available to us. Prompted by published reports describing the broad substrate specificity of various *Pseudomonas* PPTases [69, 202, 203], it was decided to express BpsA in *P. putida* KT2440, *P. aeruginosa* PAO1 and *P. syringae* pv *phaseolicola* 1448a, in order to ascertain whether the endogenous PPTases of these organisms were able to recognise the T-domain of BpsA.

6.3.1 – Construction of plasmids for expression of BpsA in E. coli and Pseudomonas species

For initial expression trials of BpsA in *E. coli* and *Pseudomonas*, it was decided to use the broad host range expression vector pSX (Section 4.2.1.1.1) The entire 3.9 kb BpsA gene was

amplified by PCR using Phusion® high fidelity polymerase using primers with an upstream NdeI site and downstream HindIII site incorporated. The PCR product was ligated into pSX, giving the final BpsA expression construct pSX:BpsA. The identity of this construct was verified by sequencing and expression tests were conducted in *E. coli* under standard growth conditions (37 °C, 200 rpm) with a final concentration of 0.5 mM IPTG used to induce protein expression. Strong induction of a band of ~140 kDa in response to IPTG was identified by SDS-PAGE, indicating that BpsA was being expressed. However, no pigmentation was observed for these cells, consistent with previous observations that the endogenous PPTases of *E. coli* are unable to activate BpsA.

6.3.2 – Assessment of BpsA activity in *Pseudomonas* species

pSX:BpsA was delivered to *P. aeruginosa*, *P. syringae* and *P. putida* by electroporation. As colonies developed, there was clear evidence of blue pigmentation for *P. putida* harbouring pSX:BpsA even though IPTG was absent from the medium. For *P. syringae* and *P. aeruginosa* there was no evidence of blue pigmentation, with or without IPTG addition. Figure 6.5A shows *P. putida* harbouring the plasmid pSX:BpsA. At the time it was assumed that the above result was due to the fact that only *P. putida* possessed a PPTase capable of recognising BpsA, however later co-expression studies in *E. coli* (6.3.1) showed that PcpS, the PPTase of *P. aeruginosa* [69] was also capable of activating BpsA. The lack of pigment production in *P. aeruginosa* was most likely due to recombination mediated loss of plasmid sequence. In *P. putida*, it was also observed that expression of BpsA was unstable. Cells that had been grown in the presence of IPTG, once sub cultured or streaked onto a fresh plate, lost the ability to express BpsA but retained gentamycin resistance. Repeated re-streaking or sub-culturing, in the absence of IPTG, also led to the loss of BpsA expression. This instability was taken as a sign that the product of BpsA, indigoidine, was toxic to the cells leading to selection of cells in which recombination mediated loss of part of the plasmid sequence had occurred. This theory was supported by the fact that plasmid DNA isolated from cells that had lost the ability to express BpsA contained a truncated version of the *bpsA* gene.

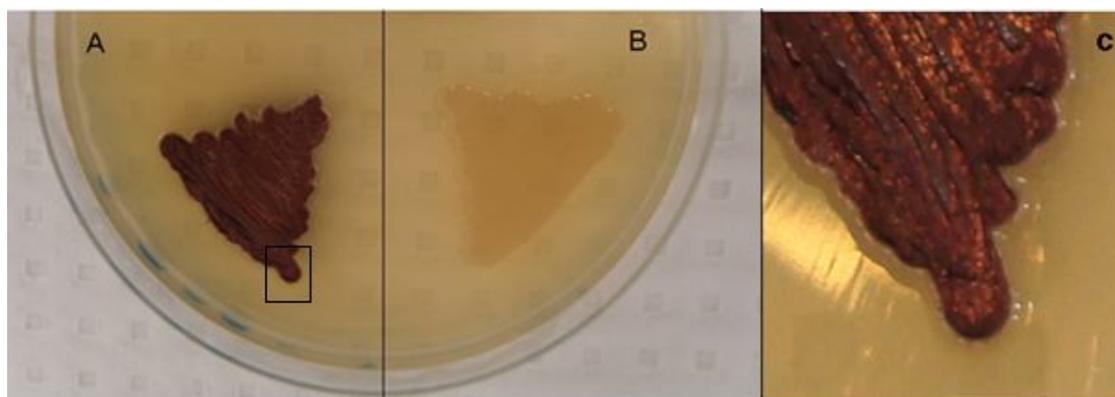


Figure 6.5 – Expression of BpsA in *P. putida*

A- *P. putida* expressing BpsA from the plasmid pSX:BpsA.

B- *P. putida* transformed with empty pSX plasmid.

C- Close up of the boxed region in panel A, it can clearly be seen that the actively replicating bacteria toward the edge of the patch have stopped producing indigoidine. Isolation of plasmid DNA from such cells showed that the *bpsA* gene was truncated.

6.4 – Development of plasmid based platforms for the co-expression of BpsA and an activating PPTase in *E. coli*

An important consideration when developing an *E. coli* BpsA expression system was tightly regulating control of BpsA expression. Although pSX:BpsA was suitable for IPTG regulated expression in *E. coli*, this plasmid was prone to leaky expression which, coupled with the toxicity of BpsA expression, was a potential source of plasmid instability. To circumvent this problem, the *bpsA* gene was sub-cloned into pET28a+, where it would be under the control of the tightly regulated T7 promoter [204, 205]. Use of pET28a+ had the additional advantage of adding an N-terminal 6his tag that would enable Ni-NTA purification of BpsA and any modified variants generated. Plasmids for IPTG regulated co-expression of three PPTases were also created. These PPTases were PcpS from *P. aeruginosa*, Sfp from *B. subtilis* and PP1183 from *P. putida*. The gene for each PPTase was cloned into both pET28a+ and pCDFduet1 expression vectors. PPTase and BpsA expression constructs were verified by sequencing and assessed for protein expression by SDS PAGE. A co-expression strain for BpsA with each PPTase was then constructed and it was found that expression of any one of

the three PPTases in *E. coli* resulted in activation of BpsA, as evidenced by strong blue colouration of cells grown on solid media and supernatant from liquid culture.

6.5 – Development and testing of T-domain swapping platforms

Once it had been established that co-expression of BpsA and an activating PPTase in *E. coli* was possible, construction and testing of plasmids for substitution of foreign T-domains into BpsA was undertaken. Three different substitution vectors (pBPSA1-3) were created, each differing in the up and downstream T-domain splice points.

6.5.1 – Substitution scheme one

For the first attempt at swapping foreign T-domains into BpsA, the definition of T-domain boundaries was set according to the prediction of the TIGR NRPS analysis server. The sequence either side of the T-domain was cloned into the expression vector pET28a+, such that the resulting construct contained the *bpsA* gene without its native T-domain and incorporated two unique restriction sites to enable introduction of foreign T-domains (Figure 6.6). Initially a positive control was created in which the native *bpsA* gene was recreated, with a NheI site introduced upstream of the T-domain and a KpnI site introduced downstream. The purpose of this control was to determine whether introduction of the two restriction sites, each of which added two foreign aas to the protein, would impair enzyme function. As described in 6.6, the resulting modified BpsA was still functional, although *in vivo* pigment synthesis efficiency was reduced. Once it had been established that introduction of restriction sites did not totally destroy function, foreign T-domains were substituted in place of the native domain. A summary of *in vivo* pigment synthesis efficiency for all modified BpsA enzymes is given in Table 6.2. A graphical overview of all swapping schemes is given in Figure 6.6.

6.5.2 – Substitution scheme two

Substitution scheme two was designed to address an inherent flaw in the first scheme. This flaw was not detected until a more detailed analysis of the splice points used in the first scheme. Alignment of the T-domain of BpsA against that of TycC3, a T-domain with solved structure, revealed that splice points had been inappropriately predicted by the TIGR NRPS analysis server, with the up and downstream points falling within the first and last α -helix of the domain respectively. Results from previous NRPS modification experiments have suggested that splice points should be introduced in the disordered linker regions that connect adjacent domains [13, 15, 16], a requirement that was addressed in the design of substitution scheme two. A positive control, consisting of the native gene with introduced restriction sites was again created and assessed for pigment synthesis efficiency. Surprisingly, the efficiency of this protein was lower than that created using scheme one. Four foreign T-domain substitutions were also carried out using scheme two, and in each case the modified protein was found to have equal or lower efficiency than the equivalent protein created using scheme one. Based on this result it was decided that substitution scheme two should not be used for any subsequent directed evolution experiments.

6.5.3 – Substitution scheme three

The third T-domain substitution scheme was created after successful directed evolution experiments had already been conducted using scheme one. The third scheme also aimed to address a specific problem that had been identified with scheme one. This problem was the presence of foreign aas as a result of restriction site introduction. In the third scheme, restriction sites were designed such that once a foreign domain had been introduced, there would be a seamless transition between native BpsA sequence and the foreign domain introduced. This was achieved by introducing silent restriction sites that did not alter the sequence of the translated protein. The design and construction of recombinant genes using substitution scheme three is outlined in 2.15.1. This method allows splicing of NRPS domains without the introduction of additional foreign protein sequence, a potentially valuable

technique for future modification experiments. Elimination of additional coding sequence introduced via restriction sites was important for interpretation of directed evolution results. Since there were no additional foreign aas present in constructs created using scheme three, it can be assumed that any improved variants from directed evolution experiments using this scheme represented adaptation of a foreign T-domain to interact with other BpsA domains, rather than adaptation to deleterious aas introduced via restriction sites. Swapping scheme three utilised pCDFduet1 as an expression vector for modified BpsA variants due to a lack of appropriate restriction sites in pET28a+.

6.5.4 – System for description of modified BpsA variants.

A total of fifteen different modified *bpsA* genes were created using swapping schemes 1-3, as outlined in table 6.1. Substitutions created using scheme one are denoted with the prefix “o”, those created using scheme 2 have the prefix “n”, those created using scheme three have the prefix “sl”. The prefix is followed by an identifier of the T-domain substitution. For example oBpsA was created using swapping scheme 1 and contains the native BpsA T-domain, nPvdD was created using swapping scheme 2 and contains the T-domain from the first module of PvdD. The protein expressed from the construct slBpsA is as identical aa sequence to WT BpsA.

Swapping scheme used	Origin of substitute T-domain						
	BpsA	PvdD	EntF	DhbF	EntB	PsT	
1	oBpsA	oPvdD	oEntF	oDhbF	oEntB	oPsT	
2	nBpsA	nPvdD	nEntF	X	X	nPsT	
3	slBpsA	slPvdD	slEntF	slDhbF	slEntB	X	

Table 6.1 – system for describing modified BpsA genes/proteins.

BpsAT= native BpsA T-domain, PvdD = T-domain from module one of *P. aeruginosa* PvdD, EntF = T-domain from *E. coli* EntF, Dhbf= T-domain from second module of *B. subtilis* Dhbf, EntB= Aryl carrier protein from *E. coli* EntB, PsT = T-domain from second module *P. syringae* Pspph1926. Prefix o, n or sl denotes swapping scheme used in generation of modified BpsA variant. Substitutions not tested in this study are denoted with an X.

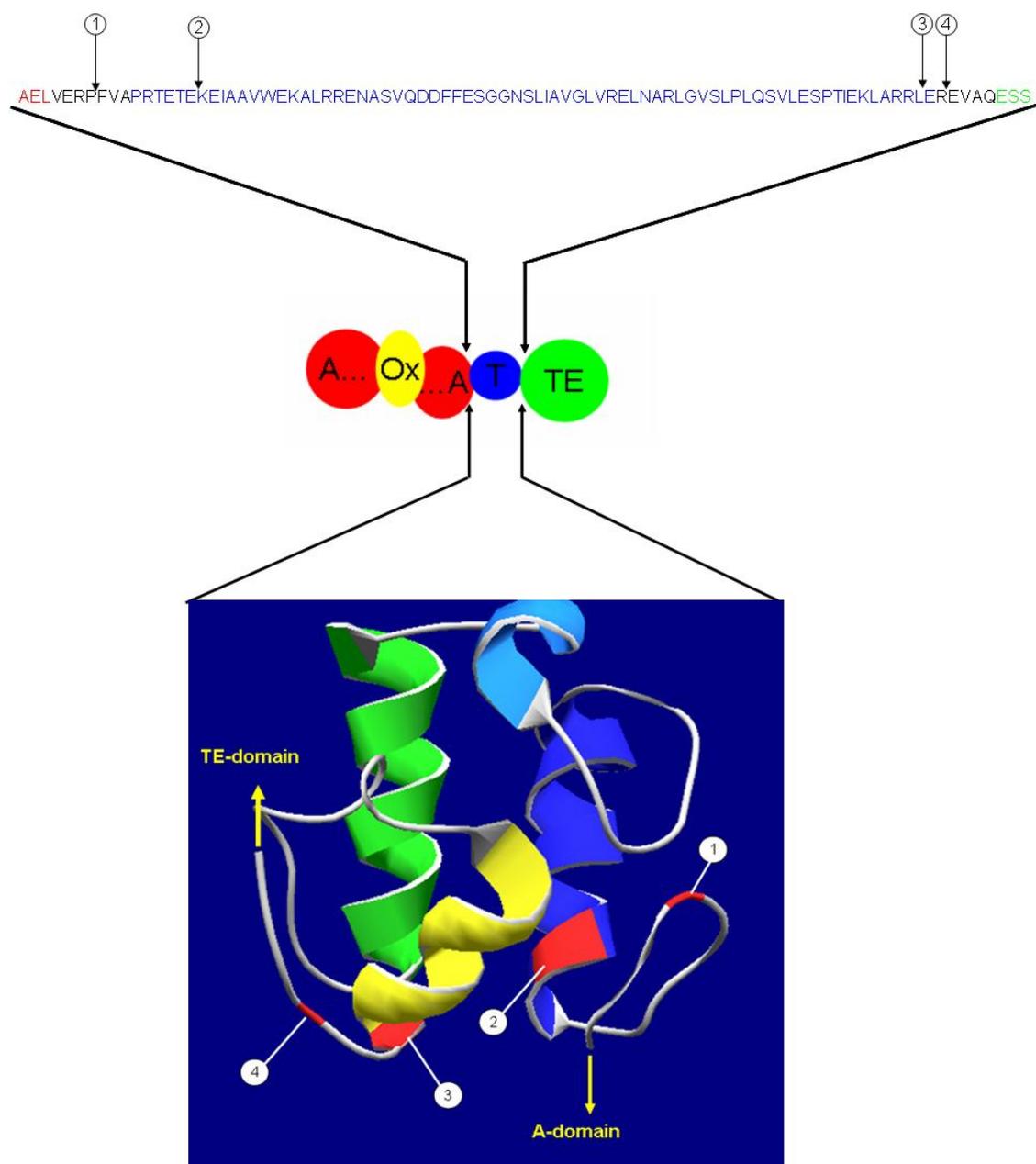


Figure 6.6 – Splice points for T-domain substitution schemes 1-3.

Substitution scheme one uses a KpnI site at (2) and an NheI site at (3), substitution scheme two uses a KpnI site at (1) and an NheI site at (4). For both schemes one and two, restriction sites resulted in insertion of aas between the residues indicated with an arrow. KpnI inserts GT, NheI inserts AS. For substitution scheme three, there were no insertions or alterations in aa sequence other than the substituted T-domain, which was inserted between (1) and (4). This was achieved by introducing silent restriction sites that did not alter the aas encoded in the *bpsA* gene. These sites were located up stream of (1) and down stream of (4). The intervening sequence between the silent restriction sites and splice points (1) and (4) was added to the primers used to amplify substitute T-domains resulting in a seamless transition between BpsA sequence and substituted T-domain sequences at points (1) and (4). The blue panel shows the location of splice points 1-4 on a structural model of the BpsA T-domain; note that splice points (2) and (3) disrupt α -helices.

6.6 – Assessment of modified BpsA variants created using swapping schemes 1-3

Each of the modified BpsA derivatives was assessed for pigment synthesis in *E. coli*. Conditions optimal for pigment synthesis were determined by systematically varying concentrations of L-glutamine (BpsA substrate) and IPTG (for induction of expression of both BpsA and the co-expressed PPTase) in culture medium. The effect of temperature on pigment production was also investigated. The reporter used for this experiment was oBpsA co-expressed in *E. coli* with PcpS. As illustrated in Figure 6.7 the optimal composition of media was LB supplemented with 100 mM L-gln, 0.5-1.0 mM IPTG and appropriate antibiotics. Growth at 18 °C resulted in markedly elevated pigment production compared to growth at 30 °C. Growth at 37 °C was not assessed in this experiment, as previous observation suggested no pigment production occurred at this temperature. Following this experiment, all future assessment of pigment production was carried out at 18 °C in LB supplemented with 100 mM L-gln, 0.5 mM IPTG and appropriate antibiotics, this medium is hence forth referred to as pigment development medium, used as both broth and agar.

Following optimisation of growth conditions, all modified BpsA proteins were assessed for pigment production capacity in *E. coli*. For each BpsA variant, the activating PPTase used was PcpS. Table 6.2 summarises the pigment production capacity of modified BpsA proteins expressed in *E. coli*. Quantification of pigment production was achieved by measuring absorbance at 590 nM, this being the previously determined absorbance maximum for indigoidine [80]. The method used for calculating *in-vitro* pigment synthesis capacity is outlined in 2.13. In some cases, pigment production was not detected using broth but was apparent after extended incubation in pigment development agar, in such cases the recombinant BpsA was presumed to have lower activity than any variant found to produce pigment in broth. In cases where pigmentation would not develop in broth or on agar, recombinant genes were deemed completely inactive.

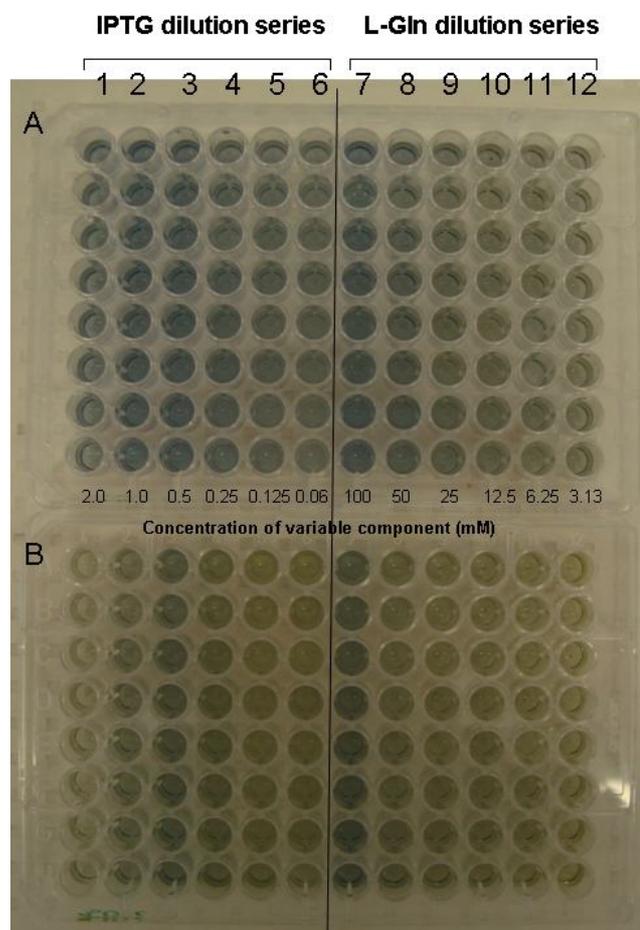


Figure 6.7 – Determination of optimal conditions for pigment synthesis

Columns 1-6 contain a two-fold serial dilution series from 2.0 to 0.0625 mM IPTG, with L-gln concentration fixed at 100 mM. **Columns 7-12** contain a two-fold serial dilution series from 100 to 3.13 mM L-gln with IPTG concentration fixed at 0.5 mM. Each well contains a final volume of 100 μ L inoculated to a uniform density with an overnight culture of *E. coli* harboring plasmids for the expression of oBpsA and *P. aeruginosa* PPTase. All wells contain 50 μ g/mL Kanamycin and 100 μ g/mL ampicillin. **Plate A** was grown for 16 h at 18 $^{\circ}$ C, **plate B** was grown for 16 h at 30 $^{\circ}$ C. Base medium was LB.

	Origin of substitute T-domain						
	BpsA	PvdD	EntF	DhbF	EntB	PsT	
Swapping scheme used	1	21.7 +/- 2.7%	0%	L	L	0%	0.773 +/- 0.278%
	2	1.62 +/- 0.14%	0%	0%	X	0%	0%
	3	100%	0%	5.14 +/- 0.28%	0%	0%	X

Table 6.2 – In vivo pigment synthesis capacity of recombinant BpsA genes.

In vivo pigment synthesis activity of all recombinant BpsA genes relative to wild type (set at 100%) as determined by measuring OD 590 of supernatant from cultures grown in pigment development medium at 18 $^{\circ}$ C. Error values are given as +/- one standard deviation. For a detailed description of how these values were determined refer to section 2.13. Substitutions not made are denoted with an X. Instances where activity was too low to be detected by OD 590 measurement, but apparent after extended incubation on plates, are denoted with an L

6.7 – Development of a high throughput screening process for directed evolution of recombinant *bpsA* genes

As outlined in Table 6.2, substitution of the native T-domain of BpsA for a foreign domain, consistently reduced the efficiency of pigment synthesis *in vivo*. This result agrees with published structural and biochemical data which suggests that T-domains undergo specific interactions with other NRPS domains during the biosynthetic process [15, 44, 47, 53, 93, 94, 192, 193]. Having established and tested a T-domain swapping platform, directed evolution experiments were conducted with the aim of identifying key residues involved in T-domain interaction with other BpsA domains. In these experiments the activity of recombinant BpsA variants created using swapping schemes one and three, that had impaired function, was improved. To achieve this, the substituted T-domains of these variants were randomly mutated using error prone (ep) PCR. The resulting products were then cloned into the corresponding BpsA T-domain swapping construct, to give a mutant library varying only in the sequence of the T-domain. Development of a robust, high throughput first tier screening procedure allowed qualitative identification of variants with improved T-domain function in BpsA. Improved variants were then subjected to a second tier screen to quantitatively assess pigment synthesis capacity *in vivo*. Based on the results of the second tier screen, selected improved clones were sequenced and the mutations responsible for improved function were determined.

6.7.1 – Development of a first tier agar plate based screening system.

To develop a first tier screen for improved pigment synthesis capacity, priority was given to screening as many clones as possible, as quickly as possible and with a minimum of manual input. Avoidance of false positives and quantitative assessment of activity were given low priority as these issues could be addressed in a second, lower throughput screen. At this stage it had already been established that pigment production was readily detectable in BpsA expressing *E. coli* patched on to agar plates, and that even weak activity (<0.7% of WT) could be detected following 24-48 h incubation on pigment development agar. To facilitate

screening however, pigmentation would have to be visible in single colonies immediately following transformation. This was now assessed.

Initial trials were run using *E. coli* harbouring a PP1183 expression plasmid. A plasmid for the expression of oBpsA was transformed into this background and plated on agar containing 100 mM L-gln and 0.5 mM IPTG. Under these conditions, it was consistently found that no transformants could be recovered. If, however, IPTG was omitted from the plates, transformation with high efficiency was achieved. This result suggested that toxicity arising from indigoidine production was rendering transformants inviable in the presence of IPTG.

To circumvent this problem, a novel system for on-plate induction was developed. Transformed cells were plated on media containing antibiotics and L-gln only, and incubated overnight to allow colonies to develop. Following colony development the entire agar slab was scooped from the plate and rested on the sterile lid. IPTG solution was then spread on the bottom of the empty plate and the agar slab replaced on top of the solution. Diffusion of IPTG into the medium then resulted in induction of protein synthesis in colonies on the top surface of the agar and development of strong pigmentation in individual colonies.

