

**Genetic manipulation of pyoverdine non-ribosomal  
peptide synthetases to identify genetic constraints to  
effective domain recombination**

By

Mark Jonathan Calcott

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  
(2014)



## Abstract

Non-ribosomal peptide synthetases (NRPSs) synthesise small highly diverse peptides with a wide range of activities, such as antibiotics, anticancer drugs, and immunosuppressants. NRPS synthesis often resembles an assembly line, in which each module acts in a linear order to add one monomer to the growing peptide chain. In the basic mechanism of synthesis, an adenylation (A) domain within each module activates a specific monomer. Once activated, the monomer is attached to an immediately downstream thiolation (T) domain via a prosthetic phosphopantetheine group, which acts as a flexible arm to pass the substrate between catalytic domains. A condensation (C) domain, upstream to the A-T domains, catalyses peptide bond formation between an acceptor substrate attached to the T domain and a donor substrate attached to the T domain of the upstream module. The peptide remains attached to the T domain of the acceptor substrate, and then acts as the donor substrate for the next C domain. When peptide synthesis reaches the final module, the peptide is released by a thioesterase (TE) domain.

The linear mode of synthesis and discrete functional domains within each module gives the potential to generate new products by substituting domains or entire modules with ones that activate alternative substrates. Attempts to create new products using domain and module substitution often result in a loss of activity. The work in this thesis focuses on identifying barriers to effective domain substitution. The NRPS enzyme *pvdD*, which adds the final residue to the eleven residue non-ribosomal peptide pyoverdine, was developed as a model for domain substitution. The primary benefit for using this model is that pyoverdine creates easily detectable fluorescent products.

The first set of experiments focused on testing the limitations of A domain and C-A domain substitutions to alter pyoverdine. Nine A domain and nine C-A domain substitution *pvdD* variants were constructed and used to complement a *P. aeruginosa* PAO1 *pvdD* deletion strain. The A domain substitutions that specified the wild type substrate were highly functional, whereas A domains that specified other substrates resulted in low levels of wild type pyoverdine production. This suggests the acceptor site substrate specificity of the C domain limited the success of A domain substitutions, rather than disruption of the C/A domain junction. In contrast, although C-A domain substitutions in *pvdD* in some cases synthesised novel pyoverdines, the majority lost function for unknown reasons. The high

success rate A domain substitutions (when not limited by the acceptor site specificity of the C domain) suggested the addition of new C domains was a likely cause for loss of function.

The second set of experiments investigated whether disrupting the protein interface between C domains and their upstream T domains may cause a loss in function of C-A domain substitutions. However, domain substitutions of T domains were found to have a high rate of success. Therefore, the results thus far confirmed that disrupting interactions of the C domain with A domains or T domains does not have a large affect on enzyme activity.

An alternative explanation for the loss in function with C-A domain substitutions is that C domains translocated to a new enzyme are unable to process the new incoming donor peptide chain because of substrate specificity or steric constraints. To develop methods to circumvent limitations caused by the C domain, the final part of this thesis examined acceptor substrate specificity of C domains. Acceptor site substrate specificity was chosen over donor site specificity as it acts on only an amino acid rather than peptide chain. The substrate specificity was narrowed down to a small subsection of the C domain. This was an initial study of C domain substrate specificity, which may guide future development of relaxed specificity C domains.

## **Acknowledgments**

I offer my sincere appreciation to my supervisor Assoc. Prof. David Ackerley for the assistance he provided me during the course of my thesis, and for creating a friendly lab environment that was enjoyable to work in. I am also thankful to all members of the Ackerley lab who provided me with assistance and support when needed.

I would like to thank my secondary supervisor Assoc. Prof. Bill Jordan for access to laboratory equipment and helpful career related advice. For assistance with mass spectrometry, I wish to acknowledge the help from Dr Jonathan Dunne and Danyl McLauchlan.

Thank you to Prof. Iain Lamont for his collaboration on this project and providing feedback on research, as well as Lois Martin for assistance with pyoverdine purification and analysis.