6.7.2 – Optimisation of library generation for directed evolution

An important consideration for directed evolution is the development of a reliable system for library generation. In this case libraries were generated by first conducting epPCR using a particular T-domain as template. The resulting PCR product was then introduced, via compatible restriction sites, into a plasmid borne copy of the *bpsA* gene, resulting in a library of *bpsA* genes with variant T-domain sequence. The most critical factor in this process is generation of high quality, restriction digested linear vector into which a digested PCR product can be introduced by ligation. In order to ensure that vector preparations for directed evolution were of consistently of high quality, a standardised protocol for preparation and quality assessment of digested linear vector was developed. This allowed consistent generation of libraries in which greater than 70 % of plasmids contained the desired insert. Library

generation protocols were progressively refined throughout the course of this study, eventually resulting in optimised protocol given in 2.11.1.

6.8 – Adaptation of a T-domain from DhbF to function in BpsA using directed evolution

The first directed evolution experiment conducted aimed to improve the function of oDhbF (i.e. the modified derivative of BpsA, in which the native T-domain had been replaced by a T-domain from the *B. subtilis* enzyme DhbF using substitution scheme 1). Assessment of the activity of oDhbF revealed initial activity was lower than 0.7% (Table 6.2). To improve the function of oDhbF, a T-domain variant library of approximately 25,000 insert containing clones was screened on pigment development agar plates, resulting in the recovery of 12 putative improved clones. Figure 6.8 gives an overview of the screening process and shows a typical improved clone on a pigment development agar plate. The 12 clones recovered from the plate base tier of screening were then subjected to a second tier screening procedure, to quantitatively assess pigment production during planktonic growth in a 96 well plate. Eleven of the initial hits were confirmed as having improved activity in this second tier screen. These were sent for sequence analysis, which revealed four different single aa changes, two of which were recovered multiple times. The procedure for library generation, screening and the controls employed for this and subsequent experiments, is described in detail in 2.11. A summary of all sequence analysis from all experiments is given in the discussion section (tables 6.3 and 6.4), complete annotated sequence alignments can be found in Appendix 1. In the second tier screens for all directed evolution experiments two controls were included. The un-improved enzyme is included as a negative control in each case and either oBpsA or WT BpsA as a positive control. It is important to note that although in some instances the activity of improved clones appears to approach or exceed that of the positive control, this is an artefact of the screening process. This artefact arises due to saturation of pigment production. This saturation is likely to be a result of indigoidine toxicity inhibiting cell growth at higher concentrations.

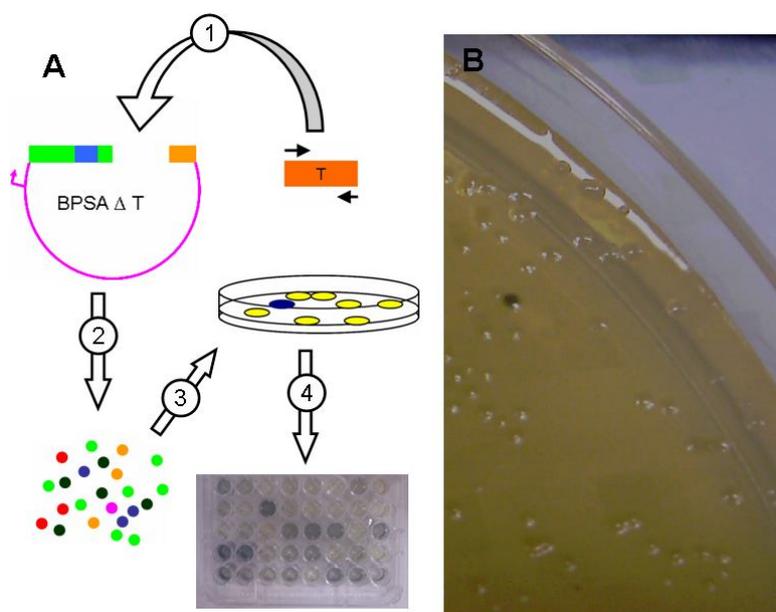


Figure 6.8 – Overview of directed evolution procedure

A) A schematic representation of the directed evolution protocol. (1) T-domain is amplified by error prone PCR and amplicon is ligated into plasmid borne copy of BpsA gene. (2) The resulting library of T-domain variants is transformed into cells harbouring an activating PPTase. (3) Transformed cells are plated on pigment development agar and improved clones are recovered. (4) Improved clones are subjected to a second, quantitative screen for activity in a 96 well plate format

B) Photograph of a pigment development plate on which an improved variant of oDhbF can be seen.

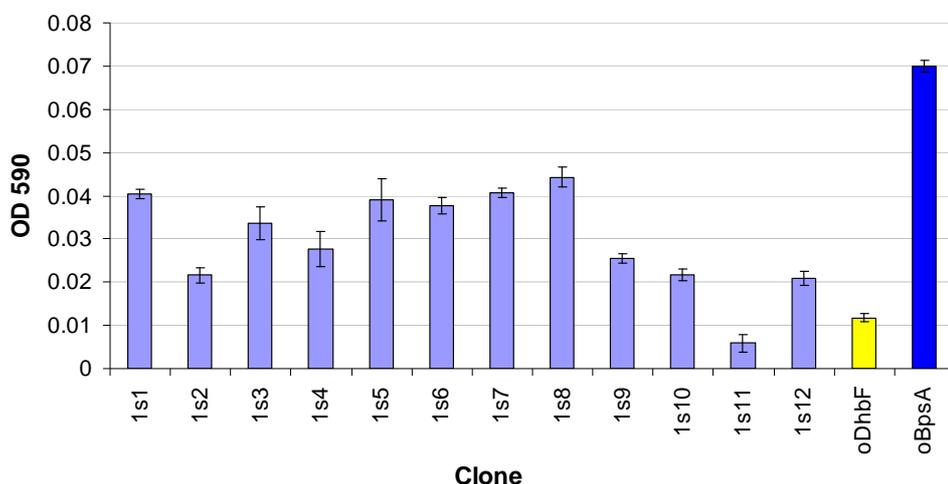


Figure 6.9 – Quantitative assessment of activity of improved variants of oDhbF

In vivo activity of improved clones was determined by measuring OD 590 of supernatant for triplicate cultures of each clone grown in a 96 well plate as described in 2.11.3. The yellow bar is unimproved oDhbF included as a negative control. The dark blue bar is oBpsA included as a positive control. Sequence of all variants except 1s11 was determined. Error bars are presented as +/- one standard deviation.

6.9 – Adaptation of a T-domain from PvdD to function in BpsA

The native T-domain of BpsA was replaced with the first T-domain from the *P. aeruginosa* NRPS PvdD using all three substitution schemes described in 6.4. In all cases this substitution was found to result in completely inactive BpsA. *E. coli* harbouring these modified BpsA variants and an activating PPTase were unable to produce any discernable blue pigment on pigment development agar or in pigment development broth regardless of incubation time, activating PPTase or temperature. Both oPvdD (swapping scheme 1) and slPvdD (swapping scheme 3) were improved using directed evolution. The cloning and transformation protocol used for evolution were modified from those used for oDhbF. This allowed generation of much larger variant libraries and an estimated 5-6 x 10⁵ clones were screened for both oPvdD and slPvdD.

6.9.2 – Recovery and assessment of hits

For each hit recovered from the agar based screen, the time for pigmentation to develop was noted and only the fastest developing hits from each library were chosen for further assessment. A total of 44 putative improved variants of slPvdD and 14 putative improved variants of oPvdD recovered from the first tier screen were subjected to second tier screening (Figures 6.10 A and B). The sequence of all 14 oPvdD clones was determined. For hits derived from slPvdD, the 8 clones indicated were analysed. The most active clone recovered from improvement of slPvdD, designated 3kF0, was chosen for further improvement by a second round of epPCR and selection.

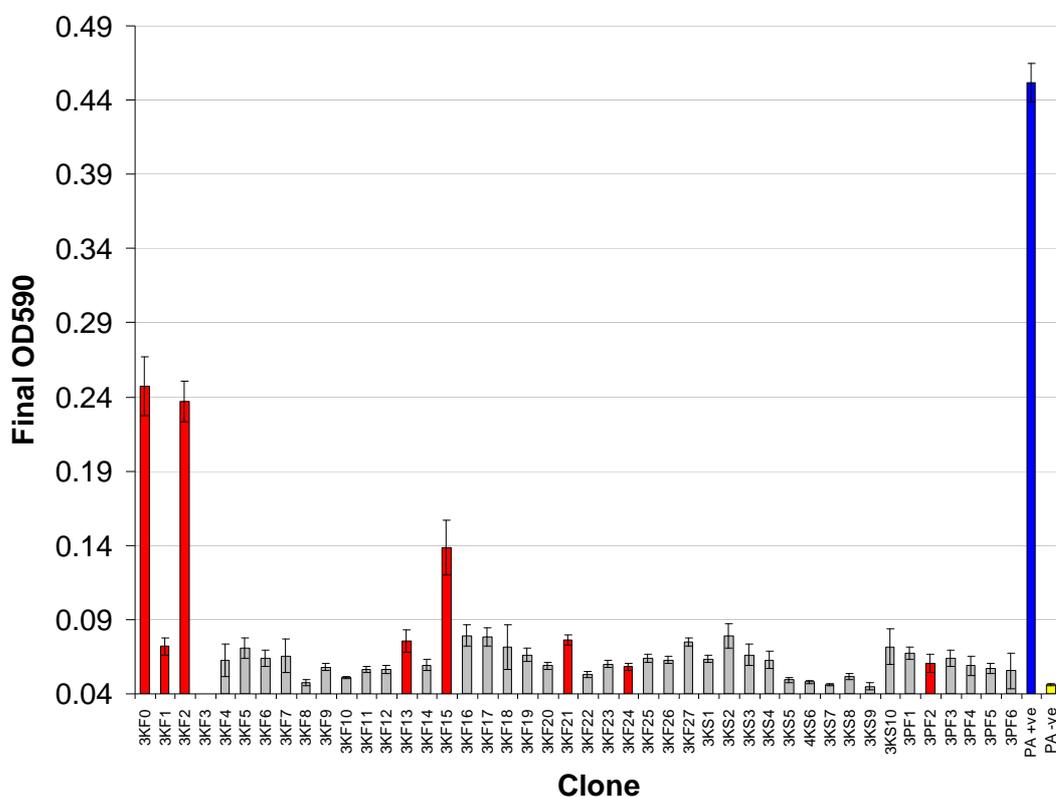


Figure 6.10A – Quantitative assessment of activity of improved variants of sIPvdD

In vivo activity of improved clones was determined by measuring OD 590 of supernatant for quadruplicate cultures of each clone grown in a 96 well plate as described in 2.11.3. The yellow bar is unimproved sIPvdD included as a negative control. The dark blue bars is WT BpsA included as a positive control. Clones for which sequence was determined are indicated by red bars. Selection aimed to determine sequence for clones with a wide variety of pigment synthesis activity. Error bars are presented as +/- one standard deviation.

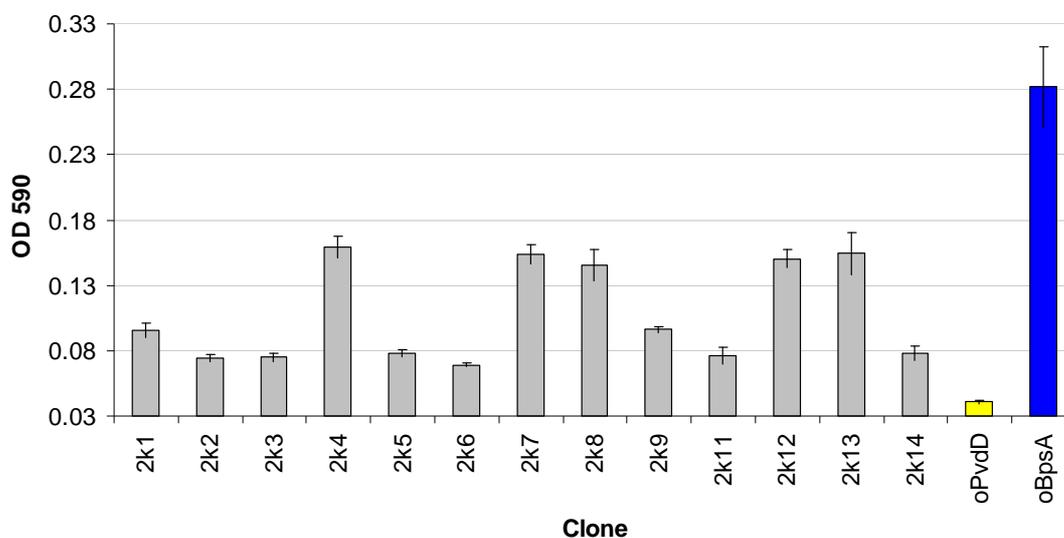


Figure 6.10B – Quantitative assessment of activity of improved variants of oPvdD

In vivo activity of improved clones was determined by measuring OD 590 of supernatant for quadruplicate cultures of each clone grown in a 96 well plate as described in 2.11.3. The yellow bar is unimproved oPvdD included as a negative control. The dark blue bar is oBpsA included as a positive control. Sequence of all variants was determined. Error bars are presented as +/- one standard deviation.

6.9.3 – Second round evolution

For second round evolution the T-domain of clone 3kF0 was used as template for the epPCR reaction instead of wt PvdD T-domain. Transformation and screening were conducted as for the first round of evolution, and 22 hits were recovered from an estimated total of 600,000 clones screened. Screening proved more difficult than for the first round of evolution due to the higher levels of pigmentation of unimproved clones. Selection of improved clones required qualitative appraisal of relative pigmentation by naked eye, including subjective assessment of local variations in induction, presumably arising due to uneven diffusion of IPTG. Nonetheless, the number of false positives recovered proved to be low with 14 of the 22 hits recovered showing a clear improvement in function in the second tier of the screen (Figure 6.11).

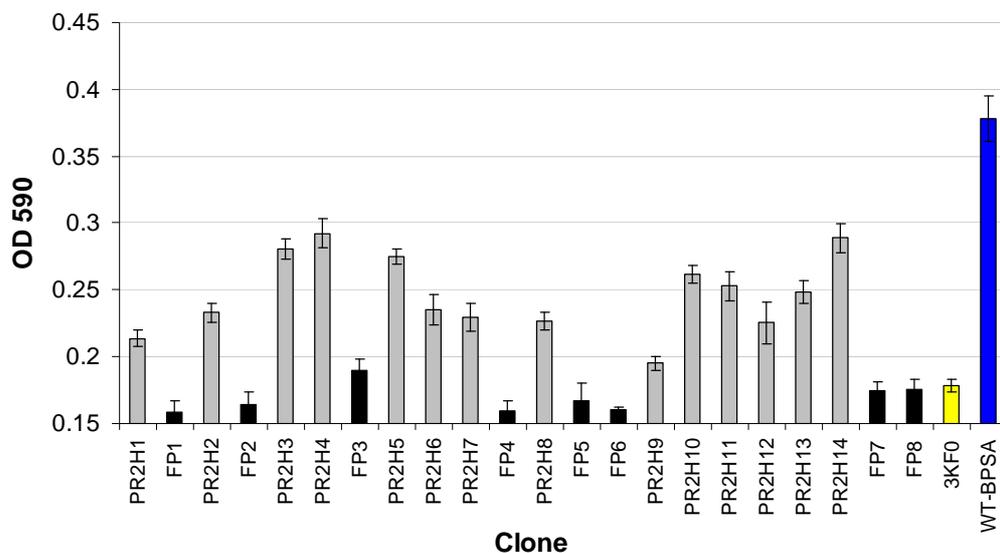


Figure 6.11 – Quantitative assessment of activity of improved variants of 3KFO

In vivo activity of improved clones was determined by measuring OD 590 of supernatant for quadruplicate cultures of each clone grown in a 96 well plate as described in 2.11.3. The yellow bar is the best hit from round one, 3kf0, included as a negative control. The T-domain of 3kf0 was the starting template for error prone library generation in this experiment. The dark blue bar is WT BpsA included as a positive control. False positives are indicated by black bars, sequence was not determined for these clones. Error bars are presented as +/- one standard deviation.

6.10 – Adaptation of the T-domain of EntF to function in BpsA

Of all the domain substitutions created, sEntF had the highest *in vivo* pigment synthesis capacity prior to directed evolution. This may be a consequence of the T-domain of EntF being the only substitute T-domain employed in swapping scheme three that is required to interact with a TE-domain in its native setting. sEntF was also improved by directed evolution, following the same protocols used for the first round evolution of oPvdD and sIPvdD. Libraries were transformed into electrocompetent cells harbouring the PPTase of *P. putida* and approximately 500,000 clones were assessed in the first tier of screening. As for the second round evolution of sIPvdD, the fact that unimproved sEntF exhibited strong pigmentation on pigment development agar made selection of improved clones more difficult than for the other first round evolution experiments described in this chapter. Nonetheless, the

screening procedure again proved to be robust, with 39 of the 46 hits recovered from the first tier showing elevated pigment synthesis in the second tier (Figure 6.12). Of the 39 improved clones recovered, the sequence of 8 was determined. Sequence analysis of plasmid DNA from an additional 12 clones failed due to poor sample quality. It was decided not to reassess these clones owing to the large data set already accumulated from the various directed evolution experiments.

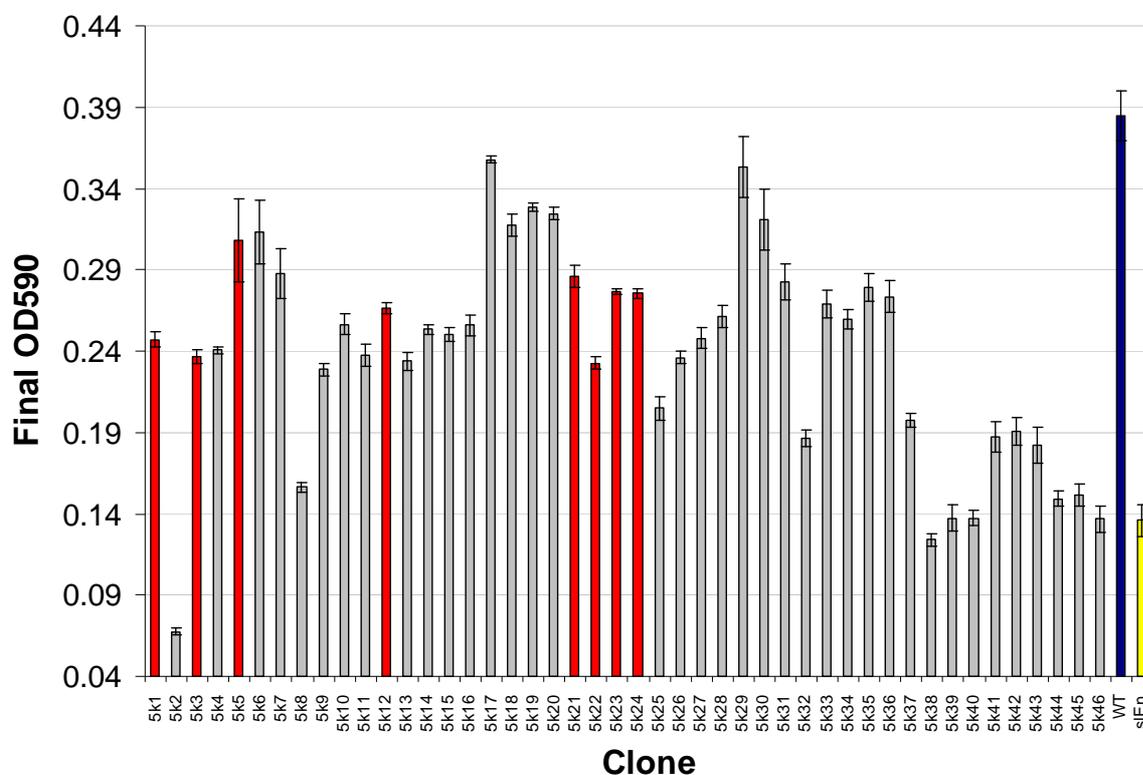


Figure 6.12 – Quantitative assessment of activity of improved variants of slEntF

In vivo activity of improved clones was determined by measuring OD 590 of supernatant for quadruplicate cultures of each clone grown in a 96 well plate as described in 2.11.3. The yellow bar is unimproved slEntF and is included as a negative control. The dark blue bar is WT BpsA included as a positive control. Clones for which sequence was determined are indicated by red bars. Error bars are presented as +/- one standard deviation.

6.11 – Summary of sequence data from directed evolution experiments

The total number of improved clones analysed from all directed evolution experiments was 59. In this set of improved clones there were a total of 26 different mutations occurring at 22 different positions. Five of these mutations had a neutral or deleterious effect on activity and were only recovered by association with a second, beneficial mutation. These were discounted from analysis. The remaining mutations all contributed positively to function of BpsA, as evidenced by improving in vivo pigment synthesis activity when present as a single mutation, or by additively improving activity when in combination with a second mutation previously shown to increase pigment synthesis.

These mutations were viewed as falling into one of two categories. The first category (type I) comprises mutations in which the sequence of the substituted domain had become more similar to the original sequence of BpsA, by reverting to a residue identical or more similar (as assessed by BLOSUM 62 matrix) to that originally found at the same location in BpsA. The second category (type II) consists of mutations which resulted in no change in similarity or the sequence of the substitute domain becoming less similar. In interpreting this data, it was assumed that the native sequence of the BpsA T-domain has evolved for optimal interaction with other BpsA domains, an assumption consistent with the observation that all substituted T-domain constructs exhibited impaired activity in this study. Bearing this assumption in mind, type I mutations would appear to highlight positions at which a specific residue is required for optimal interaction. Conversely, type II mutations may either represent optimisation of generic T-domain functions that are independent of interaction with other domains; or else provide an alternative mode of domain interaction to that which originally evolved in BpsA. A summary of the mutations recovered from first and second round directed evolution experiments is given in Tables 6.2 and 6.3 respectively.

Type I mutations					
	+4 (V)	+7 (V)	+14 (L)	+23 (V)	+24 (L)
oDhbF	A→V x 2	- M	- M	L →I x 1	- F
oPvdD	T →I x 5	- I	Q→L x 2	- V	F →L x 3
slPvdD	T→I x 3, T→A x1	- I	- Q	- V	F →I x 3
slEntF	- M	A→V x 2	- V	- V	- M

Type II mutations					
	-3 (G)	+1 (L)	+2 (I)	+6 (L)	+17 (S)
oDhbF	- G	- L	L→H x 5	- M	E→K x 4
oPvdD	- G	- L	- L	V→M x 6	- D
slPvdD	G→S x 1	- L	- L	- V	- D
slEntF	- G	L→I x 6	- L	- L	- Q

Table 6.3 – Type I and II mutations recovered from first round directed evolution experiments

The location of mutations is given in the top row with the residue present in the native T-domain of BpsA indicated in parentheses. The left hand column indicates the directed evolution experiment for which mutations were recovered. Text in bold indicates mutations which occurred in a substitute T-domain and the number of times it was recovered. Instances where a mutation was not recovered are indicated with a dash, followed by the residue present in the substitute domain.

Type I mutations						
-17 (R)	-16 (R)	-8 (D)	-1 (N)			
G→E x 1	V→A x1	N→D x3	H→Q x1			

Type II mutations						
-39 (R)	-5 (E)	+2 (I)	+8 (R)	+15 (Q)	+26 (S)	+38 (R)
R→L x2, R→C x1	E→D x1	L→Q x 2	S→F x 1	G→L x 1	R→W x 2	R→T x 1

Table 6.4 – Type I and II mutations recovered from second round evolution of 3kF0

Clone 3kF0 is an improved variant of slPvdD. This clone had two mutations, +4T→I and +24 F→I. The mutations recovered from a second round of mutagenesis and selection using 3kF0 as a precursor are shown in this table.

6.12 – Discussion

6.12.1 – Interpretation of results from T-domain swapping experiments.

Unless otherwise stated, the position of residues and mutations described in the remainder of this chapter is given relative to the 4'-PP attachment site of a T-domain, the serine at this position is absolutely conserved in all T-domains. In all instances, substitution of the WT BpsA T-domain by a foreign T-domain resulted in a drastic reduction in pigment synthesis *in vivo*. This result is consistent with disruption of native T-domain interactions impairing enzyme function. Irrespective of the swapping scheme used, two substitutions resulted in complete ablation of activity: the aryl carrier protein (ArCP) from the *E. coli* PKS enzyme EntB; and the first T-domain from the *P. aeruginosa* NRPS PvdD. In the case of the EntB swap, the loss of activity was not surprising, given that ArCPs interact with a very different set of partner domains in PKS enzymes. Restoration of function for recombinant BpsA harbouring the EntB ArCP by directed evolution was attempted (results not shown), however in spite of screening over 10^6 clones from an epPCR library, no functional variants were recovered. The reason for loss of function arising from the PvdD T-domain substitution is less obvious. Analysis of the sequence of this T-domain reveals the presence of a thr residue located at +4. As described in Appendix 1, alignment of 60 different T-domains showed that a hydrophobic residue was always present at this position in T-domains immediately upstream of a TE-domain. This finding suggests that the PvdD T-domain may have been unable to correctly interact with the TE-domain of BpsA due to the thr residue at +4 in this domain. In contrast, the remaining T-domain substitutions did not completely destroy BpsA function, and in all cases a hydrophobic residue was present at position +4 in the substituted T-domain. It is also interesting to note that the two alternative T-domains that resulted in recombinant BpsA enzymes with the highest activity (EntF and PsT) are both found upstream of TE-domains in their native setting. The results of the T-domain swapping experiments conducted therefore suggest that specific features of T-domains located upstream of a TE-domain are specialised for interaction between the two domains. Further support for this idea can be found in the bioinformatic analyses described in appendix 2 and discussed in 6.2.1. Since disruption between the T and TE-domains of BpsA is the most likely reason for loss of activity in

recombinant protein, it is reasonable to assume that many of the mutations recovered from directed evolution experiments act to reduce this disruption.

6.12.2 –Structural analysis of type I mutations from first round evolution experiments

There were five separate locations at which type I mutations occurred in first round directed evolution experiments. These positions were +4, +7, +14, +23 and +24. In the native T-domain of BpsA, each of these positions is occupied by either a val or leu residue. Figure 6.13A shows these positions mapped onto the structural model of the native BpsA T-domain in the TE-state. The leu located at position +24 forms part of the hydrophobic interface between the T and TE-domains as indicated in Figure 6.13B. Of particular note, previous mutational analysis of the T-domain of EntF found that mutation of the EntF equivalent of +24 destroyed enzyme function by disrupting T-TE communication [93].

Positions +7, +14 and +23 are occupied by val leu and val respectively. None of these residues are predicted to be involved in the interface between the T and TE-domains in the BpsA model. However, one common feature of the hydrophobic sidechains of these residues is that they are all orientated toward the core of the T-domain itself, so may be involved in stabilisation of the TE- interaction conformation of the T-domain. Alternatively they may be involved in interaction with the A or Ox-domain of BpsA. Unfortunately, due to lack of suitable templates structures, models for T-domain interaction with the A and Ox domains of BpsA could not be generated.

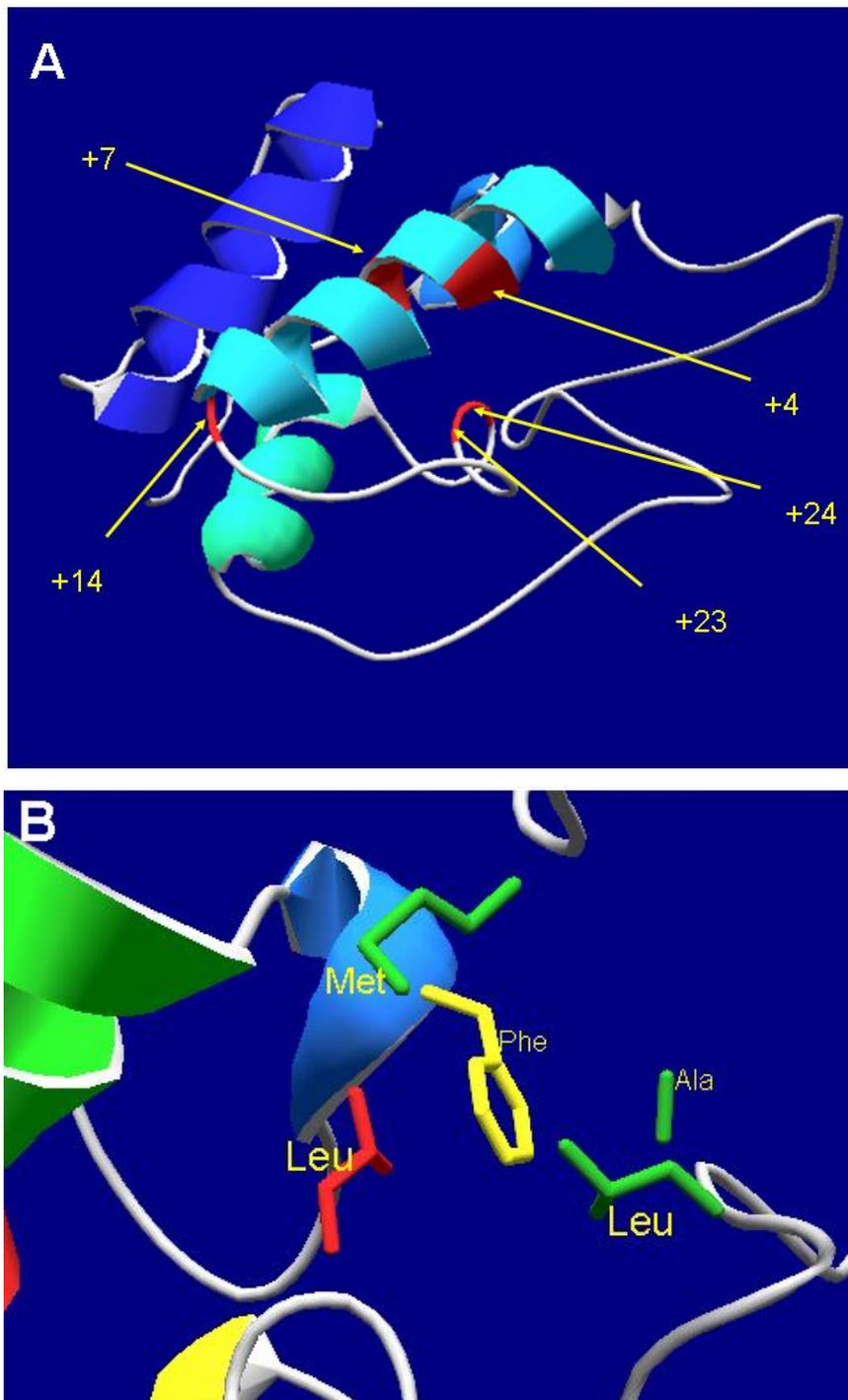


Figure 6.13 – Structural analysis of type I mutations

A- The locations of all type I mutations from first round directed evolution experiments are mapped onto a structural model of the BpsA T-domain in the TE-state.

B- leu +24, shown in red, forms part of the hydrophobic interface between the T and TE-domains in the structural model of BpsA. Side chains of other residues forming the interface are shown. Residues in green are part of the TE-domain, whereas the yellow phe is part of the T-domain.

As outlined in appendix one and section 5.4, evidence from bioinformatic analysis and domain swapping experiments suggests that position +4 in a T-domain may play an important role in the interaction with downstream domains. For interactions with TE-domains, sequence comparisons suggest that a hydrophobic residue is required at +4. Consistent with this hypothesis, mutations giving rise to a hydrophobic residue at this position were recovered multiple times in improved clones arising from the evolution of oPvdD, sIPvdD and oDhbF. Analysis of *in vivo* pigment synthesis data from the two PvdD evolution experiments reveals that a T→I mutation at this position results in the largest relative increase in pigment production of any of the mutations recovered.

In the structure 2roq, solved for the T/TE-domains of EntF [192], a met is located at +4. This methionine protrudes from the start of helix 2 of the T-domain and its side chain is oriented toward helices 4 and 5 of the TE-domain. In the elucidation of this structure, 20 similar conformers were observed. In these conformations helices 4 and 5 of the TE-domain are particularly mobile. This mobility was interpreted as providing a fluidity that might facilitate delivery of the 4'PP arm and attached substrate to the active site of the TE-domain, prior to product release [192]. A selection of conformations highlighting this fluctuating motion is shown in Figure 6.14A. It is possible the presence of a hydrophobic residue at S+4 actually serves to destabilise interaction between the T and TE-domains (i.e. preventing formation of strong interactions such as H-bonds or salt bridges), with the resulting mobility promoting entry of the 4'PP arm and attached substrate into the active site of the TE-domain (Figure 6.14B). In addition to promoting mobility that might be important for substrate transfer, this might prevent overly stable interactions of the T-domain with any individual partner domain, a situation which would be detrimental to interaction with other domains.

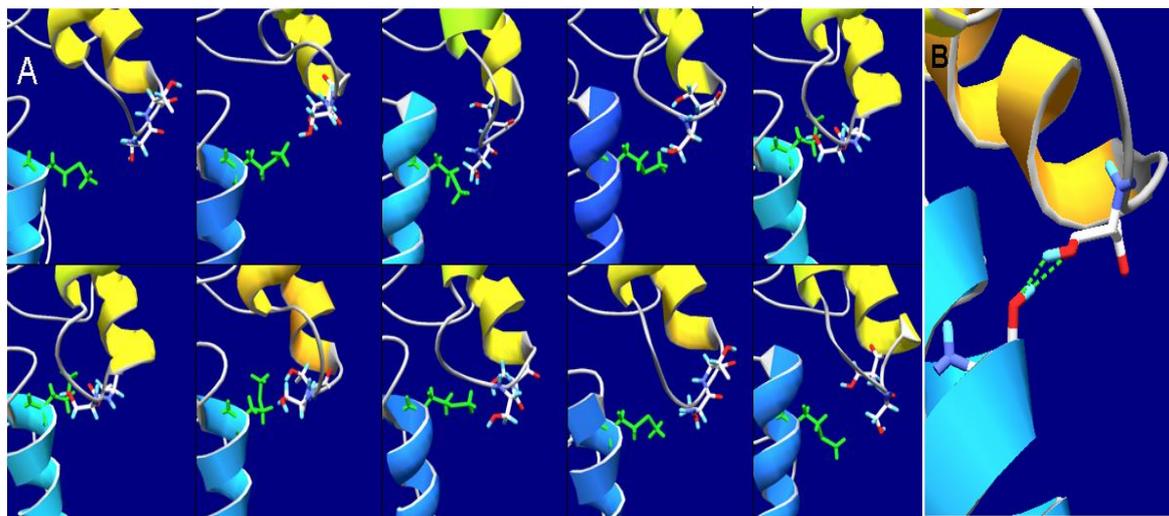


Figure 6.14 – Dynamic nature of T-TE structure in EntF

A- A selection of NMR conformers elucidated for the T and TE-domain of EntF are shown. Helices 4 and 5 of the TE-domain (yellow) were determined to rapidly fluctuate, and overlap helix 2 (blue) of the T-domain. Position +4 in this structure is occupied by a met (green). The side chain of this met comes into close proximity (<2.5 Å) to a ser and Thr residue protruding from the loop between helices 4 and 5 of the TE-domain.

B- If met +4 is replaced with a potential H-bond donor/acceptor this might result in reduction of mobility by formation of inter-domain H-bonds. This possibility was investigated by substituting a Thr residue at S+4 in the structure 2roq. As illustrated this could potentially result in formation of inter-domain H-bonds (dashed green lines). Structure was generated using the mutate, torsion and energy minimisation functions of Deep View.

6.12.3 – Interpretation of type II mutations from first round evolution experiments

Type II mutations were recovered at five separate locations in improved clones from first round directed evolution experiments. Since these mutations result in residues of a substituted domain becoming less similar to those found at the equivalent location in the native BpsA T-domain, they are unlikely to represent positions which are crucial for interaction between the native T and TE-domains of BpsA. Perhaps the most important information to be gleaned from these mutations is that there are many different ways in which T-domain function can be optimised. This situation has also been observed in directed evolution of the T-domain of EntB, where a diverse set of mutations was recovered [92]. Since interaction depends not only on direct contacts and interface with a partner domain, but also on the ability of a T-domain to

adopt a particular conformation, these mutations may represent novel solutions to the conformational demands placed on a foreign T-domain located in BpsA.

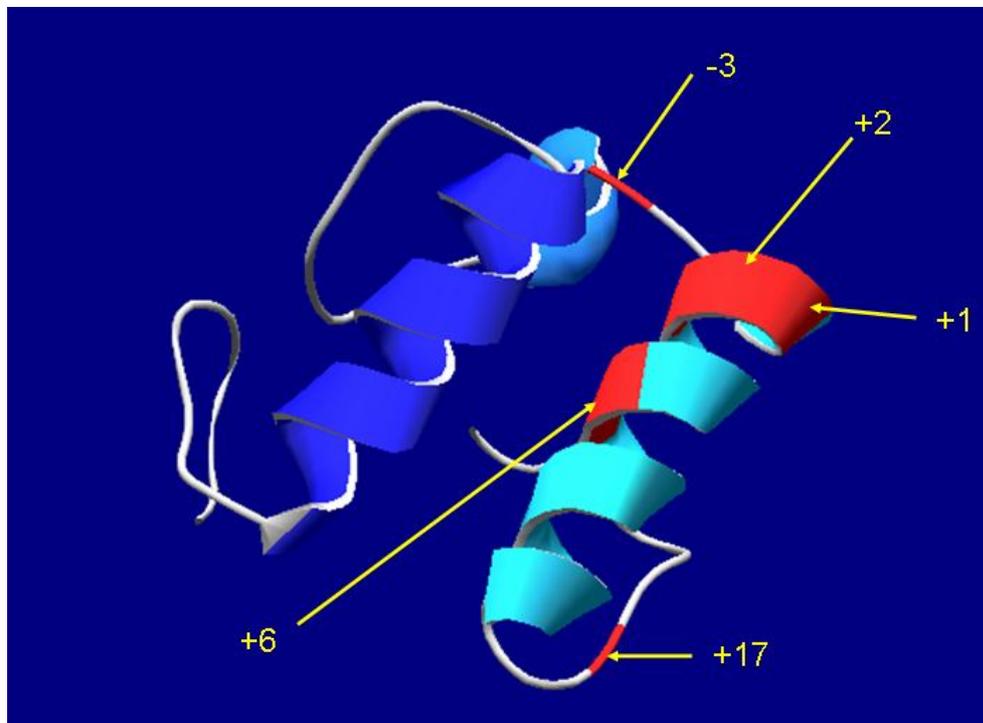


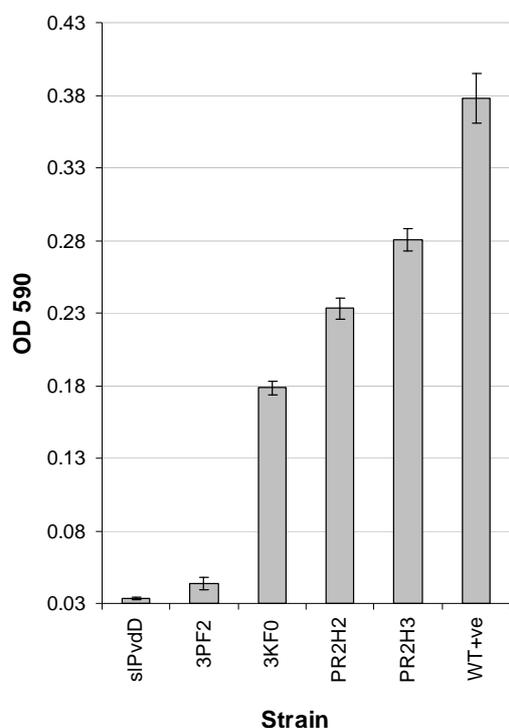
Figure 6.15 – Locations of type II mutations mapped onto a structural model of the BpsA T-domain

Identity of the mutations recovered at each location as follows -3 = G→S, +1 = L →I, +2= L →H, +6 = V →M, +17 E →K

6.12.4 –Interpretation of results from second round evolution experiment

For second round evolution experiments, the T-domain of the improved slPvdD variant 3kf0 was used as starting template. This clone showed the highest activity of any improved slPvdD variants and harboured type I mutations at positions +4 and +24. As discussed in the preceding sections, these mutations are likely to improve function by optimising interaction with the TE-domain of BpsA. The most immediately obvious difference between the patterns of beneficial mutation recovered from this second round of evolution, when compared to first round evolution is the increased diversity in location and identity of aa changes. In the 45 clones analysed from four separate first round experiments, there were a total of 10 locations at which 13 different aa changes were recovered. In the 14 clones analysed from a single second round evolution experiment, there were 11 locations at which 12 different mutations were recovered.

The location of these mutations has no overlap with those recovered from first round experiments. It is possible that the mutations recovered in the second round evolution experiment did not show up as hits in the first round experiments as they only improve function when acting synergistically with the S+4 and S+24 mutations already present in 3kf0. A subtly different possibility is that only mutations which addressed the primary barrier of T-TE communication restored activity in first round experiments and once this barrier had been overcome it revealed the effect of mutations which would otherwise have been masked. One clear conclusion that can be drawn from data arising from both first and second round evolution experiments is that the positive effect of mutations on activity is, at least in some cases, additive, as illustrated in Figure 6.17.



	+2	+4	+15	+24
sIPvdD	-	-	-	-
3pf2	-	-	-	F→L
3kf0	-	T→I	-	F→I
Pr2H2	I→Q	T→I	-	F→I
Pr2H3	I→Q	T→I	Q→L	F→I

Figure 6.16 – Additive effect of mutations on activity of sIPvdD

The graph compares *in vivo* pigment synthesis activity of sIPvdD and four improved derivatives of this protein recovered from first and second round evolution experiments. The table above shows the mutations present in each strain and gives their location relative to the site of 4'-PP attachment in the T-domain. Error bars are presented as +/- one standard deviation for quadruplicate cultures.

6.12.5 – A proposed model for T-domain interaction with the TE-domain of BpsA

The picture of T-domain interaction that is emerging, from this and previous studies is one of dynamic fluctuation. The interaction between the T-domain and other NRPS domain partners must be sufficiently stable to allow for processing of the relevant reaction intermediate. On the other hand, overly stable interaction with one partner would lead to a situation in which a T-

domain is locked in interaction with this domain at the expense of interaction with others. This chapter focused primarily on interaction between the T and TE-domains of BpsA, because this enzyme does not have a C-domain, and because no structural models for an A-domain interrupted by an Ox-domain presently exist. Furthermore, the bioinformatic analysis presented in Appendix 2 suggests that T-domain interactions with a downstream partner are more specialised than interactions with an upstream partner; and for BpsA, the only downstream interaction partner is the TE-domain. Based on the results of the directed evolution and structural modelling studies described in this chapter, as well as existing biochemical data, a model for interaction between the T and TE-domain of BpsA is now proposed.

In the structure 2roq solved for EntF, the interface between the T and TE-domains is comprised of buried hydrophobic side chains from both domains. Structural models suggest that the same situation exists in BpsA. Data from the directed evolution experiments implicates position +24 as a key residue in this interface. Mutagenesis of EntF has also indicated that this position plays a key role in T-TE interaction [93]. The T-domain appears to adopt a unique conformation, the TE-state, which allows access to the active site of the TE-domain. This conformation requires that side chains present at the core of the T-domain in the H and A/H states be shifted so as to contribute to a shared hydrophobic core between the T and TE-domains. In BpsA and EntF, a phe residue located at -6 is at the centre of this interface, and is closely associated with hydrophobic side chains extending from loops at the end of the T-domain and the start of the TE-domain. Figure 6.17 gives a schematic representation of this, for an illustration of the specific residues involved in this interaction in EntF and BpsA refer to Figures 6.4A and 6.15B respectively. This interaction is likely to depend not only on residues directly involved in the interface between the two domains, but also on residues that contribute to stability of the T-domain in the TE-state.

Position +4 was also implicated as being important for correct interaction between the T and TE-domains of BpsA. In the structure solved for EntF, this position is occupied by a met, whose sidechain extends toward helices 4 and 5 of the TE-domain. These helices are the only portion of the TE-domain that overlaps the T-domain in the structure 2roqA, and the loop

connecting them contains a ser and thr residue, the side chains of which extend toward that of met +4 in the T-domain. As illustrated in Figure 6.14, helices 4 and 5 of the TE-domain are particularly mobile and the presence of a hydrophobic residue at +4 might promote this mobility or prevent overly stable electrostatic interactions locking the T-domain in interaction with the TE-domain. Unfortunately, it was not possible to model the region corresponding to helices 4 and 5 of the BpsA TE-domain, due to low sequence identity with solved TE-domain structures. Nonetheless, it seems likely that +4 is an important position for interaction between the T and TE-domains of BpsA, although the nature of its role is unclear. As discussed in chapter 5 this position also appears to be important in correct interaction between T-domains and downstream C-domains. It is possible that residues at this position determine correct interaction between a T-domain and a downstream partner by different mechanisms in different contexts.

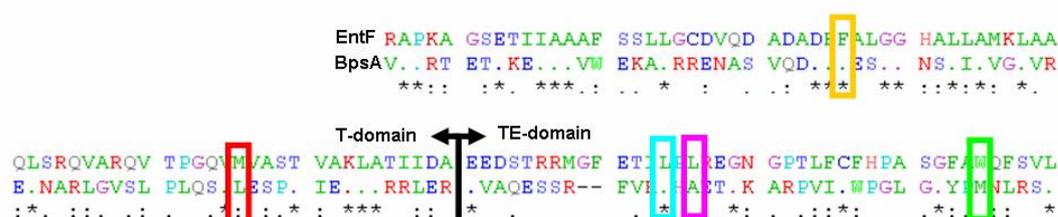
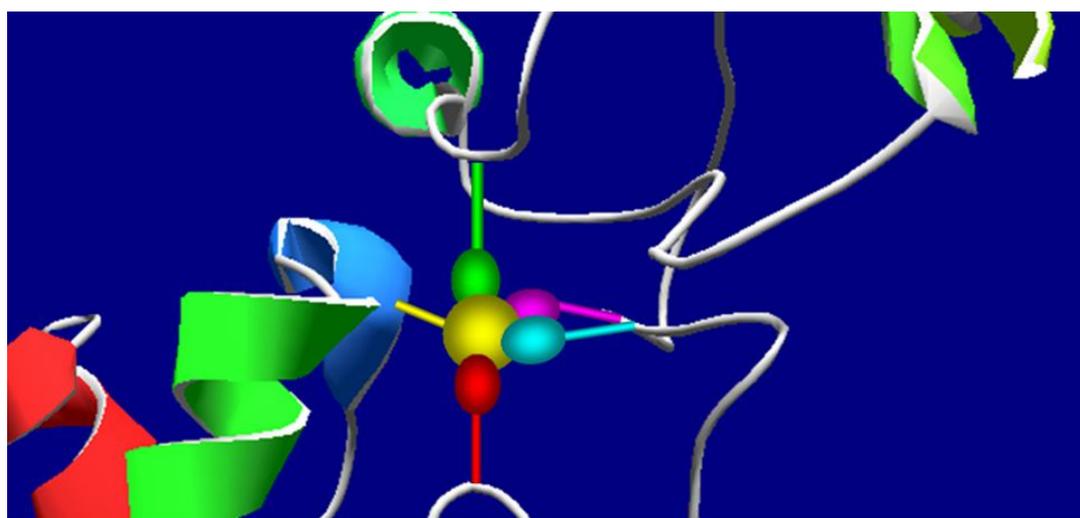


Figure 6.17 – Schematic representation of the hydrophobic interface between T and TE-domains

Residues contributing to the proposed domain interface are represented by spheres/ovals. These are colour coded to allow identification of the residues in an alignment of the EntF and BpsA T-TE interface sequences, shown in the bottom panel.