Finally, I wish to acknowledge the funding I received during my studies. This work would not have been possible without support from the Royal Society of New Zealand Marsden Fund and the Victoria University Research Fund.

## Table of Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgments</b>	<b>iii</b>
<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xiv</b>
<b>List of abbreviations</b>	<b>xvi</b>
<b>Chapter 1:</b>	<b>Introduction .....1</b>
1.1	General overview .....1
1.2	Multiple template model of non-ribosomal peptide synthesis .....3
1.2.1	Adenylation domains .....5
1.2.1.1	A domain structure and specificity.....6
1.2.2	Thiolation domains .....7
1.2.2.1	T domain structure .....8
1.2.3	Condensation domains .....9
1.2.3.1	C domain structure.....10
1.2.3.2	C domain substrate specificity .....11
1.2.3.3	Control of condensation timing.....13
1.2.3.4	C domain specialisation.....14
1.2.4	Thioesterase domains .....15
1.2.4.1	Structure of TE domains.....16
1.2.5	Type II thioesterase enzymes .....17
1.2.6	Accessory domains .....17
1.3	Engineering of novel NRPS enzymes .....18
1.3.1	Studies on A and A-T domain substitution.....18
1.3.2	Alteration of A domain binding pockets.....21
1.3.3	C-A domain substitution .....22
1.3.4	Inter-enzyme communication .....24
1.3.5	Conclusions from previous domain substitutions efforts.....25

1.4	Pyoverdine as an alternative system to study domain substitution .....	26
1.4.1	Reasons to use pyoverdine as a model for domain substitution .....	29
1.4.2	Previous work on the pyoverdine system .....	30
1.5	Aims of this study .....	31
<b>Chapter 2:</b>	<b>Methods.....</b>	<b>34</b>
2.1	Chemical, reagents, media and enzymes.....	34
2.2	Oligonucleotide primers .....	34
2.3	Plasmids .....	37
2.4	Bacterial strains .....	37
2.5	Bacterial growth and maintenance .....	38
2.5.1	Liquid media .....	38
2.5.2	Solid media .....	38
2.5.3	Media supplements .....	39
2.5.4	Growth and storage .....	39
2.6	General molecular biology methods .....	39
2.6.1	Isolation and purification of DNA .....	39
2.6.2	Confirmation of DNA products .....	40
2.6.2.1	Agarose gel electrophoresis.....	40
2.6.2.2	DNA quantification.....	40
2.6.2.3	DNA Sequencing.....	40
2.6.3	PCR protocols .....	40
2.6.3.1	Standard reactions .....	40
2.6.3.2	Overlap Extension PCR .....	41
2.6.3.3	Site-directed mutagenesis .....	42
2.6.3.4	STaggered Elongation Process PCR .....	42
2.6.4	Restriction enzyme digest .....	43
2.6.5	Ligation .....	43
2.7	Bacterial transformation.....	43
2.7.1	<i>E. coli</i> transformation by the calcium chloride method.....	43

2.7.2	<i>P. aeruginosa</i> transformation by electroporation .....	44
2.8	Creation of the domain substitution plasmids .....	45
2.8.1	Creation of the plasmids pSMC and pSMA.....	45
2.8.2	Creation of domain substitution plasmids for Chapter 5 .....	46
2.8.2.1	Creation of the plasmid pST2 .....	46
2.8.2.2	Creation of the plasmid pTCA.....	46
2.8.2.3	Creation of pST1 .....	47
2.8.3	Creation of the directed evolution plasmid and loop substitution for Chapter 6 .....	47
2.8.3.1	Creation of the plasmid pDEC .....	47
2.8.3.2	Creation of the chimeric C domain constructs.....	48
2.8.3.3	Creation of the loop substitution construct.....	48
2.8.3.4	Construction of the plasmid pTRN .....	48
2.9	Analysis of pyoverdine production by recombinant and WT <i>P. aeruginosa</i> PAO1 strains .....	49
2.9.1	Assessment of strains grown on KB agar plates .....	49
2.9.2	Pyoverdine synthesis in liquid media.....	49
2.9.3	MALDI-TOF/TOF mass spectrometry .....	49
2.9.4	Screening the StEP library .....	50
2.10	Bioinformatics .....	50
2.10.1	Prediction of NRPS substrate specificity .....	50
2.10.2	Multiple sequence alignments and phylogenies.....	50
2.10.3	Pairwise sequence alignment .....	51
2.10.4	Structural model creation and selection .....	51
2.10.4.1	Automated modelling servers .....	51
2.10.4.2	Model creation using SWISS-MODEL .....	51
2.10.4.3	Model creation using Modeller 9.11 .....	51
2.10.5	Viewing of homology and structural models .....	52