6.12.6 – Concluding comments on T-domain function and interactions

The multitude of NRPS biosynthetic systems in nature have undoubtedly arisen from a common set of ancestral proteins, which have diversified through module duplication and horizontal gene transfer events. The evolutionary history of NRPS enzymes has therefore presumably required a single ancestral T-domain protein to diversify in response to many functional niches; the result of this is a diverse set of information rich sequences sharing a common structural motif [92-94, 193, 199-201]. Elucidation of T-domain structures both in isolation, and in combination with other domains, has revealed a high degree of structural plasticity [44, 53, 192] and there is evidence that conformational switching is central to modulating interactions with other domains and proteins [3, 11, 44, 64, 192]. Bearing this in mind, a great deal of the information encoded in diverse T-domain sequences may govern these conformational dynamics. It is likely that the A and A/H structures that have been elucidated for T-domains in isolation represent stable base conformations from which other, less stable interaction conformations are adopted. There is structural evidence for the presence of such interaction conformations as already discussed in 6.2.2, and data from the domain swapping and directed evolution experiments described here, as well as directed evolution experiments previously conducted on the ArCP of EntB [92], show that there are a diverse set of sequence options that can allow a T-domain to function correctly in a particular setting. Perhaps what is most important for T-domain function is the balance between stability of various conformations, a balance which can be achieved in many different ways.

Chapter 7 – BpsA as a tool for the characterisation and discovery of phosphopantetheinyl transferase enzymes

7.1 – Introduction

As described in chapter 6, the unique ability to synthesise a readily detectable pigment in *E. coli* makes the single-module NRPS BpsA a powerful tool for probing NRPS domain interactions. In order for pigment synthesis to occur, BpsA must first be activated by post-translational attachment of a 4'-PP group to the active site serine of its T-domain. This activation is catalysed by a cognate PPTase enzyme, which is able to recognise specific sequence elements in the T-domain of BpsA. Since the native PPTase enzymes of *E. coli* are unable to fulfil this role, co-expression of an activating PPTase enzyme is necessary for BpsA-mediated pigment synthesis to occur. The dependence on PPTase-mediated activation allowed preliminary development of new assays for characterisation of these enzymes, comparing the efficiency of different PPTases both *in vivo* and *in vitro*. BpsA was also used as a reporter for the discovery of new PPTase enzymes and secondary metabolite clusters from an environmental DNA library.

7.1.2 – Aims and overview

The aim of this work was to develop BpsA based tools for the characterisation and discovery of PPTase enzymes. A test was developed using WT and modified BpsA enzymes to assess PPTase specificity *in vivo*. This provided a useful qualitative assay but did not allow accurate quantification of efficiency of T-domain activation by a particular PPTase possibly due to variation in expression levels *in vivo*. An *in vitro* assay was subsequently developed in which the rate of pigment synthesis by BpsA was monitored as an indirect measure of T-domain activation by a given PPTase. This *in vivo* assay was found to be quantitative, allowing relative rates of T-domain activation to be assigned for the three PPTases examined. Detailed kinetic analyses were also conducted for two of these PPTases. Finally, a protocol was developed that allowed high throughput screening of environmental DNA for the presence of

novel PPTase enzymes. Using this method, three previously uncharacterised PPTase enzymes capable of activating BpsA were recovered from an environmental DNA library. Analysis of one of the sequences recovered from this screen also revealed characteristic features of a secondary metabolite cluster, suggesting that the screen may also be applicable for the discovery of novel secondary metabolite clusters from uncultured organisms.

7.2 – *In vivo* characterisation of PPTase enzymes using BpsA

In order to ascertain whether BpsA could be used to assay PPTase specificity *in vivo*, plasmid-encoded copies of the WT BpsA enzyme as well as the active T-domain variants oDhbF, oPsT, sEntF were transformed into three different PPTase expression backgrounds. The construction of hybrid genes and PPTase over-expression strains is described in detail in section 6.5. The aim was to compare the ability of three different PPTases to recognise four different T-domain substrates. The PPTase enzymes examined were: Sfp from *B. subtilis*, PcpS from *P. aeruginosa* and PP1183 from *Pseudomonas putida*. All of these enzymes have previously been reported as having broad substrate specificity [40, 69, 202].

Once each test strain had been created, a master 96 well plate containing three separate glycerol stocks for each strain was established allowing rapid inoculation of triplicate overnight cultures for each strain using a 96 pin inoculating tool. The following day, assay cultures were inoculated to uniform density in pigment development medium in a 96 well plate, which was then incubated for 12-72 hrs at 18 °C. Following growth, pigment levels in culture supernatant were assessed by measuring absorbance at 590 nm in a microplate reader. Although qualitative, in that it was possible to determine whether a particular PPTase was capable of activating a particular T-domain as evidenced by presence or absence of pigmentation, quantitation of relative efficiencies proved impossible due to variation within and between assay repeats. The assay is nonetheless a valuable tool for initial assessment of substrate specificity of PPTase enzymes. It was found that each of the PPTases tested was capable of recognising the T-domains of all BpsA reporters. In the case of the reporter sEntF, which contained the T-domain from the *E. coli* NRPS EntF, pigmentation was observed even

in the absence of an exogenous activating PPTase due to this T-domain being a substrate for the native *E. coli* PPTase EntD [42]. When WT BpsA was used as a reporter, pigmentation was also observed in the absence of an exogenous activating PPTase, although this was weak and only became apparent after 48-72 h incubation. In contrast, co-expression of Sfp, PcpS or PP1183 with WT BpsA resulted in strong pigmentation appearing after 60-120 min.

7.3 – *In vitro* activity of BpsA

The first step toward developing BpsA as an *in vitro* reporter for PPTases was to establish reliable protocols for protein purification and subsequent assessment of activity. An *in vitro* protocol for measuring the rate of pigment formation was successfully established, and used to derive kinetic parameters for BpsA.

7.3.1 – Purification of his-tagged BpsA and PPTases

For purification of BpsA, expression was conducted using auto-induction medium at 16 °C as described in 2.9.1.5. Owing to the FMN group bound to the Ox domain, BpsA enzyme preparations have a characteristic bright yellow colouration [80], a feature which assisted the monitoring of enzyme solubility and purification. Following lysis and separation, the soluble fraction was discernibly yellow, indicating good yield of soluble protein. This indication was confirmed by subsequent SDS PAGE analysis which revealed a large band corresponding to BpsA in the soluble fraction and very little BpsA in the insoluble fraction (not shown).

In an initial purification attempt using 200 mL of induced culture it was found that it was difficult to apply sufficiently stringent washing conditions to separate BpsA from major contaminating species without losing the BpsA band also. For subsequent attempts a 2000 mL culture was used as described in 2.9.2.4; this gave soluble protein in vast excess to the binding capacity of the column used. This excess of target protein resulted in a much lower level of contaminants, with the final protein appearing sufficiently pure for subsequent biochemical analyses (Figure 7.1).

Three broad specificity PPTase enzymes were also purified using Ni-NTA affinity chromatography (Figure 7.1). These were *B. subtilis* Sfp, *P. aeruginosa* PcpS, and *P. putida* PP1183. Expression of PPTase enzymes was conducted at 16 °C using auto-induction medium as described in 2.9.1.5 with subsequent purification as described in 2.9.1.

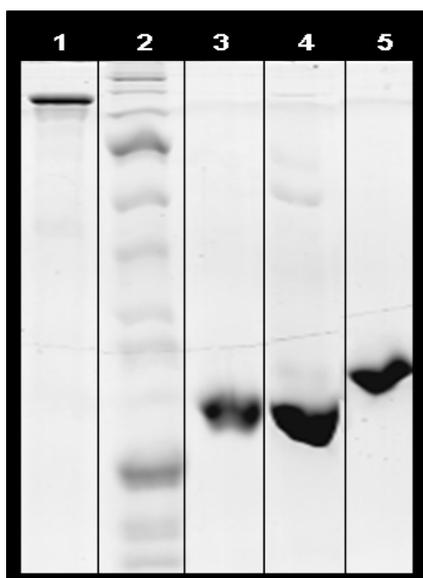


Figure 7.1 – SDS Page analysis of purified BpsA and PPTases

Purified, buffer exchanged enzymes are shown. 1-BpsA. 2- MW marker. 3-PcpS . 4- PP1183 . 5- Sfp .

7.3.2 – Initial *in vitro* characterisation of BpsA

For *in vitro* analysis of indigoidine synthesis, purified BpsA was first incubated with purified PP1183 and appropriate substrates to bring about 4'-PP attachment as described in 2.13.1. Indigoidine synthesis reactions were then set up in a 96 well plate as described in 2.13.2. Reactions were initiated by addition of activated enzyme solution followed by continuous measurement of A_{590} in a Versamax™ microplate reader. As shown in Figure 7.2, reactions exhibited a decrease in absorbance after the maximum value was reached. Interestingly, the rate of decrease in product concentration after the peak value was proportional to the rate of increase before. The reason for this phenomenon was investigated in more detail.

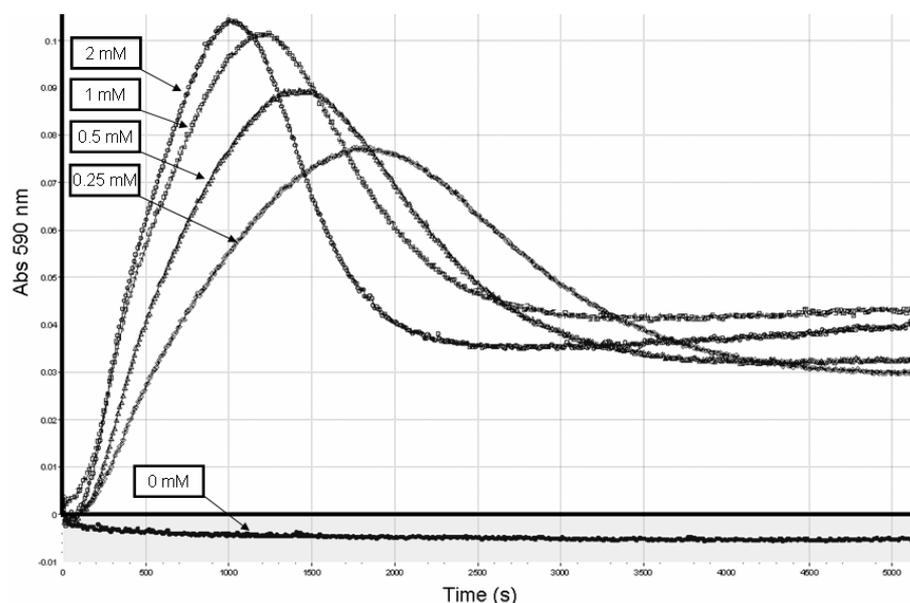


Figure 7.2 – Preliminary analysis of indigoidine formation catalysed by BpsA

Reactions were set up in a 96 well plate as described in 2.13.2 and initiated by addition of purified BpsA. Product concentration was then monitored by measuring A_{590} every 10s in a microplate reader. The final concentration of L-gln present in each reaction is indicated, the concentration of all other reactants was constant between reactions. Notice that absorbance decreases after a peak value is reached, and that this decrease is proportional to the previous rate of increase. A similar pattern was seen when using increasing concentrations of ATP at a set concentration of L-gln.

It was noted that the supernatant from BpsA-expressing cultures reached a much higher A_{590} (~1.2) than the maximum achieved from any *in vitro* experiments (~ 0.3) (not shown). It was reasoned that since indigoidine produced and exported from a cell *in vivo* would be effectively removed from exposure to intracellular enzymes, the loss of product seen *in vitro* might be due a second reaction catalysed by an enzyme present in the *in vitro* assay preparations. This could have been BpsA itself, the activating PPTase or an undetected contaminating species. To test this possibility, filter sterilised culture supernatant containing indigoidine was incubated with various concentrations of each enzyme preparation used for *in vitro* assays. However, no difference in pigment levels resulted from this treatment and it was concluded that the loss of pigmentation seen *in vitro* was not enzyme catalysed. Each of the other assay reagents was subjected to similar analysis and it was found that only addition of sodium phosphate buffer (pH 7.8) to a final concentration of 50 mM resulted in decolouration of culture supernatant. A more detailed analysis of the relationship between pH and pigment stability was then conducted. The results of this analysis, illustrated in Figure 7.3A, showed

that decolouration occurred to a greater extent at higher pH. This finding was consistent with previous observations that indigoidine is unstable at alkaline pH [206]. Although the optimal pH for BpsA function is reported to be 7.8 [80], the sensitivity of this enzyme to changes in solution pH was now tested in the hope that it might be possible to use a lower pH buffer. As illustrated in Figure 7.3B, BpsA activity was found to be very pH sensitive, with deviations of 0.2 units either side of 7.8 resulting in a severe reduction of activity. Based on the results described, it was concluded that the loss of pigmentation observed was due to non-enzymatic degradation of indigoidine by an unknown mechanism and that this degradation was unavoidable at the optimal pH for enzyme function.

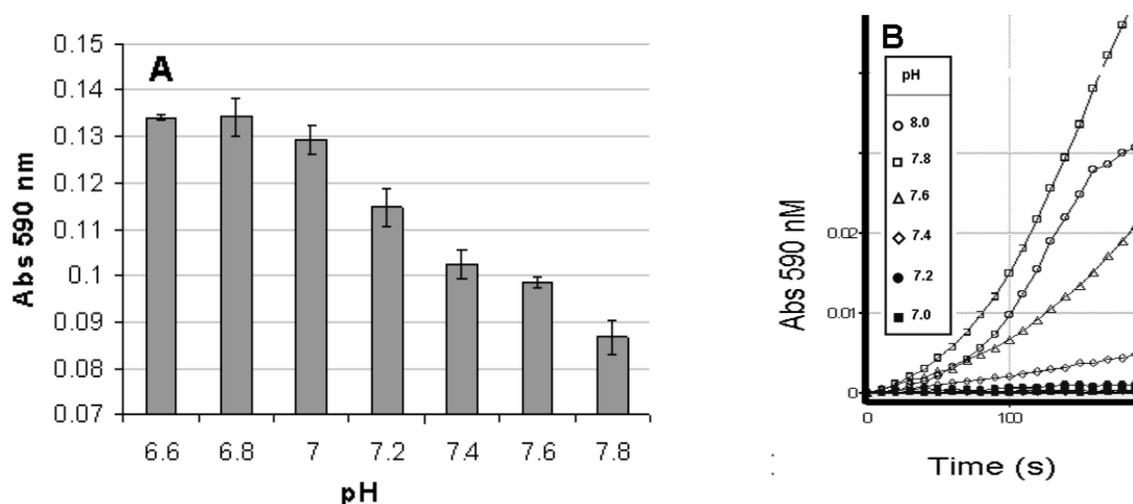


Figure 7.3 – The effect of pH on Indigoidine stability and BpsA reaction rate

A- Diluted sterile supernatant from an indigoidine producing culture was incubated for 20 min with 50 mM sodium phosphate buffer at each of the indicated pH values in triplicate. Loss of colouration was then assessed by measuring A_{590} . Error bars are presented as +/- one standard deviation.

B- Indigoidine reactions differing only in the pH of the buffer used are shown. Deviation from pH 7.8 by as little as 0.2 units was found to noticeably impair function.

7.3.3 – Kinetic studies

For the derivation of kinetic parameters for BpsA, triplicate reactions in a 96 well plate were set up across a two fold dilution series of ATP as described in 2.13.2. The maximum velocity (V_0) of product formation for each substrate concentration was derived from the five data points that gave the highest linear gradient, as illustrated in 7.4. Gradients were then converted

into rates of product formation ($\mu\text{mol}/\text{min}$) using the reported extinction coefficient of indigoidine (A_{590} of 0.1 = 8.54 μM indigoidine for 0.6 cm path length used in expt.) [206]. The resulting data was analysed using SigmaPlot® allowing derivation of K_m and V_{max} values by the one site saturation regression function (assessing fit of data to the Michaelis-Menten equation). Figure 7.5 shows plots of V_0 vs. substrate concentration for ATP. Table 7.1 gives a summary of the kinetic parameters derived from these experiments.

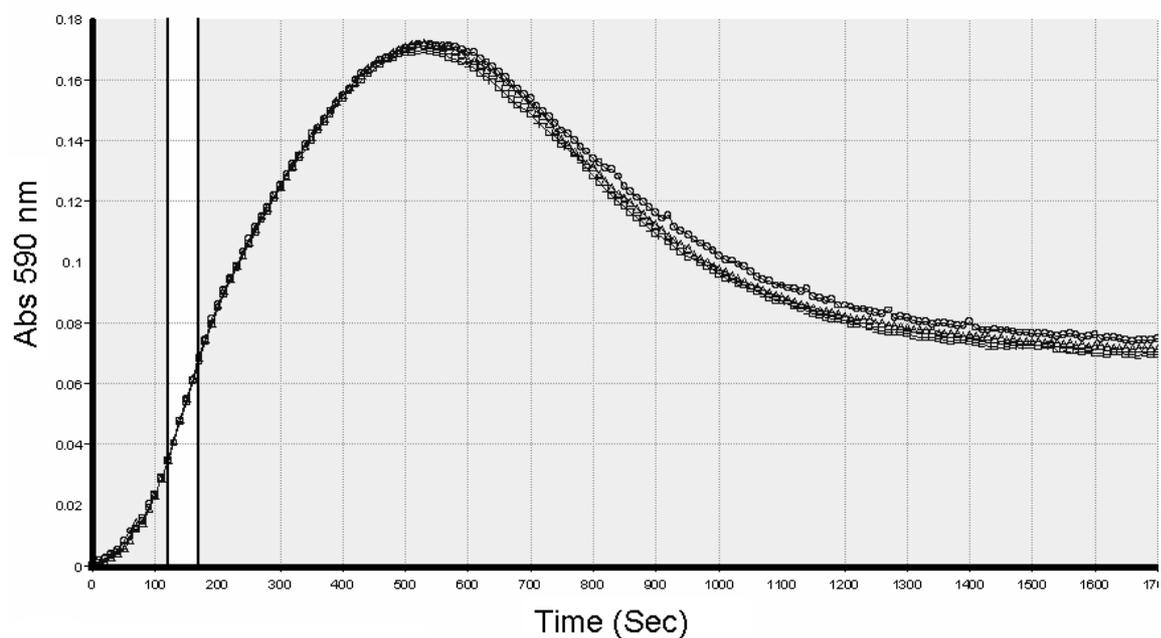


Figure 7.4 – Determination of pigment synthesis velocity

In order to determine maximum velocity for a particular substrate concentration, the six consecutive data points (50 sec) which gave the steepest gradient were determined. Illustrated here are triplicate reactions for a single concentration of ATP, the vertical black bars indicate the region which was used for maximum velocity derivation. This strategy was used to determine maximum velocity for each concentration of substrate for subsequent determination of kinetic parameters.

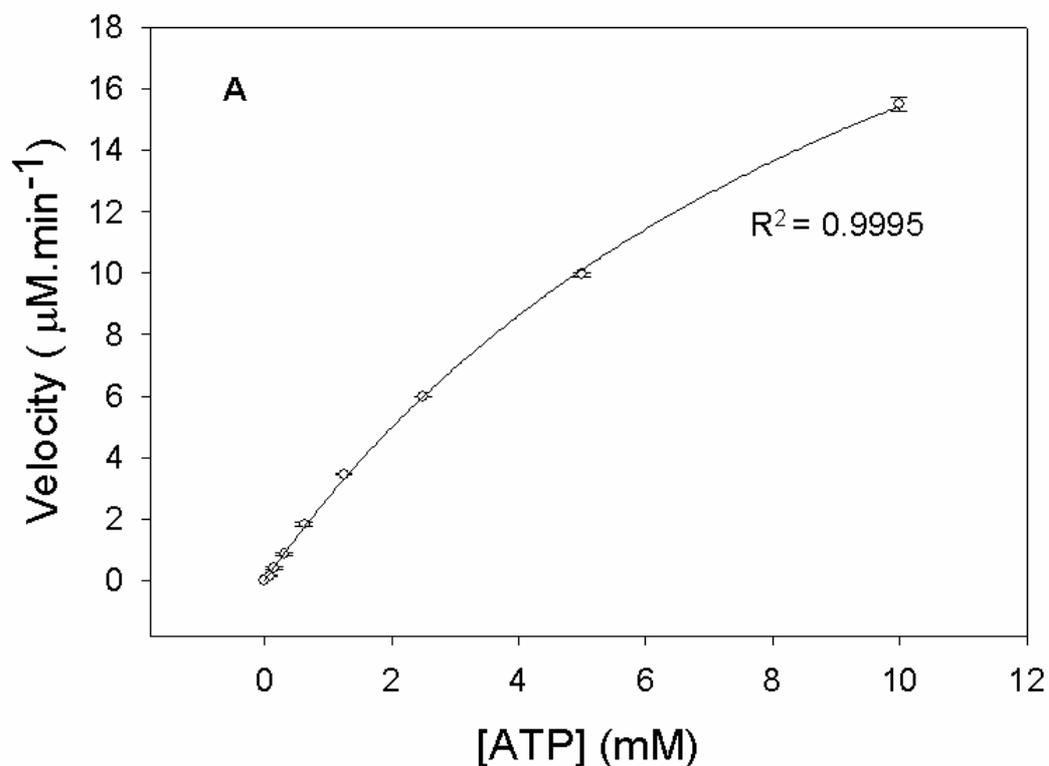


Figure 7.5 – Derivation of kinetic parameters for BpsA

Maximum velocity of indigoidine formation for a serial dilution series from 10 to 0.078 mM ATP. Curve was fitted using the Sigmaplot one site saturation regression function and used to derive kinetic parameters for BpsA with this substrate. Data points were determined from three independent repeats, and error bars are presented as +/- one standard deviation.

Table 7.1 – Kinetic parameters* for BpsA

K_m (mM)	11.26 +/- 0.3
V_{max} (nmol.min ⁻¹ .mg ⁻¹)	2791 +/- 54
k_{cat} (min ⁻¹)	241.3 +/- 4.4
k_{cat}/K_m (min ⁻¹ .mM ⁻¹)	21.62 +/- 1.02

*0.078 - 10 mM ATP with [L-gln] fixed at 2 mM,

7.4 – Determination of kinetic parameters for PPTases using a BpsA reporter system

Having established a reliable protocol for determination of pigment synthesis rates and an estimate of k_{cat} and K_m for BpsA, it was now possible to use this enzyme to quantify PPTase function. Reactions containing BpsA purified from *E. coli*, an activating PPTase, and the necessary substrates for both the 4'-PP attachment and pigment synthesis reactions were established as described in 2.13.3. Using this assay it was possible to determine kinetic parameters for PPTase enzymes by monitoring the acceleration of the indigoidine synthesis reaction of BpsA (Figure 7.6). The derivation of kinetic parameters is based on the assumption that the rate of indigoidine synthesis is directly proportional to the concentration of activated BpsA, therefore the change in rate of indigoidine synthesis is directly proportional to the rate of BpsA activation by the PPTase included in the reaction.

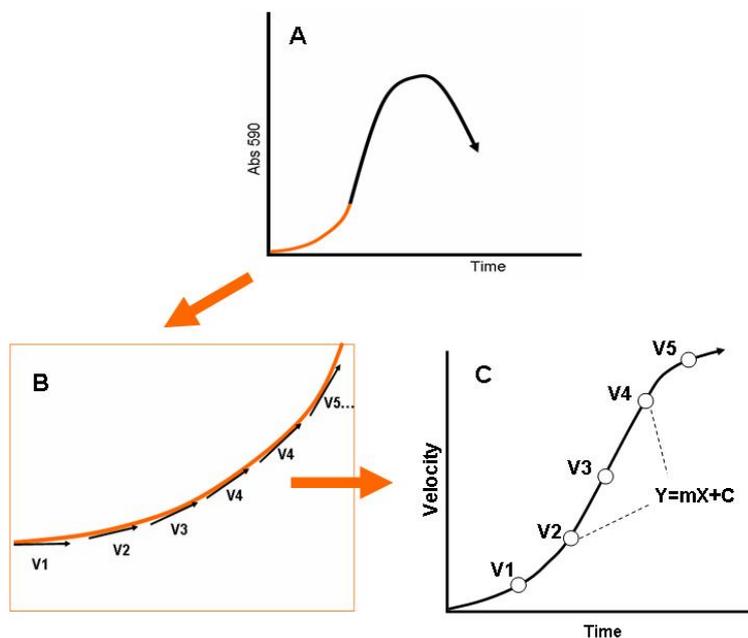


Figure 7.6 – Derivation of kinetic parameters for PPTase enzymes using the BpsA reporter assay

A) Absorbance at 590 nm for reactions containing inactive BpsA and an activating PPTase was measured every 10 s. **B)** The average gradient between every three points of the curve (20 s) was measured to provide an estimate of instantaneous velocity. **C)** Instantaneous velocity vs. time is graphed, and the steepest portion of the curve determined. This gradient represented the maximum rate of BpsA activation for an individual PPTase at a particular concentration of CoA. Rates of BpsA modification were determined across a range of substrate (CoA) concentrations; these were then used to derive kinetic parameters for the activating PPTase.

Three PPTases were assessed using the BpsA reporter assay and kinetic parameters derived for two of these. Reactions for each PPTase were set up in duplicate across a two fold serial dilution series of CoA as described in 2.13.3 and data points (A_{590}) were measured every 10 s. Figure 7.7 shows a sample of the raw data from this experiment and the gradient values that were derived from one of the reactions.

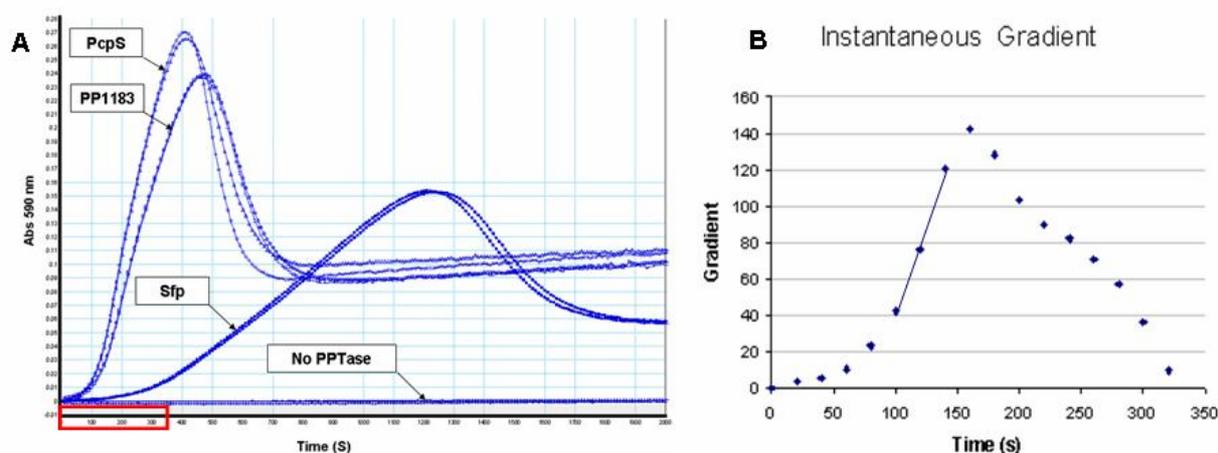


Figure 7.7 – Raw and processed data from PPTase comparison experiment

A) Data from duplicate reactions for the three PPTases and a no PPTase control for a single concentration of CoA. Measurements were taken at six different CoA concentrations (2 fold dilution series 25.0-0.78 μM). The red box indicates the region from which the gradient values shown in panel B were taken.

B) A plot of instantaneous gradient for one of the PcpS reactions shown in panel A. The points connected by a straight line were used to derive a value for the maximum velocity of PcpS modification of BpsA at this concentration of CoA. In all cases a minimum of three consecutive data points was used for determination of PPTase velocity for a particular reaction.

Using the process outlined, duplicate values for maximum velocity of BpsA activation were obtained at final concentrations of 25.0, 12.5, 6.25, 3.13, 1.56 and 0.78 μM CoA. The resulting set of data was used to derive kinetic parameters for both PPTase enzymes using the one site saturation regression analysis function of SigmaPlot. As shown in Figure 7.8 the derived data gave a very good fit with the Michaelis-Menten equation.

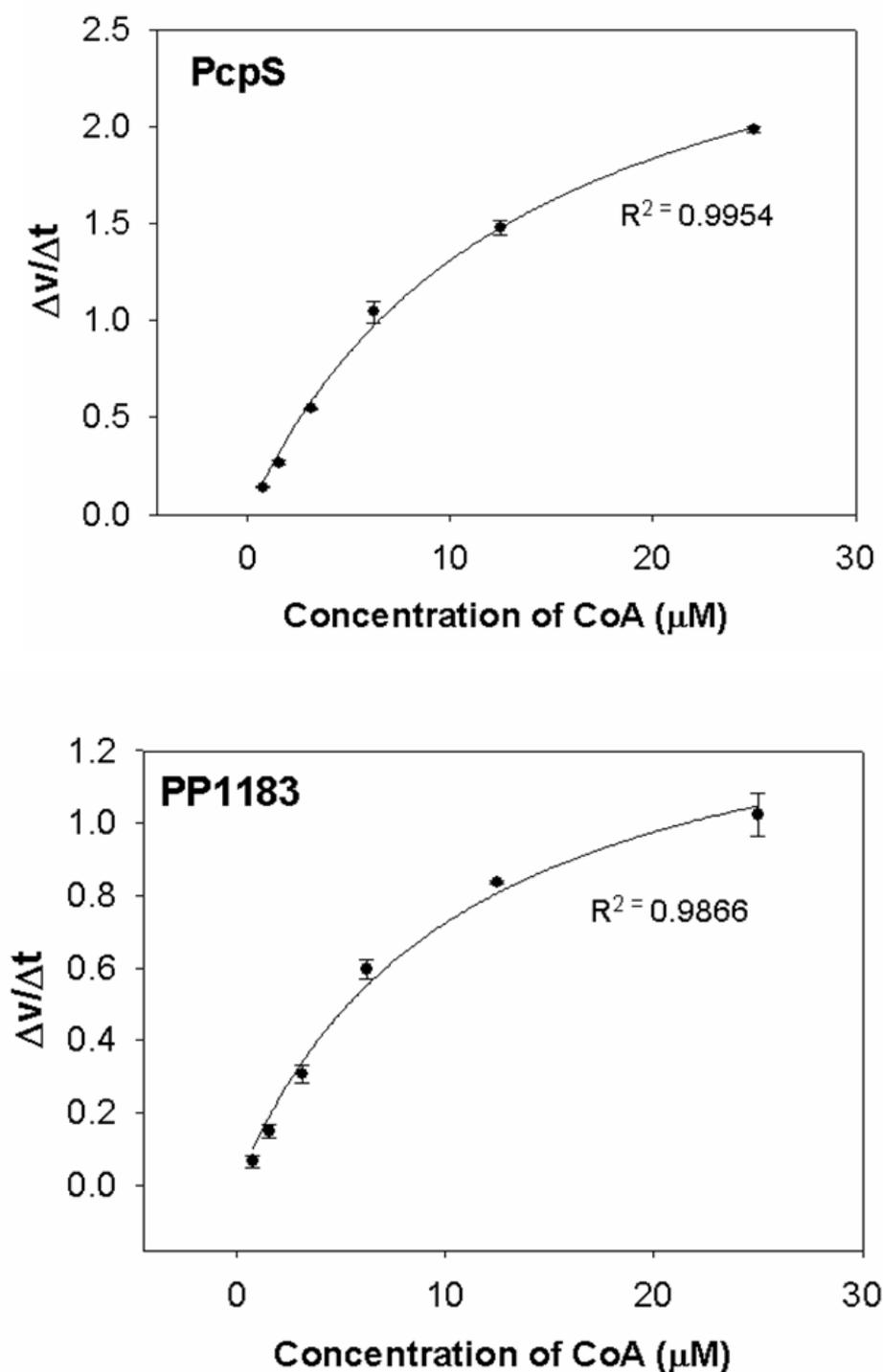


Figure 7.8 – Derivation of kinetic parameters for PcpS and PP1183

Plots of maximum acceleration in rate of pigment synthesis vs. concentration of CoA are shown. Maximum acceleration of pigment synthesis is proportional to maximum rate of BpsA modification by the PPTase being assessed. Curve was generated using the SigmaPlot one site saturation regression function and used to derive kinetic parameters for PPTase modification of BpsA. Data points were determined for duplicate independent reactions, and error bars are presented as +/- one standard deviation.

The Y axis units for the data shown in Figure 7.8 are change in reaction velocity per second. In order to convert these units into a rate of modification of BpsA the assumption was made that reaction rate was directly proportional to the amount of BpsA in a reaction that had been activated by a PPTase during the time period from which the data was derived. This is a reasonable assumption given that BpsA fits the Michaelis-Menten model (Figure 7.5) and for all acceleration values obtained, the relationship between instantaneous gradient and time was linear (Figure 7.7B). The previously determined kinetic parameters for BpsA were then used to derive a theoretical maximum velocity for pigment synthesis by that amount of BpsA under the conditions of the assay (i.e. the theoretical rate of pigment synthesis if all BpsA has been activated by a PPTase). Having established a relationship between pigment synthesis rate and amount of active BpsA present, the V_{\max} values obtained could be converted from change in gradient to amount of BpsA modified per second, which is the same as amount of CoA consumed per second. The kinetic parameters determined for PcpS and PP1183 are summarised in table 7.2. A more detailed explanation of the conversion of change in gradient values to rate of activation values is given in appendix 2.

Table 7.2 – Kinetic parameters for PcpS and PP1183

Parameter*	PP1183 value	PcpS value
K_m (μM)	10.7 +/- 1.4	13.4 +/- 1.13
V_{\max} ($\text{nmol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	27.4 +/- 1.7	56.1 +/- 2.3
k_{cat} (min^{-1})	0.733 +/- 0.045	1.54 +/- 0.060
k_{cat}/K_m ($\text{min}^{-1}\cdot\text{mM}^{-1}$)	68.7 +/- 13.2	115.5 +/- 14.2

*0.78 – 25.0 μM CoA with [BpsA] fixed at 1.64 μM

7.5 – BpsA as a tool for discovery of PPTase genes in environmental DNA libraries

Since BpsA depends on activation by an exogenous PPTase for function in *E. coli*, it has the potential to serve as a reporter for the discovery of PPTases in eDNA libraries. A protocol for recovery of PPTases was developed and used to recover novel PPTase genes from a small insert, plasmid based eDNA library.

7.5.1 – Screening of an eDNA library and recovery of hits

For screening purposes, an *E. coli* strain harbouring a plasmid-encoded copy of the *bpsA* gene was used as a reporter. The eDNA library used for these experiments was kindly provided by Nadia Skorupa Parachin of Lund University. This library contained small insert genomic DNA fragments in the plasmid pRSET1. The genomic DNA was isolated from a soil sample and prepared so that the size inserts would be between 500 and 3000 bp. Library DNA was delivered to reporter cells by electroporation and approximately 1×10^6 of the resulting clones were assessed for colour development on pigment production agar as described in 2.14. Nine of these clones developed colouration and the plasmid-encoded eDNA fragment from these was isolated and sequenced. Figure 7.9 shows a representative screening plate on which a hit was recovered.

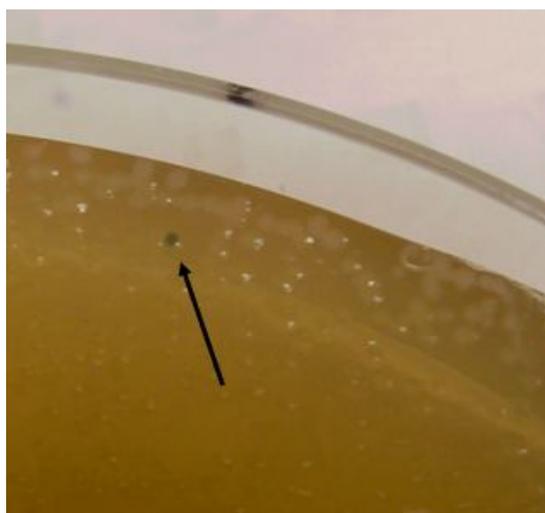


Figure 7.9 – Recovery of PPTases from an eDNA library

Pigmentation arising from PPTase mediated activation of BpsA allowed rapid and unambiguous identification of PPTase harbouring clones. A plate on which such a clone was recovered is pictured.

7.5.2 – Sequence analysis of hits from eDNA library screen

Plasmid DNA was isolated from each of the nine clones recovered from the screen and the nucleotide sequence of the eDNA inserts analysed. It was found that three unique sequences had been recovered multiple times. Each of the three unique nucleotide sequences was translated in all six reading frames and the resulting aa sequences analysed by BLAST search against the non-redundant protein sequence collection of NCBI. This enabled the identification of entire PPTase genes for all three fragments, as well as additional partial and complete genes isolated by virtue of proximity to the PPTase genes. The organisation and identity of partial and complete genes identified in each eDNA fragment is illustrated in Figure 7.10. Of particular interest were two partial genes recovered in one of the eDNA fragments which indicated that of part of a NRPS/PKS secondary metabolite cluster had been recovered.

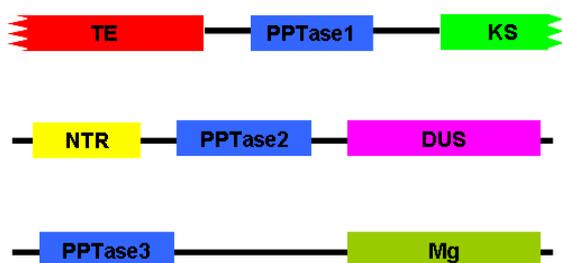


Figure 7.10 – Genomic DNA fragments isolated from eDNA library screen.

Arrangement and identity of partial and complete genes found in the three unique eDNA fragments isolated. Refer to table 7.3 for a description of the genes. PPTase1 was isolated along with partial sequences indicative of a secondary metabolite cluster.

Table 7.3 – partial and complete genes found in eDNA fragments

Gene	Top match from BLAST search	Species in which top match is found	Percent identity
PPTase1	Sfp type PPTase	<i>Nostoc punctiforme</i> PCC 73102	43 %
PPTase2	Sfp type PPTase	<i>Streptomyces sviveus</i> ATCC 29083	35 %
PPTase3	Sfp Type PPTase	<i>Burkholderia oklahomensis</i> C6786	40 %
TE	Type II thioesterase – partial	<i>Planktothrix agardhii</i>	47 %
KS	PKS KS-domain – partial	<i>Cyanothece sp.</i> CCY0110	52 %
NTR	Nitroreductase	<i>Nitrococcus mobilis</i> Nb-231	50%
DUS	Dihydrouridine synthase	<i>Nocardia farcinica</i> IFM 10152	55 %
Mg	Magnesium transporter	<i>Xanthomonas campestris</i>	55%

7.6 – Discussion

The results outlined in this chapter establish the BpsA protein as a powerful tool for characterisation and discovery of PPTase enzymes. The use of BpsA and modified variants to probe PPTase specificity *in vivo* allows rapid determination of the compatibility of a T-domain-PPTase combination by a simple co-transformation experiment, whereas the *in vitro* activity assay allows rapid and accurate assessment of kinetic parameters for PPTase enzymes. Finally, the eDNA mining procedure allows unambiguous identification of novel PPTase enzymes with a minimum of manual input. The validity of the various BpsA assays and the possible applications are now discussed.

7.6.1 – Interpretation of BpsA kinetic data

BpsA presents a unique opportunity to study the kinetics of an entire NRPS module using a simple colourimetric assay. The final version of the protocol as described in this chapter and 2.13.2 allowed triplicate maximum velocity values for eight substrate concentrations to be determined simultaneously in a single 96 well plate. The standard deviation between replicates was very small and the final data set fit very well with the Michaelis-Menten equation (Figure 7.5, $R^2 = 0.9995$). Formation of indigoidine requires a number of enzymatic steps catalysed by BpsA (Figure 6.2). The first of these steps is activation of L-gln by the A-domain, with ATP being a second substrate in this reaction. The K_m value derived for ATP therefore pertains specifically to the A-domain of BpsA at a fixed concentration of L-gln. However, activation of a substrate by an A-domain is not the rate limiting step in NRPS directed biosynthesis, with subsequent transfer to the T-domain believed to be slower than activation [30, 169, 207]. The V_{max} derived for BpsA therefore pertains to the maximum possible rate of indigoidine synthesis by the entire enzymatic complex, dictated by whichever step in the process is rate limiting. The derivation of kinetic parameters for BpsA was an important enabling step in the subsequent development of a PPTase kinetic assay, as they allowed a theoretical maximum velocity value to be assigned for fully activated BpsA under the conditions used in this assay. This value is the theoretical maximum velocity of the pigment synthesis reaction, if all BpsA present had been activated by 4'-PP attachment. Assuming that pigment synthesis rate is

directly proportional to amount of active BpsA in a reaction, this allows the conversion of change in gradient values to rates of modification of BpsA by a PPTase as outlined in appendix 2.

7.6.2 – Interpretation of PPTase kinetic data and assay validation

The kinetic analysis revealed that PcpS is a more efficient PPTase for the modification of BpsA than PP1183. The data derived for PPTases using the BpsA assay system fits well with the Michaelis-Menten equation (Figure 7.8), indicating that the technique provides an accurate assessment of PPTase reaction rate. Further validation of the assay comes from comparing the data derived from this experiment to that from alternative forms of PPTase assay. PcpS has been previously characterised using a standard HPLC based protocol for determination of modification rate. A K_m for CoA was not determined in these experiments; however a k_{cat} of $3.0 \pm 0.1 \text{ min}^{-1}$ was determined for PcpS with an NRPS T-domain substrate [40]. This value is in good agreement with the value of $2.08 \pm 0.08 \text{ min}^{-1}$ obtained using the BpsA assay.

7.6.2 – Utility of BpsA for recovery of novel PPTases and secondary metabolite clusters.

BpsA proved to be an effective tool for recovering PPTase enzymes from eDNA samples. Using the protocol developed it was possible to screen and recover hits from 10^6 clones in less than one week. There were no false positives recovered from this screening process, with each of the nine clones recovered containing a PPTase gene upon sequence analysis. There were only three unique sequences represented in these nine clones, each of which was recovered multiple times. A BlastP search using translated gene sequences revealed that none of the three PPTase genes recovered had been previously characterised. Of particular interest was the recovery of one of the PPTase enzymes along with partial NRPS/PKS gene sequences. This is in accordance with previous observations that PPTase genes are commonly associated with NRPS secondary metabolite clusters, and to a lesser extent, with PKS clusters [4, 208]. The insert size of the plasmid-based genomic DNA library used was relatively small (0.5-3.0 kb), limiting the number of associated sequences recovered with PPTase genes. By using BAC,

cosmid or fosmid vectors, soil eDNA libraries with average insert sizes in excess of 40 kb are easily achievable [209]. The creation and screening of such libraries using the BpsA reporter system would be an effective means for recovery of novel NRPS and PKS genes and could even be used for the recovery of entire, functional secondary metabolite clusters. Even if entire clusters were not recovered in a single clone, subsequent primer walking experiments could be used to recover missing cluster regions. In its present format clone numbers in excess of 10^6 can easily be assessed by a single researcher, without automation, in a single week. With further refinement, this screen has the potential to comprehensively assess large eDNA libraries for novel PPTases and associated secondary metabolite clusters.

Chapter 8 – Conclusions

8.1- Research motivation

The study of NRPS enzymes is of immense scientific interest due not only to the unique chemistry catalysed by these enzymes, but also to their integral role in bacterial physiology. NRPS enzymes are comprised of a series of interacting catalytic centres, the individual function of which is well understood. It is becoming increasingly apparent however, that understanding the complex mechanisms of interaction between these centres presents a rich new avenue of scientific investigation. The research described in this thesis investigates the function of NRPS enzymes at a molecular level as well as their natural role in bacteria.

The initial aim of this research was to first characterise the pyoverdine biosynthetic NRPS of *P. syringae*, which were to serve as raw material for subsequent domain swapping experiments. This afforded the opportunity to investigate not only the mechanism of pyoverdine synthesis, but also the physiological significance of the siderophores of *P. syringae*. As part of this characterisation, determination of the substrate specificity of seven A-domains was attempted. This proved to be a challenging exercise, requiring systematic investigation of protein expression, purification and characterisation techniques.

Two separate but related aspects of NRPS domain interaction were then investigated. The first of these was specificity of the C-domain for the donor substrate of a condensation reaction. The second was interaction between the T-domain and other NRPS domains, with particular focus on T-TE interaction. Understanding both of these phenomena is an important prerequisite for developing NRPS enzymes as a robust platform for the synthesis of novel bioactive compounds.

8.2 – Key findings

8.2.1 – Biosynthesis and physiological significance of siderophores in *P. syringae*

Ps1448a has the genetic capacity to produce three different siderophores however only two of these, pyoverdine and achromobactin, were detectable as active siderophores under the various conditions examined. Pyoverdine, synthesised in part by NRPS enzymes, has a higher affinity for iron than achromobactin and is more efficient at scavenging iron from the environment. Analysis of the Ps1448a genome resulted in identification of five candidates for pyoverdine biosynthetic NRPS genes. The role of each of these in pyoverdine synthesis was confirmed by gene deletion. For all five genes, deletion completely abrogated pyoverdine production, as evidenced by loss of fluorescence and increased sensitivity to iron starvation.

The specificity of the seven A-domains directing aa incorporation into the pyoverdine peptide sidechain could not be determined biochemically owing to an inability to purify active recombinant enzyme. In order to assign A-domain specificity a combination of *in silico* analysis of the pyoverdine locus and MS analysis of the purified molecule was used. The results from these two analyses were congruent allowing confident allocation of specificity for each A-domain. Interestingly, MS analysis revealed that Ps1448a produces two pyoverdine isoforms differing in mass by 71 units. Fragmentation analysis of these species suggested that the heavier isoform was an as yet uncharacterised pyoverdine variant which may contain an extra alanine residue in the peptide side chain, located between the chromophore and the lysine residue of the molecule.

Achromobactin is synthesised by an NRPS-independent pathway. MS analysis of the purified siderophore, coupled with *in silico* analysis of its biosynthetic locus, suggests that Ps1448a synthesises achromobactin of the same structure and by the same mechanism as in *D. dadantii* [134]. Production of achromobactin in *Ps1448a* is temperature regulated, with growth at 22 °C resulting in higher levels of achromobactin than growth at 28 °C. Deletion of the achromobactin biosynthetic gene *acsA* did not result in a detectable phenotype under any of

the conditions examined, with pyoverdine able to completely compensate for loss of this secondary siderophore. Deletion of *acsA* from a pyoverdine null background however, resulted in increased sensitivity to iron starvation. Neither siderophore is required for pathogenicity of *Ps1448a* in *Phaseolus vulgaris*.

8.2.3 – Manipulation of pyoverdine biosynthetic NRPS in *P. aeruginosa*

Previous module swapping experiments conducted using the pyoverdine biosynthetic NRPS of *P. aeruginosa* led to the proposal that C-domain specificity for peptide donor substrates in a condensation reaction is a major barrier to NRPS manipulation [85, 143]. The experiments described in chapter five aimed to more rigorously test this hypothesis. Recombinant *pvdD* genes, in which the C and A-domains of the terminal module were substituted for foreign domains were created and assessed for their ability to complement a *pvdD* mutation. Using this strategy, it was found that a C-domain known to process L-fhOrn as a donor substrate in nature, was also able to efficiently process L-thr. This result clearly refuted the initial hypothesis of C-domain donor specificity. A second recombinant gene, in which the terminal A-domain specified ser instead of thr was found to restore fluorescence to a PAO1:: $\Delta pvdD$ indicating that the manipulation had successfully resulted in production of a novel pyoverdine molecule with a new aa installed at the terminal position of the peptide sidechain. Analysis of the pyoverdine produced this strain is presently being conducted as part of the doctoral research of Mark Calcott. If confirmed this will be the first example of residue swapping in pyoverdine using artificially recombined NRPS modules. Finally, a recombinant gene containing a C-domain that processes L-thr as a donor substrate in its native setting was not able to perform the same role when substituted into *pvdD*. This led to the proposal that the inactivity observed in this recombinant NRPS may have stemmed from failure of the substituted C-domain to correctly interact with one or both of the T-domains in PvdD. This hypothesis was supported by bioinformatic analysis of T-domain sequences, which found that the identity of residues at specific locations in T-domains correlated with the type of domain located immediately downstream.