**Chapter 3: Construction and characterisation of *pvdD* A domain substitution variants 53**

3.1	Introduction .....	53
3.2	Construction of the A domain substitution staging plasmid and A domain substitution variants .....	54
3.3	Evaluation of pyoverdine synthesis by A domain substitution strains .....	58
3.3.1	Growth of A domain substitution strains on solid KB media.....	58
3.3.2	Pyoverdine production by A domain substitution strains grown in liquid media	59
3.4	MALDI-TOF MS/MS confirmation of pyoverdine production.....	61
3.4.1	Pyoverdine produced by Thr specifying A domain substitutions.....	61
3.4.2	Pyoverdine produced by non-Thr specifying A domain substitutions .....	64
3.5	Discussion .....	67
<b>Chapter 4:</b>	<b>Substitution of C-A domains in <i>pvdD</i> .....</b>	<b>71</b>
4.1	Introduction .....	71
4.2	Creation of C-A domain substitution strains.....	72
4.3	Evaluating levels of pyoverdine synthesis .....	73
4.3.1	Fluorescence of C-A domain substitution strains on solidified KB media..	73
4.3.2	Pyoverdine production of C-A domain substitution strains in liquid media	74
4.4	MALDI-TOF MS/MS confirmation of pyoverdine production.....	75
4.4.1	Pyoverdine produced by highly functional C-A domain substitutions.....	75
4.4.2	Pyoverdine produced by C-A domain substitutions with limited activity...	77
4.5	Discussion .....	80
<b>Chapter 5:</b>	<b>Investigation of the influence of T domains on the activity of C-A domain substitutions.....</b>	<b>85</b>
5.1	Introduction .....	85
5.1.1	Additional nomenclature.....	85
5.1.2	Evidence of T domain interactions being inhibited by domain substitution	86
5.1.2.1	T domain substitutions and evolution in BpsA .....	88
5.2	Chapter overview and aims .....	91

5.3	Substitution of evolved <i>pvdD</i> T domains back into <i>pvdD</i> .....	92
5.3.1	Growth on solid media of T2 domain substitution strains .....	93
5.3.2	Pyoverdine production in liquid media by T2 domain substitution strains ..	94
5.3.3	Mass spectrometry of T2 domain substitution strains .....	95
5.4	Bioinformatics of T domains.....	95
5.5	Attempts to avoid disruption of the T/C interface via T-C-A domain substitutions .....	97
5.5.1	Growth on solid media of T-C-A and C-A domain substitution strains ....	100
5.5.2	Growth in liquid media of T-C-A and C-A domain substitution strains ...	101
5.5.3	Mass spectrometry of T-C-A and C-A domain substitution strains.....	101
5.6	Analysing the effect of disrupting the T-C junction via T domain substitutions	104
5.6.1	Creation of T domain substitution variants.....	105
5.6.2	Pyoverdine production by T domain substitution strains.....	107
5.6.2.1	Growth of T domain substitution strains on solid media .....	107
5.6.2.2	Growth in liquid media of T1 domain substitution strains.....	108
5.6.3	Mass spectrometry of T domain substitution strains .....	110
5.6.4	Amino acid sequence alignments of T domains .....	111
5.7	Discussion .....	113
<b>Chapter 6:</b>	<b>Investigation of the acceptor site substrate specificity of a</b>	
<b>C domain</b>	<b>117</b>	
6.1	Introduction .....	117
6.2	Prediction of the residues most likely to be involved in substrate specificity ..	119
6.2.1	Sequence comparison of JC <sub>Lys</sub> and DC <sub>Thr</sub> .....	119
6.2.1	Structural comparison of variable regions of JC <sub>Lys</sub> and DC <sub>Thr</sub> .....	119
6.2.1.1	Creation and selection of a structural model of DC <sub>Thr</sub> .....	120
6.2.1.2	Analysis of non-identical residues between JC <sub>Lys</sub> and DC <sub>Thr</sub> .....	124
6.2.1.3	Analysis of residues within 15 Å of the conserved catalytic His residue.....	125
6.3	Probing the role of residues close to the catalytic His residue .....	126

6.3.1	Testing the function of chimeric C domains.....	127
6.3.1.1	Creation of a vector suitable for directed evolution.....	127
6.3.1.2	Creation of chimeric C domains .....	128
6.3.1.3	Analysis of the strains expressing chimeric C domain constructs on solid media and in liquid media.....	130
6.3.1.4	MALDI-TOF detection of pyoverdine from chimeric C domain strains.....	132
6.3.2	Altering the loop covering the active site of the C domain .....	133
6.3.2.1	Testing pyoverdine production by the C domain loop substitution strain .....	134
6.3.3	Recombination of the third variable region of JC <sub>Lys</sub> and DC <sub>Thr</sub> .....	135
6.3.3.1	Evaluation of the J12-D3 chimeric C domain strain .....	139
6.4	Discussion .....	141
<b>Chapter 7:</b>	<b>Key findings and future directions .....</b>	<b>147</b>
7.1	Research motivation.....	147
7.2	Key findings .....	147
7.3	Limitations of this research.....	151
7.3.1	Limitations in the use of pyoverdine NRPS enzymes .....	151
7.3.2	Limitations in testing pyoverdine production.....	152
7.4	Future directions.....	153
7.4.1	Improve the activity of A domain substitutions.....	153
7.4.2	Perform T1 domain substitutions of T <sub>C</sub> domains from non-pyoverdine NRPS enzymes .....	154
7.4.3	Identify residues involved in acceptor site specificity of the C domain....	155
7.4.4	Characterise the donor site constraints of the C domain .....	156
7.4.5	Development of large scale screens.....	156
7.5	Concluding remarks .....	157
<b>Appendix A.</b>	<b>Accession numbers of NRPS genes used as templates for domain substitution, and the substrate specificity of key A domains.....</b>	<b>159</b>
<b>Appendix B.</b>	<b>Optimisation of liquid media based assay for measuring pyoverdine levels</b>	<b>161</b>

<b>Appendix C.</b>	<b>Alignments of T domains from non-<i>Pseudomonas</i> strains</b>
<b>163</b>	
<b>Appendix D.</b>	<b>Comparative modelling quality assessment.....167</b>
<b>Appendix E.</b>	<b>Pyoverdine production by strains containing the directed</b>
<b>evolution construct</b>	<b>171</b>
<b>References</b>	<b>173</b>

## List of figures

Figure 1-1. Examples of non-ribosomal peptides. ....	2
Figure 1-2. The domain architecture and mechanism of synthesis for a hypothetical NRPS. ....	4
Figure 1-3. Two step adenylation domain reaction. ....	5
Figure 1-4. Peptide bond formation catalysed by the C domain reaction.....	9
Figure 1-5. Formation of a DKP product using PheATE and ProCAT.....	12
Figure 1-6. TE domain reactions. ....	16
Figure 1-7. Pyoverdine from <i>P. aeruginosa</i> PAO1 .....	26
Figure 1-8. Synthesis of the pyoverdine precursor. ....	28
Figure 3-1. Domain arrangement within substitution plasmids, and plasmid map of pSW196. ....	55
Figure 3-2. Location the of upstream restriction site for pSMC.....	56
Figure 3-3. NRPS genes with modules used for domain substitution highlighted.....	57
Figure 3-4. Fluorescence of A domain substitutions strains on KB agar media in the absence (-) or presence (+) of EDDHA. ....	59
Figure 3-5. Percentage of pyoverdine production from A domain substitution strains grown in liquid media. ....	60
Figure 3-6. Mass spectra obtained from the supernatant of each of the Thr specifying A domain substitution strains and Thr-WT .....	61
Figure 3-7. CID spectra of pyoverdine synthesised by Thr-WT and the three Thr specifying A domain substitution strains.....	62
Figure 3-8. Nomenclature for pyoverdine ions formed by CID of pyoverdine .....	63
Figure 3-9. MALDI-TOF spectra obtained using the supernatant of non-Thr specifying A domain substitution strains. ....	65
Figure 3-10. CID spectra of pyoverdine synthesised by non-Thr specifying A domain substitution strains. ....	66
Figure 4-1. Fluorescence of C-A domain substitution strains on KB agar plates .....	73
Figure 4-2. Liquid media assays of relative levels of pyoverdine production by C-A domain substitution strains. ....	74
Figure 4-3. MALDI-TOF spectra obtained using the supernatant from functional C-A domain substitution variants. ....	75