8.2.4 – Substitution and directed evolution of T-domains in BpsA

The interaction between T-domain and partner NRPS domains was more closely examined using a novel reporter system based on the pigment-synthesising NRPS BpsA. This single-module enzyme is capable of synthesising a strong blue coloured pigment in *E. coli* without the need for additional synthetase enzymes, making it a powerful tool for investigation of NRPS function *in vivo*. It was found that substitution of the native T-domain for foreign T-domains resulted in severe or complete disruption of enzyme function, consistent with the idea that T-domains have evolved specifically for optimal interaction with other NRPS domains in their native setting. It was particularly interesting to note that the alternative T-domains that yielded recombinant BpsA enzymes with the highest activity were taken from NRPS modules that are also required to interact with a TE-domain immediately downstream in their native setting. This result suggests that T-domain communication with a TE-domain is particularly specialised, and that there are particular features of T-domains immediately adjacent to a TE-domain that are unique, facilitating interaction between the two domains. This observation was supported by previous biochemical data [92-94] and the bioinformatic analysis described above.

A directed evolution approach was then used to improve *in vivo* function of recombinant *bpsA* genes by random mutagenesis of the substituted T-domain, coupled to a pigment based screen for improved function. Using this approach, it was possible to recover improved variants of four different recombinant *bpsA* genes and derive a structural map of mutations that resulted in increased function. The screen developed is an extremely powerful tool for T-domain directed evolution, allowing clone numbers in excess of 10^6 to be qualitatively assessed in a matter of days, with a facile second tier of screening permitting subsequent quantification of the activity of improved clones. In combination with structural modelling and sequence alignment analyses, the results of these directed evolution experiments were used to derive a model for interaction between the T and TE-domains of BpsA. In this model, interaction requires that the T-domain adopt a specialised conformation, in which hydrophobic side chains previously at the core of the T-domain are shifted so that they form a shared hydrophobic interface between the T and TE-domains. A set of key residues involved in this interface was proposed based on

the directed evolution data. A more general model for T-domain interaction was also derived from results of this study as well as previous structural and biochemical data [53, 92-94, 192]. In this model interactions are mediated by conformational changes, and optimal function requires that a balance between the various conformational states be reached. On one hand formation of a particular interaction state must be stable enough to allow processing of a T-domain linked intermediate, but on the other hand, if a single state is overly stable this might be detrimental to formation of other interaction states required for function. Interaction between a T-domain and other NRPS domains is therefore dependent not only on residues that stabilise or destabilise the interface between two domains, but also on residues within the T-domain that govern the stability of a particular conformation. The models generated fit well with the data derived in this and other studies, and will provide a basis for subsequent biochemical studies of T-domain interactions in NRPS enzymes. Interaction models aside, this work has clearly shown that achieving optimal T-domain interactions is critical for the success of NRPS manipulation experiments; with directed evolution a powerful tool for optimising T-domain function in recombinant NRPS.

8.2.5 – Discovery and characterisation of PPTases

BpsA was also used as both an *in vivo* and *in vitro* reporter for characterisation and discovery of PPTase enzymes. *In vivo* characterisation showed that the three broad specificity PPTase enzymes (Sfp, PcpS and PP1183) were all capable of recognising and activating T-domains from *P. aeruginosa*, *P. syringae*, *E. coli*, *B. subtilis* and *S. lavendulae*. An *in vitro* characterisation assay was developed in which the rate of T-domain modification in BpsA could be indirectly determined by measuring the rate of increase in velocity of the pigment synthesis catalysed by BpsA. Using this assay it was possible to derive kinetic parameters for PcpS and PP1183. This assay presents a novel means for rapidly assessing the relative activity of PPTase enzymes that is considerably more convenient than existing HPLC and radiolabelling techniques. This same assay could also be applied to measure rates of regeneration of misprimed T-domains by type II thioesterases.

A high throughput *in vivo* screen was developed and implemented to recover novel PPTase enzymes from eDNA libraries. Using this technique three previously uncharacterised PPTases were recovered based on their ability to activate BpsA. Sequence analysis revealed the presence of NRPS and PKS genes in association with one of the recovered PPTase genes, indicating that this technique is also applicable to the discovery of novel secondary metabolite clusters from eDNA.

8.3 – Research applications

This research finds its most exciting application in the manipulation of NRPS enzymes to generate novel bioactive compounds. NRPS enzymes provide a diverse set of modular catalytic functions which can, in theory, be recombined to produce new molecules, either by targeted rational modification or combinatorial biosynthesis [13, 20, 81]. Combinatorial biosynthesis has the potential to generate vast libraries of bioactive molecules, which when combined with high-throughput activity screening would serve as a rich source of lead compounds for new pharmaceuticals [14-16]. An ideal tool box for NRPS mediated combinatorial biosynthesis would consist of a set of domains, each catalysing a desired reaction, that could be randomly combined in any order [82]. The diversity of molecules arising from such a tool box would derive primarily from the number, specificity and order of A-domains. This diversity could be augmented by alternative chain release mechanisms catalysed by different TE-domains [54, 60, 210], and the action of various tailoring enzymes [71, 72]. The main barrier to achieving this ideal situation is a lack of understanding of how NRPS domains interact during synthesis.

This study addressed two key issues impeding the development of combinatorial biosynthesis platforms. The first of these issues is that of C-domain specificity. In order for a recombinant NRPS to be functional, all C-domains contained therein must be able to process the donor and acceptor substrates encountered in their new setting. The *in vivo* results described in this study support findings from previous *in vitro* studies [48, 49], suggesting that C-domains do not exhibit specificity for the aa sidechains of donor substrates. Instead, the evidence points to

sub-optimal T-domain interactions being responsible for the inability of hybrid *pvdD* genes to restore pyoverdine synthesis in this and previous [85, 143] studies.

The domain swapping experiments and directed evolution experiments conducted using BpsA provide information on T-domain interactions that is highly relevant to potential combinatorial biosynthesis applications of NRPS. It was found that T-domain communication is central to correct function in recombinant NRPS enzymes, and that directed evolution of T-domain sequences in such enzymes can be used to improve function. The models developed for T-domain interaction and the map of key residues derived from directed evolution provide insight into why recombination of NRPS modules might fail, and how best to avoid or rectify this failure. This research contributes toward a more complete understanding of the role and mechanism of T-domain interaction in NRPS function, an understanding which may find application in the design of combinatorial biosynthesis platforms. Certain T-domain residues appear to be crucial for correct interaction with a specific partner domains [93, 94] however as evidenced by the results of directed evolution conducted in this and a previous study [92], a diverse set of sequence solutions exists for a single interaction. It would appear that the success or failure of NRPS manipulation experiments is dependent to a large extent on the dynamics of T-domains contained therein; accordingly determination of the rules governing T-domain dynamics and interactions is an essential pre-requisite for the development of NRPS based combinatorial biosynthesis platforms.

Another essential pre-requisite for the activity of hybrid NRPS systems, is that all carrier proteins in an enzyme system be activated by post translational attachment of a 4'-PP group by a cognate PPTase [40]. As the application of recombinant NRPS systems to production of new bioactive molecules increases, so to will the demand for PPTase enzymes capable of efficiently activating a diverse set of T-domains within a single host. The BpsA based system described in this thesis addresses this need by providing a convenient tool for rapidly assessing the activity of PPTases with a variety of T-domain substrates both *in vivo* and *in vitro*. It has also proven to be a simple and robust method for discovery of novel PPTase genes present in environmental DNA. Since PPTase genes are often found in association with the NRPS and

PKS genes they serve to activate, this system also has great potential for the discovery of new secondary metabolite clusters.

8.4 – Future directions

8.4.1 – Siderophore characterisation

The immediate continuation of this research into siderophore biosynthesis could focus on answering two questions. First, why do *P. syringae* and other fluorescent pseudomonads produce more than one siderophore? Pyoverdine was found to be superior to achromobactin for the scavenging of iron, and no phenotype could be detected for deletion of an achromobactin synthetase unless pyoverdine was also absent. So what is the physiological role of achromobactin and other secondary siderophores in these bacteria? A plausible hypothesis is that achromobactin acts as a first response to iron starvation, which is always present at low levels and allows cells to cope with sudden onset of iron starvation while synthesis of pyoverdine is established. A similar hypothesis has been put forward to explain the role of achromobactin in *D. dadantii*, however this was not experimentally verified [134]. An effective experimental approach to shed light on the functional role of achromobactin and pyoverdine would be the creation of *lacZ* fusion constructs to report on expression from the promoters of achromobactin and pyoverdine biosynthetic genes. This would allow determination of expression levels in response to varying degrees of iron starvation and would reveal whether low level constitutive expression of achromobactin biosynthetic genes is present. The role of other environmental cues such as temperature, nutrient availability and population density could also be investigated using this approach. Quantitative RTPCR is another method for determining transcript abundance that would be applicable. The second question is: What is the structure of the heavier of the two pyoverdine iso-forms detected, and how is it synthesised? The best starting point for investigation would be the separation of the two isoforms by HPLC followed by elucidation of a solution structure using NMR.

8.4.2 – T-domain function and interaction

A combination of structural modelling and bioinformatic analysis allowed meaningful interpretation of results from the domain swapping and directed evolution experiments described in this chapter however more a more detailed biochemical examination is required to test the hypotheses derived from this *in silico* study. In particular, the question of whether T-domains are specialised for interaction with a particular type of downstream domain should be addressed. A starting point for such examination would be to substitute a variety of additional T-domains into BpsA, to determine whether recombinant genes harbouring T-domains that interact with a TE-domain in nature have higher activity. Improvement of additional non-functional substitutions by directed evolution should also be conducted in order to refine the set of important residues already derived.

The role of individual residues in inter-domain communication could be determined by site directed mutagenesis of the native T-domain in BpsA. Non-functional enzymes recovered from mutagenesis studies could then be subjected to further biochemical characterisation to determine the precise cause for loss of functionality. By incubating non-functional enzymes with radiolabeled L-gln, it should be possible to determine whether function loss is a result of failure of a T-domain to interact with the A-domain of BpsA. If this were the case, the radiolabeled substrate would not be tethered to the enzyme and as a result would not be recovered along with precipitated protein. If on the other hand, product release was impaired due to inability of a mutant T-domain to interact with the TE-domain of BpsA one would expect to recover elevated levels of radiolabeled L-gln along with precipitated protein, due to the stalling of the substrate on the enzyme.

The hypothesis that the inactivity of the recombinant *pvdD* gene Ps(Thr)Thr is due to a breakdown in interaction between the substituted C-domain and the T-domains of PvdD will be tested using a directed evolution approach similar to that described in this thesis. This work will be conducted as part of the doctoral research of Mark Calcott. Experiments have already been designed to test this hypothesis. In these experiments each of the T-domains of Ps(Thr)Thr will be subjected to random mutagenesis followed by selection of variants that have restored function. The first tier of this selection process will be qualitative selection of

viable clones able to grow under severe iron starvation. The use of KB agar containing 200 µg/mL EDDHA will provide unambiguous selection for the first tier as only clones which have gained the ability to synthesise functional pyoverdine will be viable. If viable clones are recovered from this first tier, the intrinsic fluorescence of pyoverdine will enable a second, quantitative tier of screening. The establishment of an alternative directed evolution system for identification of residues involved in T-domain interactions will enable the generality of the models proposed in this thesis to be tested. In particular if successful, this second system will facilitate identification of T-domain residues important for interaction with both TE and C-domains. By substituting only the A-domain of *pvdD* for foreign domains specifying L-thr, it should be possible to determine whether improper communication between A and T-domains can also abrogate function in recombinant NRPS. If so, a map of key residues involved in this process could also be derived by directed evolution.

8.4.3 – Discovery and characterisation of PPTases

Concerning the application of BpsA for discovery and characterisation of PPTase enzymes, future experiments could aim to recover further PPTases enzymes from environmental DNA samples. The most interesting result from this will not necessarily be discovery of the PPTase enzymes themselves, but rather novel NRPS and PKS clusters that are usually found in association with PPTase enzymes. By creating eDNA libraries with a larger insert size in a BAC vector, it may even be possible to recover and characterise entire secondary metabolite clusters.

This system also lends itself to high throughput screening of libraries of compounds for inhibitors of PPTase enzymes *in vivo*. Since activation of BpsA is dependent on the activity of a co-expressed PPTase, inhibition of such a PPTase could be detected based on reduced pigment synthesis. The *in vitro* screen for PPTase characterisation could then serve as a second tier of screening allowing precise determination of the efficacy of compounds as well as their mode of inhibition. PPTases play an integral role in both primary and secondary metabolism of bacteria, and are in many cases essential proteins [40, 69]. As such, the discovery of inhibitors is of particular interest from the perspective of drug discovery [211].

8.5 – Concluding remarks

The role of NRPS in bacterial physiology, as well as the molecular mechanisms governing the function of these enzymes, presents a challenging and multi-faceted avenue of scientific investigation. The present understanding of the function of these enzymes has required an amalgamation of molecular biological, structural and chemical approaches and has reached the point where detailed mechanistic aspects of function are being illuminated. The microbiological research described in this thesis characterises the biosynthetic pathway and physiological significance of the NRPS derived siderophore pyoverdine in *P. syringae*. The domain swapping and directed evolution experiments conducted demonstrate the utility of two model systems for studying NRPS function. These model systems were the pyoverdine biosynthesis NRPS of *P. aeruginosa* and the pigment synthesising NRPS BpsA from *S. lavendulae*. Common to both of these models is the production of readily detectable products that are indicative of function of recombinant proteins. The results from these experiments contribute toward the understanding of how individual NRPS modules interact during the biosynthetic process, an understanding which is crucial for the success of future manipulation endeavours aimed at generating novel bioactive compounds. The protein BpsA also provided the basis for novel methods of PPTase characterisation and discovery which have potential applications in the discovery of secondary metabolite clusters from eDNA libraries as well as the discovery of PPTase inhibitors from libraries of chemical compounds.

Appendix One – Sequence alignments

A1.1 – Phylogenetic analysis of T-domains

It was theorised that since T-domains do not differ in their core function as carriers of substrates and intermediates, it is likely that the diversity of sequence information represents specialisation to interact with a particular domain or set of domains in a synthetase enzyme. In order to test this theory a simple bioinformatic analysis was carried out. Sequence diversity of NRPS T-domains from the genus *Pseudomonas* was compared to diversity of carrier proteins (ACPs) involved type two fatty acid synthesis (type II FAS) in the same genus. In contrast to NRPS T-domains, type II FAS ACPs interact with a limited and invariant set of proteins during synthesis of fatty acids [212, 213] so based on the theory outlined above would be expected to have much lower sequence diversity. As shown in Figure A1.1 the sequence of ACPs from 16 different species of *Pseudomonas* is highly conserved; in contrast the sequence of T-domains is highly divergent. This analysis is somewhat biased in that the single ACP in each species is likely to have arisen from a common ancestral protein whereas the multiple T-domains present in each species have been shaped by horizontal gene transfer and gene duplication events [124], however a comparison of ACP sequences from phylogenetically disparate bacteria (A1.1c) reveals that ACP sequences from multiple phyla are more highly conserved than T-domain sequences taken only from the genus *Pseudomonas*.

The initial analysis described, supported the theory that T-domains have evolved for optimal interaction with particular partner domains prompting a more detailed bioinformatic assessment of T-domain sequences. T-domains are required to interact with both upstream and downstream catalytic partners in an NRPS assembly line. There is less variation in the variety of upstream interaction partners, with most T-domains being required to interact with upstream A and C-domains. Downstream domain interactions are more varied, with three commonly occurring variations in downstream interaction partner: T-domains located in an elongation module must interact with a C domain, a subset of T-domains within elongation modules are also required to interact with a downstream E-domain and T-domains in a termination module have only a TE-domain as a downstream partner. In order to determine whether the type of downstream interaction partner correlated with specific sequence elements in a T-domain, a set of 16-20 sequences was compiled for each of the three downstream domain possibilities outlined. All T-domain sequences were taken from genome sequenced *Pseudomonas* species. Sequences were limited to a single genus in order to maximise the proportion of sequence variation arising from differences in downstream interaction partner. The alignment resulting from this set of sequences is shown in Figure A1.2; selected features that are specific to T-domains with a particular downstream interaction partner are highlighted. A subset of these positions, at which mutations were recovered during directed evolution, is also indicated. Table A1.1 gives a summary of the alignment analysis.

		-12	-8	-1	+1...+5	+10	+12	+24	+27
O41509A_T_C	RQLAAIWASV LNL SQV G I T D	D P F E L G G H S L	L A T Q V V S F I R	--	K Q L N V D M A	L R T I F E A P T I			
Psyr3722-a-t-c	HA. E. QAL . G. ER. RH. H.	
Psyr3722-a-t-c2	.A. D. QEL . GVDRA. RN. H.	
spTO4699-a-t-c	MA. E. QEL . DVER. RH. H.	
Pfl01_2213-a-t-c	TT. R. GEM . QVER. RR. H.	
psyr1959-a-t-c	Q. DL . K. ER. RS. N.	
Psyr2576-a-t-c	IAM. Q. . GIQR. RH. N.	
P	TM. ... L. CE. . SVER. RQ. H.	
P	QA. D. . KVG. . LS. S.	
sp	K. I. . V. RQ. . GVPR. LR.	
al	. RI. SE. . G. PR. LR.	
al	Q. V. DI . GVER. LH. H.	
Pflu2554-a-t-c	Q. ... L. QD. . VA. . LD. N.	
Pput1674-a-t-c	KDV. ... GEL . GVER. LD. N.	
pal4_pvdJ-a-t-c	. V. . T. . EL . KAER. LG.	
619_3541-a-t-c	C. ... Q. . E. . K. EK. LA. N.	
Pseen3230-a-t-c	K. I. . V. RQ. . GVPR. LR.	
Pspph1911-a-t-c	EK. ... V. . D. . K. E. . S. . N.	
pto2105_A_T_E	Q. I. D. . K. E. . A. . N.	
f4093_A_T_E2	Q. I. D. . K. E. . A. . N.	
f4093_A_T_E1	AR. VQV. QR. . KREH. . D. N.	
619_3542-a-t-e	. EI. D. . K. ER. LA. N.	
tgbl_4083-a-t-e	QR. D. . K. GR. LD. N.	
a7_2859-A-T-E	S. ... N. . QEL . GVER. RD. N.	
P	Q. ... T. . E. . K. ER. L. . N.	
al	EK. ... V. . D. . K. E. . S. . N.	
ps	Q. I. QDL . K. ER. RS. N.	
P	Q. ... I. . . QDL . K. ER. RS. N.	
pspto2135-a-t-e	Q. CA. . VEK. LN. N.	
alespvd_L-a-t-e	Q. GV. RE. . VER. LG. N.	
619_3562-a-t-e	E. ... Q. . RE. . V. R. VQ. N.	
Pflu2554-a-t-e	QKI. SD. . R. PR. LH. N.	
Pput1674-a-t-e-1	QR. ... L. QR. . K. P. . VS. N.	
Pput1674-a-t-e-2	Q. CA. . K. E. . LD. N.	
Pspph1911-a-t-e	Q. QD. . VEK. LN. N.	
r2608C_T_TE	IA. . NL. KEL . K. EK. SRH.	
TO4519A_T_Te	KSV. . V. KNT . . EFISLN. N.	
T02105A_T_TE	QR. ... L. QQ. . QVE. . LN. N.	
f4093_A_T_Te	Q. I. . L. SQ. . GVD. . LD. N.	
a_t-te_?dom_up	. T. . GLLGEL . DRPR. . R. S.	
T_TE_?domup	. T. . GLLGEL . DRPR. . R. S.	
1849_A_T_Te	EA. . GL. QQS . . IER. N.	
t1680_A_T_TE	QA. . TL. QTA . VERISRD. H.	
P	HA. E. QAL . G. ER. RH. H.	
sp	HA. E. KAL . G. ER. RH. H.	
61	. A. D. . G. E. . VS. H.	
tgbl_1805-a-t-te	LA. . H. . QD. . GVEST. VD. N.	
seen2717-a-t-te	LT. YQ. KGL . LAPHI. . R. N.	
spto_4519-a-t-te	KSV. . V. KNT . . EFISLN. N.	
7_2859-A-T-TE	QA. . G. . QE. . GIARI. VH. N.	
Pflu3225?-a-t-te	AL. EV. . Q. . G. ER. E. . A.	
a14_pvdD-a-t-te	QRI. EI . GVER. LD. N.	
PP4219-a-t-te	LAI. R. . QD. . KVE. N.	
ales_pvdD-a-t-te	KD. . RL. RE. . GVER. LH. N.	
Psyr1960-a-t-te	QR. . VL. EQ. . HVER. . LN. N.	
Identity	:	*	**	**	*	*	*	*	

Figure A1.2 – Interaction specific features of T-domain sequences

Sequences are grouped according to downstream interaction partner (C, E or TE-domain) as indicated on the left. Boxes indicate positions at which residue identity correlates with downstream interaction partner. Blue boxes indicate positions at which mutations leading to improved function were recovered in directed evolution experiments (Chapter 6)

Table A1.1 – Summary of alignment features shown in Figure A1.2

	-12	-8*	-1*	+1*	+2*	+3
C	Always G	Variable	Always H	Always L	Always L	A (1 x V)
E	Always G	Always N	Always D	Always I	V, L, I, M	Always S
TE	G or S	Variable	Variable	L or I	Variable	Variable
	+4**	+10	+12	+24***	+27	
C	Variable	Variable	Insert	F (1 x L)	Variable	
E	L, I	Always Q	Gap	Always F	Always Q	
TE	V, L, I, M	Variable	Insert	V, L, I, M	Variable	

Left hand column indicates downstream interaction domain. Top row indicates position in T-domain. Positions marked * were altered in directed evolution experiments. Position marked ** was altered in directed evolution experiments and implicated in T-TE communication using structural modelling. Position marked *** was altered in directed evolution experiments implicated in T-TE communication using structural modelling (this study) and found to be important in T-TE communication in a previous mutagenesis study [93]

A1.1.1 – Examination of a module duplication event in PvdD

The two modules of the *P. aeruginosa* pyoverdine NRPS PvdD both specify L-thr. The high level of identity between the CAT domains of these two modules (90%) suggests the occurrence of a module duplication event in the evolution of PvdD. Examining the individual domains in these putative duplicated modules reveals that the A-domains have the highest sequence identity (99.5 %) followed by the C-domains (80 %) and finally the T-domains (72 %). Interestingly, the two T-domains of PvdD are 97 % identical in the region before the 4'-PP attachment site and only 20 % identical thereafter (Figure A1.3). The divergence in the sequence of the two T-domains over this region may have been driven by requirement of each to interact with a different downstream partner (C-domain vs. TE-domain) following duplication.