Figure 4-4. Fragmentation spectra of the main pyoverdine species from functional C-A domain substitution strains.....	76
Figure 4-5. MALDI-TOF spectra obtained using the supernatant from functional C-A domain substitution variants with low levels or no fluorescence. ....	78
Figure 4-6. CID spectra of truncated pyoverdines obtained from C-A domain substitution strains. ....	79
Figure 5-1. Nomenclature for T domain types and interactions shown for a hypothetical three module NRPS.....	86
Figure 5-2. BpsA and indigoidine synthesis. ....	89
Figure 5-3. The domain organisation within the substitution plasmid pST2.....	93
Figure 5-4. Fluorescence of T2 domain substitution strains on KB agar media in the (A) absence or (B) presence of EDDHA. ....	93
Figure 5-5. Pyoverdine production of T2 domain substitutions detected when grown in liquid media.....	94
Figure 5-6. Mass analysis of T2 domain substitution strains.....	95
Figure 5-7. Sequence logos of T domains from bacterial strains other than pseudomonads .....	96
Figure 5-8. Domain organisation within the T-C-A domain substitution plasmid. ....	99
Figure 5-9. Location of domains used to source T-C-A domains for substitution. ....	99
Figure 5-10. Growth of C-A versus T-C-A domain substitution strains on solid media. ....	100
Figure 5-11. Pyoverdine production in liquid media from T-C-A versus C-A domain substitution strains.....	101
Figure 5-12. Mass spectra obtained from the supernatant of each of the T-C-A and C-A domain substitution strains.....	102
Figure 5-13. CID spectra of the pyoverdine synthesised by the strains TCAhfOrn-J and CAhfOrn-J.....	103
Figure 5-14. Comparison of T domain substitutions to C-A domain substitutions. ....	105
Figure 5-15. Location of T domains in <i>Pseudomonas</i> NRPS pathways used for T1 domain substitution. ....	106
Figure 5-16. Growth of T1 domain substitution strains on solid media .....	107
Figure 5-17. Pyoverdine production from T1 domain substitution strains grown in liquid media.....	109
Figure 5-18. Mass spectra obtained from the supernatant of each of the T1 domain substitution strains.....	111

Figure 5-19. Alignment of T domains substituted into the first module of <i>pvdD</i> . ....	112
Figure 6-1. Nucleotide sequence alignment of JC <sub>Lys</sub> and DC <sub>Thr</sub> . ....	120
Figure 6-2. Analysis of the structural model RaptorXmsa. ....	123
Figure 6-3. Residues differing between JC <sub>Lys</sub> and DC <sub>Thr</sub> mapped onto the model of DC <sub>Thr</sub> . .....	124
Figure 6-4. Residues non-identical to JC <sub>Lys</sub> and within 15 Å of the catalytic His residue of DC <sub>Thr</sub> . ....	126
Figure 6-5. Domain architecture within the plasmid pDEC. ....	128
Figure 6-6. Domain architecture of <i>pvdD</i> and the chimeric C domains created to replace DC <sub>Thr</sub> . ....	129
Figure 6-7. Pyoverdine production by chimeric C domain strains measured on solid media and in liquid media. ....	131
Figure 6-8. Mass spectrometry of the supernatant from strains expressing <i>pvdD</i> variants that contain chimeric C domains. ....	133
Figure 6-9. Analysis of pyoverdine production by the loop substitution strain. ....	134
Figure 6-10. Fluorescence based screening of StEP library. ....	136
Figure 6-11. Fluorescence of strains that were sequenced from the StEP library. ....	137
Figure 6-12. Sequence alignments of the third variable region from strains that were highly or moderately active in pyoverdine synthesis. ....	139
Figure 6-13. Analysis of pyoverdine synthesised by D12-J3 compared with J12-D3. ...	141
Figure 6-14. Location of residues potentially important for high levels of activity mapped onto 2VSQ .....	144
Figure B-1. Fluorescence (left) and absorbance (right) of the culture and supernatant of A domain substitution strains .....	161
Figure B-2. Absorbance and fluorescence taken from a 2x serial dilution of the supernatant from strain Thr-WT. ....	162
Figure C-3. ClustalW sequence alignments of T domains from non- <i>Pseudomonas</i> NRPS enzymes. ....	165
Figure E-4. L-arabinose induction of pyoverdine synthesis for the pDEC restriction site control strain. ....	171
Figure E-5. L-arabinose induction of pyoverdine synthesis for the chimeric C domain strains from Chapter 6. ....	172

## List of tables

Table 2-1. Oligonucleotide primers used in this study .....	35
Table 2-2. Plasmids used in this study .....	37
Table 2-3. <i>E. coli</i> strains used in this study .....	37
Table 2-4. <i>Pseudomonas</i> strains used in this study* .....	37
Table 2-5. Concentrations of antibiotics in media .....	39
Table 2-6. Kits used for DNA purification .....	40
Table 2-7. PCR protocol used for Phusion™ polymerase .....	41
Table 2-8. PCR protocol used for Biomix Red™ polymerase.....	41
Table 2-9. Overlap extension protocol used for Biomix Red™ polymerase.....	42
Table 2-10. Protocol for StEP PCR recombination .....	43
Table 2-11. Automated structural modelling servers.....	52
Table 3-1. Ions detected by CID of pyoverdine from Thr specifying A domain substitution strains.....	63
Table 3-2. Ions detected by CID of pyoverdine from non-Thr specifying A domain substitution strains.....	66
Table 3-3. Amino acid sequence identity of introduced A domains with the module 2 PvdD A domain that was substituted. ....	68
Table 4-1. Ions detected by CID of major pyoverdine species produced from highly functional C-A domain substitution strains. ....	76
Table 4-2. CID of truncated pyoverdine produced by C-A domain substitution strains. ..	79
Table 5-1. <i>In vivo</i> synthesis of indigoidine by BpsA variants. ....	90
Table 5-2. Residues found at the +4 and +24 positions of <i>Pseudomonas</i> pyoverdine NRPS T domains relative to the 4'-pp attachment site. ....	91
Table 5-3. Ions detected by CID of the pyoverdine synthesised by the strains TCAhfOrn-J and CAhfOrn-J.....	103
Table A-1. Accession numbers of genes used for domain substitution .....	159
Table A-2. Residues lining the A domain binding pockets and substrate specificity predictions.....	160
Table C-3: Accession numbers of enzymes from which T domain sequences were sourced. ....	164
Table D-4. Ranking of comparative structural models of the Thr-WT C domain by QMEAN6.....	167

Table D-5. Ranking of comparative structural models of the Thr-WT C domain by QMEANclust. ....	169
--	-----