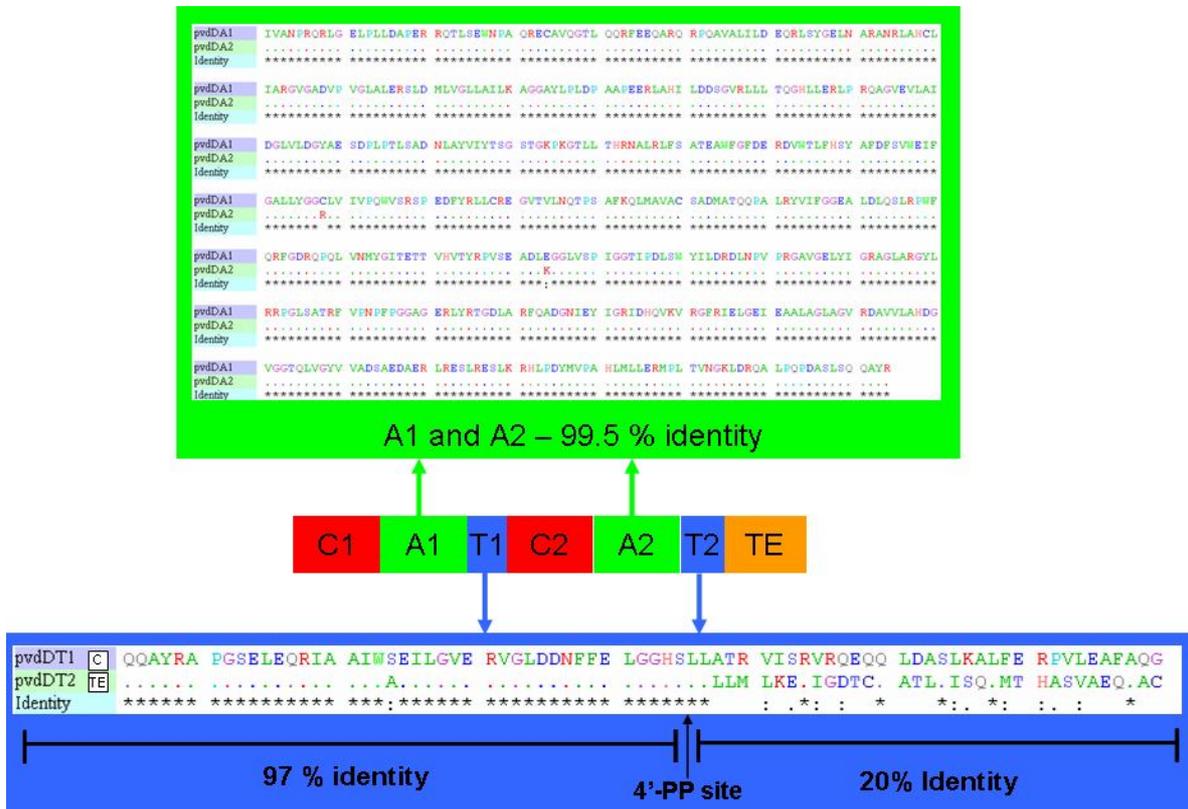


Figure A1.3 – T-domain sequence divergence following a putative module duplication event in PvdD

Alignments of the two A-domains and the two T-domains of PvdD are shown. It is likely, given the high level of identity between them; that the two CAT modules of PvdD are the product of a module duplication event. The divergence in T-domain sequence after the 4'-PP site may reflect specialisation of the domains for interaction with different downstream partners.

A1.2 – Complete alignments for all improved clones

Complete alignments for all improved clones are shown, in each case the top row is the unimproved sequence, immediately below this is the T-domain of BpsA, and the remaining sequences are improved clones. Alignments are annotated according to the key shown on the following page.

pvdDT1	YRA	PGSELEQRIA	AIWSEILGVE	RVGLDDNFFE	LGGHSLLAIR	VISRVRQEQQ	LDASLKALIE	RPVLEAFAQG
BPSAT	FV	.RT.T.KE..	.V.EKA.RR.	NASVQ.D...	S..N..I.VG	LVRELNARLG	VSLP.QSVL.	S.TI.KL.RR
3kf0					I		I	
3kf1					I		I	V
3kf2					I		I	S
3kF13					A		I	
3kf15				S	I		I	
3kf21					I		I	
3kf24							I	A
3pf2							I	
Identity	:	* * * *	* * * *	* * * *

-3 +4 +24

Improved clones from evolution of sIPvD

entfT	PGRA	PKAGSETIIA	AAPSSLLGCD	VQDADADFFA	LGGHSLLAMK	IAAQLSRQVA	RQVTPGQVMV	ASTVAKLATI	ID
BPSAT	FV	.RTET.KE..	.VWEKA.RRE	NASVQD...	E S..N..I.VG	VRE.NARLG	VSLPLQS.LE	SP.IE...	RR LE
5k1					I				
5k3					I				
5k5					I				
5k17					I	V			
5k21					I	V			
5k22					I				
5k23					I				
5k24					I				
Identity	*	* * * *	* * * *	* * * *

+1 +7

Improved clones from evolution of sIEntF

3kf0Taa	AELVERPYRA	PGSELEQRIA	AIWSEILGVE	RVGLDDNFFE	LGGHSLLAIR	VISRVRQEQQ	LDASLKALIE	RPVLEAFAQG	LER
BPSATaa	FV	.RT.T.KE..	.V.EKA.RR.	NASVQ.D...	S..N..I.VG	LVRELNARLG	VSLP.QSVL.	S.TI.KL.RR	...
pr2h1Taa			E		C				
Pr2h2Taa					C				
Pr2h3Taa									
Pr2h4Taa	L			D					
Pr2h5Taa									
Pr2h6Taa									
Pr2h7Taa	C								
Pr2h8Taa									P
Pr2h9Taa									
Pr2h10Taa									
Pr2h11Taa	P								M
pr2h12Taa				D	D				
pr2h13Taa			A						
Pr2h14Taa					C				
Pr2h15Taa									
PvdT*					T			F	
Identity	*****	* * * *	* * * *	* * * *

- 39 -16 - 8 - 5 -1 +2 +8 +15 +26 + 33 + 38

Improved clones from second round evolution of 3kF0

Appendix 2 – Unit conversion for PPTase kinetic parameters

The raw data input into SigmaPlot for derivation of kinetic parameters for the PPTases PcpS and PP1183 was maximum change in indigoidine (Ind) synthesis reaction rate (change in gradient per second) vs. Concentration of CoA present in the reaction. The Vmax values given by SigmaPlot therefore had the units “Change in gradient per second” The assumptions made and calculations used to convert this into a slightly more useable unit are given below.

Analysis of Ind synthesis reaction rates for various concentrations of ATP had previously revealed a good fit to the Michaelis-Menten equation. As such the assumption can be made that reaction rate for this enzyme is directly proportional to the concentration of active enzyme according to the equation:

$$V_0 = K_{cat} [E] [S] / K_m + [S]$$

For the reactions used in derivation of kinetic parameters of PPTases, [BpsA] = [E] = 1.64 uM. [S] = [ATP] = 1mM. K_{cat} (BpsA) = 241.34 min⁻¹, K_m (ATP/BpsA) = 11.12 mM.

Substitution of these values into the above equation gives a theoretical maximum velocity of 32.487 uM/min if all BpsA present in the reaction is activated by the PPTase.

The gradient units for the data input into SigmaPlot were gradient of 1.0 = change in A590 of 0.001/min (assigned by SoftmaxPro, don't blame me!).

The relationship between [indigoidine] and A590 for the path length used in the assay is A590 of 0.001 = 0.0854 uM Indigoidine. Therefore the maximum gradient attainable if all BpsA is active is 32.487/0.0854 = 380.409.

A concentration of 1.64 uM BpsA in a 200 uL reaction means the amount of BpsA present is 328 pmoles.

Going back to the original assumption that velocity is proportional to concentration of active BpsA in a reaction. If 328 pmoles active BpsA give a gradient of 380.4 then the theoretically the amount of active BpsA that would give a gradient of 1.0 is $328/380.4 = 0.86$ pmoles

This establishes a relationship between gradient and concentration of active BpsA which allows the conversion of “change in gradient per second values to” to amount of BpsA activated per second. For example if the instantaneous gradient were measured at a given point, and one second later and found to have increased by 1.0, the above relationship tells us that an extra 1.1587 pmoles of active BpsA are present.

This relationship was used to convert the units of the V_{max} values given by SigmaPlot from “Change in gradient pre second” to “pmoles BpsA activated per second”. Since BpsA has only one T-domain, one molecule of CoA is used to activate one molecule of BpsA and the amount of BpsA activated per second = amount of CoA consumed by PPTase per second.

Bibliography

1. Finking, R. and M.A. Marahiel. Biosynthesis of nonribosomal peptides. Annual Review of Microbiology, 2004. **58**: p. 453-88.
2. Challis, G.L. and J.H. Naismith. Structural aspects of non-ribosomal peptide biosynthesis. Current Opinion in Structural Biology, 2004. **14**: p. 748-56.
3. Marahiel, M.A. and L.O. Essen. Nonribosomal peptide synthetases: Mechanistic and structural aspects of essential domains. Methods in Enzymology, 2009. **458**: p. 337-351.
4. Marahiel, M.A., T. Stachelhaus, and H.D. Mootz. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chemical Reviews, 1997. **97**: p. 2651-2674.
5. Schweizer, H.P. and T.T. Hoang. An improved system for gene replacement and *xyIE* fusion analysis in *Pseudomonas aeruginosa*. Gene, 1995. **158**: p. 15-22.
6. Stein, T., J. Vater, V. Kruff, et al. The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates. Journal of Biological Chemistry, 1996. **271**: p. 15428-15435.
7. Stachelhaus, T. and M.A. Marahiel. Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. FEMS Microbiology Letters, 1995. **125**: p. 3-14.
8. J.A. Vizcaíno, L.S., R.E. Cardoza, E. Monte, S. Gutiérrez,. Detection of putative peptide synthetase genes in *trichoderma* species: Application of this method to the cloning of a gene from *T. harzianum* CECT2413. FEMS Microbiology Letters, 2005. **244**: p. 139-148.
9. Caboche, S., M. Pupin, V. Leclere, et al. Norine: A database of nonribosomal peptides. Nucleic Acids Research, 2007. **36**. D326-D331
10. Meier, J.L. and M.D. Burkart. The chemical biology of modular biosynthetic enzymes. Chemical Society Reviews, 2009. **38**: p. 2012-2045.
11. Koglin, A. and C.T. Walsh. Structural insights into nonribosomal peptide enzymatic assembly lines. Natural Product Reports, 2009. **26**: p. 987-1000.

12. Caboche, S., M. Pupin, V. Leclere, et al. Structural pattern matching of nonribosomal peptides. BMC Structural Biology, 2009. **9**: p. 15.
13. Stevens, B.W., T.M. Joska, and A.C. Anderson. Progress toward re-engineering non-ribosomal peptide synthetase proteins: A potential new source of pharmacological agents. Drug Development Research, 2005. **66**: p. 9-18.
14. Shen, B., M. Chen, Y. Cheng, et al., Prerequisites for combinatorial biosynthesis: Evolution of hybrid NRPS/PKS gene clusters, in Biocombinatorial Approaches for Drug Finding. 2005, Springer. p. 107-126.
15. Doekel, S., M.-F. Coeffet-Le Gal, J.-Q. Gu, et al. Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. Microbiology, 2008. **154**: p. 2872-2880.
16. Baltz, R.H., P. Brian, V. Miao, et al. Combinatorial biosynthesis of lipopeptide antibiotics in *Streptomyces roseosporus*. Journal of Industrial Microbiology and Biotechnology, 2006. **33**: p. 66-74.
17. Stein, D.B., U. Linne, and M.A. Marahiel. Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. FEBS Journal, 2005. **272**: p. 4506-20.
18. Duerfahrt, T., K. Eppelmann, R. Muller, et al. Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. Chemistry and Biology, 2004. **11**: p. 261-71.
19. Cane, D.E., C.T. Walsh, and C. Khosla. Harnessing the biosynthetic code: Combinations, permutations, and mutations. Science, 1998. **282**: p. 63-68.
20. Stachelhaus, T., A. Schneider, and M.A. Marahiel. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. Science, 1995. **269**: p. 69-72.
21. Doekel, S. and M.A. Marahiel. Dipeptide formation on engineered hybrid peptide synthetases. Chemistry and Biology, 2000. **7**: p. 373-384.
22. Mootz, H. and M. Marahiel. The tyrocidine biosynthesis operon of *Bacillus brevis*: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. Journal of Bacteriology, 1997. **179**: p. 6843-6850.
23. von Dohren, H., U. Keller, J. Vater, et al. Multifunctional peptide synthetases. Chemical Reviews, 1997. **97**: p. 2675-2706.

24. Stachelhaus, T., H.D. Mootz, V. Bergendahl, et al. Peptide bond formation in nonribosomal peptide biosynthesis: Catalytic role of the condensation domain. Journal of Biological Chemistry, 1998. **273**: p. 22773-22781.
25. Conti E, Stachelhaus T, Marahiel MA, et al. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. The EMBO Journal, 1997. **16**: p. 4174-4183.
26. May, J.J., N. Kessler, M.A. Marahiel, et al. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. Proceedings of the National Academy of Sciences U S A, 2002. **99**: p. 12120-5.
27. Yonus, H., P. Neumann, S. Zimmermann, et al. Crystal structure of DltA: Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. Journal of Biological Chemistry, 2008. **283**: p. 32484-32491.
28. Weber, T. and M.A. Marahiel. Exploring the domain structure of modular nonribosomal peptide synthetases. Structure, 2001. **9**: p. R3-9.
29. von Dohren, H., R. Dieckmann, and M. Pavela-Vrancic. The nonribosomal code. Chemistry and Biology, 1999. **6**: p. R273-R279.
30. Dieckmann, R., Pavela-Vrancic, M., Pfeifer E., Döhren, H. and Kleinkauf, H. The adenylation domain of tyrocidine synthetase 1. European Journal of Biochemistry, 1997. **247**: p. 1074-1082.
31. Saito, M., K. Hori, T. Kurotsu, et al. Three conserved glycine residues in valine activation of gramicidin S synthetase 2 from *Bacillus brevis*. Journal of Biochemistry, 1995. **117**: p. 276-282.
32. Gocht, M. and M.A. Marahiel. Analysis of core sequences in the D-phe activating domain of the multifunctional peptide synthetase tyca by site-directed mutagenesis. Journal of Bacteriology, 1994. **176**: p. 2654-2662.
33. Tokita, K., K. Hori, T. Kurotsu, et al. Effect of single base substitutions at glycine-870 codon of gramicidin s synthetase 2 gene on proline activation. Journal of Biochemistry, 1993. **114**: p. 522-527.
34. Pavela-Vrancic, M., E. Pfeifer, H. van Liempt, et al. ATP binding in peptide synthetases: Determination of contact sites of the adenine moiety by photoaffinity

- labeling of tyrocidine synthetase 1 with 2-azidoadenosine triphosphate. *Biochemistry*, 1994. **33**: p. 6276-6283.
35. Challis, G.L., J. Ravel, and C.A. Townsend. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry and Biology*, 2000. **7**: p. 211-24.
 36. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, 1999. **6**: p. 493-505.
 37. Ackerley, D.F., T.T. Caradoc-Davies, and I.L. Lamont. Substrate specificity of the nonribosomal peptide synthetase PvdD from *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 2003. **185**: p. 2848-2855.
 38. Husi, H., K. Schörgendorfer, G. Stempfer, et al. Prediction of substrate-specific pockets in the cyclosporin synthetase. *FEBS Letters*, 1997. **414**: p. 532-536.
 39. Rausch, C., T. Weber, O. Kohlbacher, et al. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMS). *Nucleic Acids Research*, 2005. **33**: p. 5799-5808.
 40. Mootz, H.D., R. Finking, and M.A. Marahiel. 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *Journal of Biological Chemistry*, 2001. **276**: p. 37289-37298.
 41. Walsh, C.T., A.M. Gehring, P.H. Weinreb, et al. Post-translational modification of polyketide and nonribosomal peptide synthetases. *Current Opinion in Chemical Biology*, 1997. **1**: p. 309-315.
 42. Lambalot, R.H., A.M. Gehring, R.S. Flugel, et al. A new enzyme superfamily the phosphopantetheinyl transferases. 1996. **3**: p. 923-936.
 43. Weber, T., R. Baumgartner, C. Renner, et al. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure*, 2000. **8**: p. 407-418.
 44. Koglin, A., M.R. Mofid, F. Lohr, et al. Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science*, 2006. **312**: p. 273-276.

45. Bergendahl, V., U. Linne, and M.A. Marahiel. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry*, 2002. **269**: p. 620-629.
46. De Crecy-Lagard, V., P. Marliere, and W. Saurin. Multienzymatic non ribosomal peptide biosynthesis: Identification of the functional domains catalysing peptide elongation and epimerisation. *Comptes Rendus de l'Académie des Sciences. Série III*, 1995. **318**: p. 927-936.
47. Samel, S.A., G. Schoenafinger, T.A. Knappe, et al. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure*, 2007. **15**: p. 781-792.
48. Belshaw, P.J., , C.T. Walsh, et al. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, 1999. **284**: p. 486.
49. Ehmann, D.E., J.W. Trauger, T. Stachelhaus, et al. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chemistry and Biology*, 2000. **7**: p. 765-772.
50. Rausch, C., I. Hoof, T. Weber, et al. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, 2007. **7**: p. 78.
51. Roongsawang, N., S.P. Lim, K. Washio, et al. Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. *FEMS Microbiology Letters*, 2005. **252**: p. 143-151.
52. Keating, T.A., C.G. Marshall, C.T. Walsh, et al. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, 2002. **9**: p. 522-6.
53. Tanovic, A., S.A. Samel, L.-O. Essen, et al. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science*, 2008. **321**: p. 659-663.
54. de Ferra, F., F. Rodriguez, O. Tortora, et al. Engineering of peptide synthetases. Key role of the thioesterase like domain for efficient production of recombinant peptides. *Journal of Biological Chemistry*, 1997. **272**: p. 25304-25309.

55. Bruner, S.D., T. Weber, R.M. Kohli, et al. Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure*, 2002. **10**: p. 301-310.
56. Trauger, J.W., R.M. Kohli, H.D. Mootz, et al. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature*, 2000. **407**: p. 215-218.
57. Lawson, D.M., U. Derewenda, L. Serre, et al. Structure of a myristoyl-*acp*-specific thioesterase from *Vibrio harveyi*. *Biochemistry*, 1994. **33**: p. 9382-9388.
58. Nardini, M. and B.W. Dijkstra. α/β hydrolase fold enzymes: The family keeps growing. *Current Opinion in Structural Biology*, 1999. **9**: p. 732-737.
59. Shaw-Reid, C.A., N.L. Kelleher, H.C. Losey, et al. Assembly line enzymology by multimodular nonribosomal peptide synthetases: The thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chemistry and Biology*, 1999. **6**: p. 385-400.
60. Keating, T.A., D.E. Ehmann, R.M. Kohli, et al. Chain termination steps in nonribosomal peptide synthetase assembly lines: Directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *ChemBioChem*, 2001. **2**: p. 99-107.
61. Li, J., R. Szittner, Z.S. Derewenda, et al. Conversion of serine-114 to cysteine-114 and the role of the active site nucleophile in acyl transfer by myristoyl-*acp* thioesterase from *Vibrio harveyi*. *Biochemistry*, 1996. **35**: p. 9967-9973.
62. Samel, S.A., B. Wagner, M.A. Marahiel, et al. The thioesterase domain of the fengycin biosynthesis cluster: A structural base for the macrocyclization of a non-ribosomal lipopeptide. *Journal of Molecular Biology*, 2006. **359**: p. 876-889.
63. Lautru, S. and G.L. Challis. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology*, 2004. **150**: p. 1629-1636.
64. Marahiel, M.A. Working outside the protein-synthesis rules: Insights into non-ribosomal peptide synthesis. *Journal of Peptide Science*, 2009. **15**: p. 799-807.
65. Kopp, F. and M.A. Marahiel. Where chemistry meets biology: The chemoenzymatic synthesis of nonribosomal peptides and polyketides. *Current Opinion in Biotechnology*, 2007. **18**: p. 513-520.

66. Schwarzer, D., H.D. Mootz, U. Linne, et al. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proceedings of the National Academy of Sciences U S A*, 2002. **99**: p. 14083-8.
67. Heathcote, M.L., J. Staunton, and P.F. Leadlay. Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & Biology*, 2001. **8**: p. 207-220.
68. Yeh, E., R.M. Kohli, S.D. Bruner, et al. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-s-enzyme that cannot be elongated. *Chembiochem*, 2004. **5**: p. 1290-1293.
69. Finking, R., J. Solsbacher, D. Konz, et al. Characterization of a new type of phosphopantetheinyl transferase for fatty acid and siderophore synthesis in *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, 2002. **277**: p. 50293-50302.
70. Reuter, K., M.R. Mofid, M.A. Marahiel, et al. Crystal structure of the surfactin synthetase-activating enzyme Sfp: A prototype of the 4'-phosphopantetheinyl transferase superfamily. *EMBO Journal*, 1999. **18**: p. 6823-6831.
71. Samel, S.A., M.A. Marahiel, and L.-O. Essen. How to tailor non-ribosomal peptide products—new clues about the structures and mechanisms of modifying enzymes. *Molecular Biosystems*, 2008. **4**: p. 387-393.
72. Walsh, C.T., H. Chen, T.A. Keating, et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current Opinion in Chemical Biology*, 2001. **5**: p. 525-534.
73. Linne, U. and M.A. Marahiel. Control of directionality in nonribosomal peptide synthesis: Role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry*, 2000. **39**: p. 10439-10447.
74. Luo, L., R.M. Kohli, M. Onishi, et al. Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: Kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry*, 2002. **41**: p. 9184-9196.

75. Weber, G., K. Schörgendorfer, E. Schneider-Scherzer, et al. The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Current Genetics*, 1994. **26**: p. 120-125.
76. OBrien, D.P., P.N. Kirkpatrick, S.W. OBrien, et al. Expression and assay of an N-methyltransferase involved in the biosynthesis of a vancomycin group antibiotic. *Chemical Communications*, 2000. 103-104.
77. Lauer, B., R. Russwurm, and C. Bormann. Molecular characterization of two genes from *Streptomyces tendae* required for the formation of the 4-formyl-4-imidazolin-2-one-containing nucleoside moiety of the peptidyl nucleoside antibiotic nikkomycin. *European Journal of Biochemistry*, 2000. **267**: p. 1698-1706.
78. Singh, G.M., P.D. Fortin, A. Koglin, et al. Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: Characterization of two candidate hydroxylases AspH and SyrP in *Pseudomonas syringae*. *Biochemistry*, 2008. **47**: p. 11310-11320.
79. Du, L., M. Chen, C. Sánchez, et al. An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *streptomyces verticillus* ATCC15003. *FEMS Microbiology Letters*, 2000. **189**: p. 171-175.
80. Takahashi, H., T. Kumagai, K. Kitani, et al. Cloning and characterization of a streptomyces single module type non-ribosomal peptide synthetase catalyzing a blue pigment synthesis. *Journal of Biological Chemistry*, 2007. **282**: p. 9073-9081.
81. Wenzel, S.C. and R. Muller. Formation of novel secondary metabolites by bacterial multimodular assembly lines: Deviations from textbook biosynthetic logic. *Current Opinion in Chemical Biology*, 2005. **9**: p. 447-458.
82. Keller, U. and F. Schauwecker. Combinatorial biosynthesis of non-ribosomal peptides. *Combinatorial Chemistry and High Throughput Screening*, 2003. **6**: p. 527-40.
83. de Boer, A.L. and C. Schmidt-Dannert. Recent efforts in engineering microbial cells to produce new chemical compounds. *Current Opinion in Chemical Biology*, 2003. **7**: p. 273-278.
84. Schneider, A., T. Stachelhaus, and M.A. Marahiel. Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Molecular Genetics and Genomics*, 1998. **257**: p. 308-318.

85. Ackerley, D.F. and I.L. Lamont. Characterization and genetic manipulation of peptide synthetases in *Pseudomonas aeruginosa* PAO1 in order to generate novel pyoverdines. Chemistry and Biology, 2004. **11**: p. 971-80.
86. Eppelmann, K., T. Stachelhaus, and M.A. Marahiel. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. Biochemistry, 2002. **41**: p. 9718-26.
87. Baltz, R.H. Biosynthesis and genetic engineering of lipopeptides in *Streptomyces roseosporus*. methods in enzymology, 2009. **58**: p. 511-531.
88. Miao, V., M.-F. Coëffet-Le Gal, K. Nguyen, et al. Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. Chemistry and Biology, 2006. **13**: p. 269-276.
89. Gal, M.-F.C.-L., L. Thurston, P. Rich, et al. Complementation of daptomycin *dptA* and *dptD* deletion mutations *in-trans* and production of hybrid lipopeptide antibiotics. Microbiology, 2006. **152**: p. 2993-3001.
90. Gu, J.-Q., K.T. Nguyen, C. Gandhi, et al. Structural characterization of daptomycin analogues produced by a recombinant streptomyces roseosporus strain. Journal of Natural Products, 2007. **70**: p. 233-240.
91. Nguyen, K.T., D. Ritz, J.-Q. Gu, et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin. Proceedings of the National Academy of Sciences U S A, 2006. **103**: p. 17462-17467.
92. Zhou, Z., J.R. Lai, and C.T. Walsh. Directed evolution of aryl carrier proteins in the enterobactin synthetase. Proceedings of the National Academy of Sciences U S A, 2007. **104**: p. 11621-11626.
93. Zhou, Z., J.R. Lai, and Christopher T. Walsh. Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. Chemistry and Biology, 2006. **13**: p. 869-879.
94. Lai, J.R., M.A. Fischbach, D.R. Liu, et al. A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. Proceedings of the National Academy of Sciences U S A, 2006. **103**: p. 5314-5319.
95. Rivas, L.A., J. Mansfield, G. Tsiamis, et al. Changes in race-specific virulence in *Pseudomonas syringae* pv. *Phaseolicola* are associated with a chimeric transposable

- element and rare deletion events in a plasmid-borne pathogenicity island. Applied and Environmental Microbiology, 2005. **71**: p. 3778-3785.
96. Milton, D.L., R. O'Toole, P. Horstedt, et al. Flagellin A is essential for the virulence of *vibrio anguillarum*. Journal of bacteriology, 1996. **178**: p. 1310-1329.
97. Baynham, P.J., D.M. Ramsey, B.V. Gvozdyev, et al. The *Pseudomonas aeruginosa* ribbon-helix-helix DNA-binding protein AlgZ (Amrz) controls twitching motility and biogenesis of type IV pili. Journal of Bacteriology, 2006. **188**: p. 132-140.
98. West, S.E.H., H.P. Schweizer, C. Dall, et al. Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from puc18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. Gene, 1994. **148**: p. 81-86.
99. Prosser, G.A., J.N. Copp, S.P. Syddall, et al. Discovery and evaluation of *Escherichia coli* nitroreductases that activate the anti-cancer prodrug cb1954. Biochemical Pharmacology. **79**: p. 678-687.
100. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, 1983. **166**: p. 557-580.
101. Sambrook, J. and D.W. Russell. Transformation of *E. coli* by electroporation. Cold Spring Harbor Protocols, 2006 – doi:10.1101/pdb.prot3933.
102. Choi, K.-H., A. Kumar, and H.P. Schweizer. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. Journal of Microbiological Methods, 2006. **64**: p. 391-397.
103. Schwyn, B. and J. Neilands. Universal chemical assay for the detection and determination of siderophores. . Analytical Biochemistry 1987 **160**: p. 47-56.
104. Berti, A.D. and M.G. Thomas. Analysis of achromobactin biosynthesis by *pseudomonas syringae* pv. *Syringae* b728a. Journal of Bacteriology., 2009. **191**: p. 4594-4604.
105. Meyer, J.M., A. Stintzi, D. De Vos, et al. Use of siderophores to type pseudomonads: The three pseudomonas aeruginosa pyoverdine systems. Microbiology, 1997. **143**: p. 35-43.

106. de Marco, A. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. Nature Protocols, 2007. **2**: p. 2632-2639.
107. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 1970. **227**: p. 680-685.
108. Sambrook, J. and D.W. Russell. Molecular cloning: A laboratory manual. 3 ed. 2003: Cold Spring Harbour Laboratory Press.
109. Lee, S.G. and F. Lipmann. Isolation of amino acid activating subunit-pantetheine protein complexes: Their role in chain elongation in tyrocidine synthesis. Proceedings of the National Academy of Sciences U S A, 1977. **74**: p. 2343-2347.
110. Upson, R.H., R.P. Haugland, M.N. Malekzadeh, et al. A spectrophotometric method to measure enzymatic activity in reactions that generate inorganic pyrophosphate. Analytical Biochemistry, 1996. **243**: p. 41-45.
111. McQuade, T.J., A.D. Shallop, A. Sheoran, et al. A nonradioactive high-throughput assay for screening and characterization of adenylation domains for nonribosomal peptide combinatorial biosynthesis. Analytical Biochemistry, 2009. **386**: p. 244-250.
112. Benkert, P., M. Künzli, and T. Schwede. Qmean server for protein model quality estimation. Nucleic Acids Research, 2009. **37**: p. W510–W514.
113. Braun, V. and H. Killmann. Bacterial solutions to the iron-supply problem. Trends in Biochemical Sciences, 1999. **24**: p. 104-109.
114. Cornelis, P. and S. Matthijs, *Pseudomonas* siderophores and their biological significance, in Microbial siderophores. 2007. p. 193-203.
115. Höfte, M. and P. Vos, Plant pathogenic pseudomonas species, in Plant-associated bacteria. 2006. p. 507-533.
116. He, J., R.L. Baldini, E. Dalziel, et al. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. Proceedings of the National Academy of Sciences U S A, 2004. **101**: p. 2530-2535.
117. Meyer, J., A. Neely, A. Stintzi, et al. Pyoverdinin is essential for virulence of *pseudomonas aeruginosa*. Infection and Immunity, 1996. **64**: p. 518-523.

118. Weber, T., C. Rausch, P. Lopez, et al. Clusean: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. Journal of Biotechnology, 2009. **140**: p. 13-17.
119. Ravel, J. and P. Cornelis. Genomics of pyoverdine-mediated iron uptake in pseudomonads. Trends in Microbiology, 2003. **11**: p. 195-200.
120. Meyer, J. and M. Abdallah. The fluorescent pigment of *pseudomonas fluorescens*: Biosynthesis, purification and physicochemical properties. Journal of General Microbiology 1978. **107**: p. 319-328.
121. Visca, P., F. Imperi, and I. Lamont, Pyoverdine synthesis and its regulation in fluorescent pseudomonads, in Microbial siderophores. 2007. p. 135-163.
122. Budzikiewicz, H. Siderophores of the pseudomonadaceae sensu stricto (fluorescent and non-fluorescent *pseudomonas* spp.). Progress in the Chemistry of Organic Natural Products, 2004. **87**: p. 81 - 237.
123. Smith, E., E. Sims, D. Spencer, et al. Evidence for diversifying selection at the pyoverdine locus of *Pseudomonas aeruginosa*. Journal of Bacteriology, 2005. **187**: p. 2138 - 2147.
124. Tummler, B. and P. Cornelis. Pyoverdine receptor: A case of positive darwinian selection in *Pseudomonas aeruginosa*. J. Bacteriol., 2005. **187**: p. 3289-3292.
125. Crosa, J.H. and C.T. Walsh. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiology and Molecular Biology Reviews, 2002. **66**: p. 223-49.
126. Greenwald, J., G. Zeder-Lutz, A. Hagege, et al. The metal dependence of pyoverdine interactions with its outer membrane receptor FpvA. Journal of Bacteriology, 2008. **190**: p. 6548-6558.
127. Visca, P., F. Imperi, and I.L. Lamont. Pyoverdine siderophores: From biogenesis to biosignificance. Trends in Microbiology, 2007. **15**: p. 22-30.
128. Visca, P., L. Leoni, M.J. Wilson, et al. Iron transport and regulation, cell signalling and genomics: Lessons from *Escherichia coli* and *Pseudomonas*. Molecular Microbiology, 2002. **45**: p. 1177-1190.
129. Swingle, B., D. Thete, M. Moll, et al. Characterization of the PvdS-regulated promoter motif in *pseudomonas syringae* pv. *Tomato* DC3000 reveals regulon members and

- insights regarding pvds function in other pseudomonads. *molecular microbiology*, 2008. **68**: p. 871-889.
130. Lamont, I., P. Beare, U. Ochsner, et al. Siderophore-mediated signaling regulates virulence factor production in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences U S A*, 2002. **99**: p. 7072 - 7077.
131. Schmelz, S., N. Kadi, S.A. McMahon, et al. AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nature Chemical Biology*, 2009. **5**: p. 174-182.
132. Challis, G. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. *Chembiochem*, 2005. **6**: p. 601-611.
133. Gulick, A.M. Ironing out a new siderophore synthesis strategy. *Nature Chemical Biology*, 2009. **5**: p. 143-144.
134. Franza, T., B. Mahe, and D. Expert. *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. *Molecular Microbiology*, 2005. **55**: p. 261-275.
135. Ochsner, U.A., P.J. Wilderman, A.I. Vasil, et al. Genechip expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: Identification of novel pyoverdine biosynthesis genes. *Molecular Microbiology*, 2002. **45**: p. 1277-1287.
136. Bultreysa, A., I. Gheysen, B. Wathélet, et al. The pyoverdins of *pseudomonas syringae* and *pseudomonas cichorii*. *Zeitschrift für Naturforschung*, 2004. **59**: p. 613-618.
137. Jülich, M., K. Taraz, H. Budzikiewicz, et al. The structure of the pyoverdin isolated from various *pseudomonas syringae* pathovars. *Zeitschrift für Naturforschung.*, 2001. **56**: p. 687-694.
138. Joardar, V., M. Lindeberg, R.W. Jackson, et al. Whole-genome sequence analysis of *Pseudomonas syringae* pv. *Phaseolicola* 1448a reveals divergence among pathovars in genes involved in virulence and transposition. *Journal of Bacteriology.*, 2005. **187**: p. 6488-6498.
139. Moon, C., X.-X. Zhang, S. Matthijs, et al. Genomic, genetic and structural analysis of pyoverdine-mediated iron acquisition in the plant growth-promoting bacterium *Pseudomonas fluorescens* SBW25. *BMC Microbiology*, 2008. **8**: p. 7.

140. Mossialos, D., U. Ochsner, C. Baysse, et al. Identification of new, conserved, non-ribosomal peptide synthetases from fluorescent pseudomonads involved in the biosynthesis of the siderophore pyoverdine. *Molecular Microbiology*, 2002. **45**: p. 1673 - 1685.
141. Horton, R., H. Hunt, S. Ho, et al. Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. *Gene*, 1989. **77**: p. 61 - 68.
142. King, E.O., M.K. Ward, and D.E. and Raney. Two simple media for the demonstration of pyocyanin and fluorescein. *Journal of Laboratory and Clinical Medicine*, 1954. **44**: p. 301-307.
143. Ackerley, D.F. Non-ribosomal peptide synthesis. PhD thesis, in School of biological sciences. 2001, University of Otago: Dunedin.
144. Lopez-Lopez, K., J.L. Hernandez-Flores, M. Cruz-Aguilar, et al. In *Pseudomonas syringae* pv. *Phaseolicola*, expression of the *argK* gene, encoding the phaseolotoxin-resistant ornithine carbamoyltransferase, is regulated indirectly by temperature and directly by a precursor resembling carbamoylphosphate. *Journal of Bacteriology*, 2004. **186**: p. 146-153.
145. Winsor, G.L., T. Van Rossum, R. Lo, et al. *Pseudomonas* genome database: Facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Research*, 2009. **37**: p. D483-488.
146. Mahé, B., C. Masclaux, L. Rauscher, et al. Differential expression of two siderophore-dependent iron-acquisition pathways in *Erwinia chrysanthemi* 3937: Characterization of a novel ferrisiderophore permease of the abc transporter family. *Molecular Microbiology*, 1995. **18**: p. 33-43.
147. Fourie, D. Characterization of halo blight races on dry beans in south africa. *Plant Disease*, 1998. **82**: p. 307-310.
148. Garner, B.L., J.E.L. Arceneaux, and B.R. Byers. Temperature control of a 3,4-dihydroxybenzoate (protocatechuate)-based siderophore in *Bacillus anthracis*. *Current Microbiology*, 2004. **49**: p. 89-94.
149. Colquhouna, D.J. and H. Sørumb. Temperature dependent siderophore production in *Vibrio salmonicida*. *Microbial Pathogenesis*, 2001. **31**: p. 213-219

150. Bachhawat, A.K. and S. Ghosh. Temperature inhibition of siderophore production in *Azospirillum brasilense*. Journal of Bacteriology, 1989. **171**: p. 4092-4094.
151. Bender, C.L., F. Alarcon-Chaidez, and D.C. Gross. *Pseudomonas syringae* phytotoxins: Mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiology and Molecular Biology Reviews, 1999. **63**: p. 266-292.
152. Expert, D., C. Enard, and C. Masclaux. The role of iron in plant host-pathogen interactions. Trends in Microbiology, 1996. **4**: p. 232-237.
153. Cody, Y. and D. Gross. Outer membrane protein mediating iron uptake via pyoverdinpss, the fluorescent siderophore produced by *Pseudomonas syringae* pv. *Syringae*. Journal of Bacteriology, 1987. **169**: p. 2207-2214.
154. Hirano, S.S. and C.D. Upper. Bacteria in the leaf ecosystem with emphasis on *pseudomonas syringae*: a pathogen, ice nucleus, and epiphyte. Microbiology and Molecular Biology Reviews, 2000. **64**: p. 624-653.
155. Visca, P., G. Colotti, L. Serino, et al. Metal regulation of siderophore synthesis in *pseudomonas aeruginosa* and functional effects of siderophore-metal complexes. Applied and Environmental Microbiology 1992. **58**: p. 2886-2893.
156. Sandra Matthijs, Kourosch Abbaspour Tehrani, George Laus, et al. Thioquinolobactin, a *Pseudomonas* siderophore with antifungal and anti-*pythium* activity. Environmental Microbiology, 2007. **9**: p. 425-434.
157. Guenzi, E., G. Galli, I. Grgurina, et al. Characterization of the syringomycin synthetase gene cluster: A link between prokaryotic and eukaryotic peptide synthetases. Journal of Biological Chemistry, 1998. **273**: p. 32857-32863.
158. Hengen, P.N. Purification of his-tag fusion proteins from *Escherichia coli*. Trends in Biochemical Sciences, 1995. **20**: p. 285-286.
159. Hochuli, E., W. Bannwarth, H. Döbeli, et al. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Nature Biotechnology, 1988. **6**: p. 1321-1325.
160. Couch, R., S.E. O'Connor, H. Seidle, et al. Characterization of CmaA, an adenylation-thiolation didomain enzyme involved in the biosynthesis of coronatine. J. Bacteriol., 2004. **186**: p. 35-42.

161. Sørensen, H. and K. Mortensen. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories*, 2005. **4**: p. 1.
162. Gasser, B., M. Saloheimo, U. Rinas, et al. Protein folding and conformational stress in microbial cells producing recombinant proteins: A host comparative overview. *Microbial Cell Factories*, 2008. **7**: p. 11.
163. Baneyx, F. and M. Mujacic. Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology*, 2004. **22**: p. 1399-1408.
164. Keating, T.A., Z. Suo, D.E. Ehmann, et al. Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry*, 2000. **39**: p. 2297-2306.
165. Dieckmann, R., Y.-O. Lee, H. van Liempt, et al. Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Letters*, 1995. **357**: p. 212-216.
166. Blackwell, J.R. and R. Horgan. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Letters*, 1991. **295**: p. 10-12.
167. Lee, S.G., F. Lipmann, and H.H. John, Tyrocidine synthetase system. *Methods in Enzymology*. 1975, Academic Press. p. 585-602.
168. Gonzalez, C.F., D.F. Ackerley, S.V. Lynch, et al. ChrR, a soluble quinone reductase of *Pseudomonas putida* that defends against H₂O₂. *Journal of Biological Chemistry*, 2005. **280**: p. 22590-22595.
169. Ehmann, D.E., C.A. Shaw-Reid, H.C. Losey, et al. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences U S A*, 2000. **97**: p. 2509-2514.
170. Reichert, J., M. Sakaitani, and C.T. Walsh. Characterization of EntF as a serine-activating enzyme. *Protein Science*, 1992. **1**: p. 549-556.
171. Geladopoulos, T.P., T.G. Sotiroudis, and A. Evangelopoulos. A malachite green colorimetric assay for protein phosphatase activity. *Analytical Biochemistry*, 1991. **192**: p. 112-116.
172. Mogk, A., M.P. Mayer, and E. Deuerling. Mechanisms of protein folding: Molecular chaperones and their application in biotechnology. *ChemBiochem*, 2002. **3**: p. 807-814.

173. Hartl, F.U. and M. Hayer-Hartl. Molecular chaperones in the cytosol: From nascent chain to folded protein. Science, 2002. **295**: p. 1852-1858.
174. Tsumoto, K., D. Ejima, I. Kumagai, et al. Practical considerations in refolding proteins from inclusion bodies. Protein Expression and Purification, 2003. **28**: p. 1-8.
175. Anfinsen, C.B. Principles that govern the folding of protein chains. Science, 1973. **181**: p. 223-230.
176. Oganessian, N., S.-H. Kim, and R. Kim. On-column protein refolding for crystallization. Journal of Structural and Functional Genomics, 2005. **6**: p. 177-182.
177. Li, M., Z.-G. Su, and J.-C. Janson. In vitro protein refolding by chromatographic procedures. Protein Expression and Purification, 2004. **33**: p. 1-10.
178. Studier, F.W. Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification, 2005. **41**: p. 207-234.
179. Cisar, J.S., J.A. Ferreras, R.K. Soni, et al. Exploiting ligand conformation in selective inhibition of non-ribosomal peptide synthetase amino acid adenylation with designed macrocyclic small molecules. Journal of the American Chemical Society, 2007. **129**: p. 7752-7753.
180. Finking, R., A. Neumüller, J. Solsbacher, et al. Aminoacyl adenylyate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases. Chembiochem, 2003. **4**: p. 903-906.
181. Kaplan, W. and T.G. Littlejohn. Swiss-pdb viewer (deep view). Briefings in Bioinformatics, 2001. **2**: p. 195-197.
182. de Crecy-Lagard, V., V. Blanc, P. Gil, et al. Pristinamycin I biosynthesis in streptomyces pristinaespiralis: Molecular characterization of the first two structural peptide synthetase genes. Journal of. Bacteriology, 1997. **179**: p. 705-713.
183. de Marco, A., L. Vigh, S. Diamant, et al. Native folding of aggregation-prone recombinant proteins in escherichia coli by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. Cell Stress and Chaperones, 2005. **10**: p. 329-39.
184. Baneyx, F. Recombinant protein expression in Escherichia coli. Current Opinion in Biotechnology, 1999. **10**: p. 411-421.

185. Martínez-Alonso, M., N. González-Montalbán, E. García-Fruitós, et al. Learning about protein solubility from bacterial inclusion bodies. *Microbial Cell Factories*, 2009. **8**: p. 4.
186. Gonzalez-Montalban, N., E. Garcia-Fruitos, and A. Villaverde. Recombinant protein solubility-does more mean better? *Nature Biotechnology*, 2007. **25**: p. 718-720.
187. González-Montalbán, N., E. García-Fruitós, S. Ventura, et al. The chaperone Dnak controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells. *Microbial Cell Factories*, 2006. **5**: p. 26.
188. Christian, P.S. Illuminating folding intermediates. *Nature Structural Biology*, 2000. **7-10**: p. 7.
189. Hoang, T.T., A.J. Kutchma, A. Becher, et al. Integration-proficient plasmids for pseudomonas aeruginosa: Site-specific integration and use for engineering of reporter and expression strains. *Plasmid*, 2000. **43**: p. 59-72.
190. Meyer, J., A. Stintzi, D. De Vos, et al. Use of siderophores to type pseudomonads: The three pseudomonas aeruginosa pyoverdine systems. *Microbiology*, 1997. **143**: p. 35 - 43.
191. Abdallah, M. Pyoverdines and pseudobactins, in *CRC handbook of microbial iron chelates*, 1991. p. 139 - 152.
192. Frueh, D.P., H. Arthanari, A. Koglin, et al. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature*, 2008. **454**: p. 903-906.
193. Lai, J.R., A. Koglin, and C.T. Walsh. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry*, 2006. **45**: p. 14869-14879.
194. Stein, T. Whole-cell matrix-assisted laser desorption/ionization mass spectrometry for rapid identification of bacteriocin/lantibiotic-producing bacteria. *Rapid Communications in Mass Spectrometry*, 2008. **22**: p. 1146-1152.
195. Neuhof, T., R. Dieckmann, I.S. Druzhinina, et al. Intact-cell MALDI-TOF mass spectrometry analysis of peptaibol formation by the genus *trichoderma/hypocrea*: Can molecular phylogeny of species predict peptaibol structures? *Microbiology*, 2007. **153**: p. 3417-3437.

196. Vater, J., C. Wilde, and H. Kell. In situ detection of the intermediates in the biosynthesis of surfactin, a lipopeptide from *Bacillus subtilis* okb 105, by whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in combination with mutant analysis. Rapid Communications in Mass Spectrometry, 2009. **23**: p. 1493-1498.
197. Hicks, L., P. Weinreb, D. Konz, et al. Fourier-transform mass spectrometry for detection of thioester-bound intermediates in unfractionated proteolytic mixtures of 80 and 191 kDa portions of bacitracin synthetase. Analytica Chimica Acta, 2003. **496**: p. 217-224.
198. Reverchon, S., C. Rouanet, D. Expert, et al. Characterization of indigoidine biosynthetic genes in *Erwinia chrysanthemi* and role of this blue pigment in pathogenicity. Journal of Bacteriology, 2002. **184**: p. 654-665.
199. Kim, C.-Y., V.Y. Alekseyev, A.Y. Chen, et al. Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. Biochemistry, 2004. **43**: p. 13892-13898.
200. Wu, N., D.E. Cane, and C. Khosla. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. Biochemistry, 2002. **41**: p. 5056-5066.
201. Wu, N., S.Y. Tsuji, D.E. Cane, et al. Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. Journal of the American Chemical Society, 2001. **123**: p. 6465-6474.
202. Gross, F., D. Gottschalk, and R. Müller. Posttranslational modification of myxobacterial carrier protein domains in *Pseudomonas* sp. by an intrinsic phosphopantetheinyl transferase. Applied Microbiology and Biotechnology, 2005. **68**: p. 66-74.
203. Seidle, H.F., R.D. Couch, and R.J. Parry. Characterization of a nonspecific phosphopantetheinyl transferase from *Pseudomonas syringae* pv. *Syringae* ff5. Archives of Biochemistry and Biophysics, 2006. **446**: p. 167-174.
204. Novagen, PET system manual, 10th edition. 2003.

205. Tabor, S. and C.C. Richardson. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the U S A*, 1985. **82**: p. 1074-1078.
206. Kuhn, R., M.P. Starr, D.A. Kuhn, et al. Indigoidine and other bacterial pigments related to 3,3'-bipyridyl. *Archives of Microbiology*, 1965. **51**: p. 71-84.
207. Gulick, A.M. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chemical Biology*, 2009. **4**: p. 811-827.
208. Du, L., Hybrid peptide-polyketide natural products, in *Genetic engineering: Principles and methods*. 2003. p. 231-236.
209. Daniel, R. The metagenomics of soil. *Nature reviews microbiology*, 2005. **3**: p. 470-478.
210. Schwarzer, D., H.D. Mootz, and M.A. Marahiel. Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chemistry and Biology*, 2001. **8**: p. 997-1010.
211. Yasgar, A., T.L. Foley, A. Jadhav, et al. A strategy to discover inhibitors of bacillus subtilis surfactin-type phosphopantetheinyl transferase. *Molecular biosystems*, 2010. **6**: p. 365-375.
212. Schujman, G.E. and D. de Mendoza. Regulation of type ii fatty acid synthase in gram-positive bacteria. *Current Opinion in Microbiology*, 2008. **11**: p. 148-152.
213. Ying-Jie, L., Z. Yong-Mei, and O.R. Charles, Product diversity and regulation of type II fatty acid synthases. *Biochemistry and Cell Biology*, 2004. **82**: p. 145-156.