## List of abbreviations

3-HBA	3-hydroxybenzoic acid
3mGlu	3-methyl-glutamic acid
4'-pp	4'-phosphopantetheine
A domain	adenylation domain
SNAC	<i>N</i> -acetylcysteamine thioester
ArCP	aryl-carrier protein
ASA	aspartate $\beta$ -semialdehyde
bp	basepair
C domain	condensation domain
CAS	chrome azurol S
CAT	chloramphenicol acetyltransferases
CHCA	$\alpha$ -Cyano-4-hydroxycinnamic acid
CID	collision induced dissociation
CoA	Coenzyme A
COM domain	communication-mediating domain
Cy domain	cyclisation domain
Dab	diaminobutyrate
DC <sub>Thr</sub>	The C domain from the second module of <i>pvdD</i>
DHB	2,3-dihydroxybenzoic acid
DKP	diketopiperazine
E domain	epimerisation domain
E2p	dihydrolipoyl transacetylases
EDDHA	ethylenediamine-N,N'-bis(2-hydroxyphenylacetic acid)
EDTA	ethylenediaminetetraacetic acid
FACS	fluorescence activated cell sorting
HDTMA	hexadecyl(trimethyl)azanium chloride
IPTG	isopropyl $\beta$ -D-1-thiogalactoside
JC <sub>Lys</sub>	The C domain from the first module of <i>pvdD</i>
kb	kilobase
KB media	King's B media
kyn	kynurenine

LB	Luria broth
MALDI	matrix-assisted laser desorption/ionisation
hfOrn	<i>N</i> <sup>5</sup> -formyl- <i>N</i> <sup>5</sup> -hydroxyornithine
NRPS	non-ribosomal peptide synthetase
Orn	ornithine
PCR	polymerase chain reaction
PheATE	phenylalanine specifying initiation module containing A-T-E domains
PP <sub>i</sub>	pyrophosphate
PPTase	phosphopantetheine transferase
ProCAT	Pro specifying elongation module containing C-A-T domains
Suca	succinamide
T domain	thiolation domain
T <sub>C</sub> domain	T domain with a C domain immediately downstream
T <sub>Ct</sub> domain	T domain with a C domain <i>in trans</i> downstream
T <sub>E</sub> domain	T domain with an E domain immediately downstream
TE domain	thioesterase domain
TE <sub>II</sub>	Type II thioesterase
T <sub>TE</sub> domain	T domain with a TE domain immediately downstream
TOF	time of flight
UV	ultra-violet
v/v	volume per volume
w/v	weight per volume



## Chapter 1: **Introduction**

### 1.1 **General overview**

Non-ribosomal peptides are a class of natural products synthesised mainly by bacteria and fungi. They are generally very small peptide or mixed-peptide products; the largest non-ribosomal peptide to have been identified to date contains 49 residues, with all others known containing under 23 residues, and more commonly under ten residues (Caboche et al., 2010). Despite their small size, they are highly diverse in terms of the monomers that can be incorporated. As of 2010, over 500 unique monomers have been detected in non-ribosomal peptides (Caboche et al., 2010), including both proteinogenic and non-proteinogenic D- and L- amino acids, carboxylic acids and amines (Marahiel et al., 1997; Sieber and Marahiel, 2005). In addition to substrate diversity, non-ribosomal peptides exhibit high structural diversity with only 27 % being linear; the remainder having cyclic, branched or other complex primary structures (Caboche et al., 2010). Figure 1-1 highlights this diversity by showing a number of peptides relevant to this introduction.

The diversity of non-ribosomal peptides imparts on them many useful bioactivities, and peptides have been identified with antibiotic, anticancer, anti-inflammatory, immunosuppressant and antiviral activities (Felnagle et al., 2008; Marahiel et al., 1997; Sieber and Marahiel, 2005). Often natural products need to be modified to improve clinical properties and/or bypass resistance mechanisms (Bush, 2012; O'Connell et al., 2013). Due to difficulty in their synthesis, most clinical natural product derivatives are created by means of semisynthesis; a process whereby the natural product is chemically modified after isolation from fermentation or biological sources (Kirschning and Hahn, 2012; O'Connell et al., 2013; Walsh, 2003). An alternative that would open up a range of structural diversity is protein engineering to create combinatorial libraries of existing natural products. For non-ribosomal peptides, their modular mode of synthesis, which will be discussed below, makes them potentially amenable to rational manipulation at the genetic level. However, modifications often greatly impair activity, thereby limiting the ability to create new non-ribosomal peptides using protein engineering. The work in this thesis focuses on identifying barriers to effective domain recombination for the creation of novel non-ribosomal peptides.

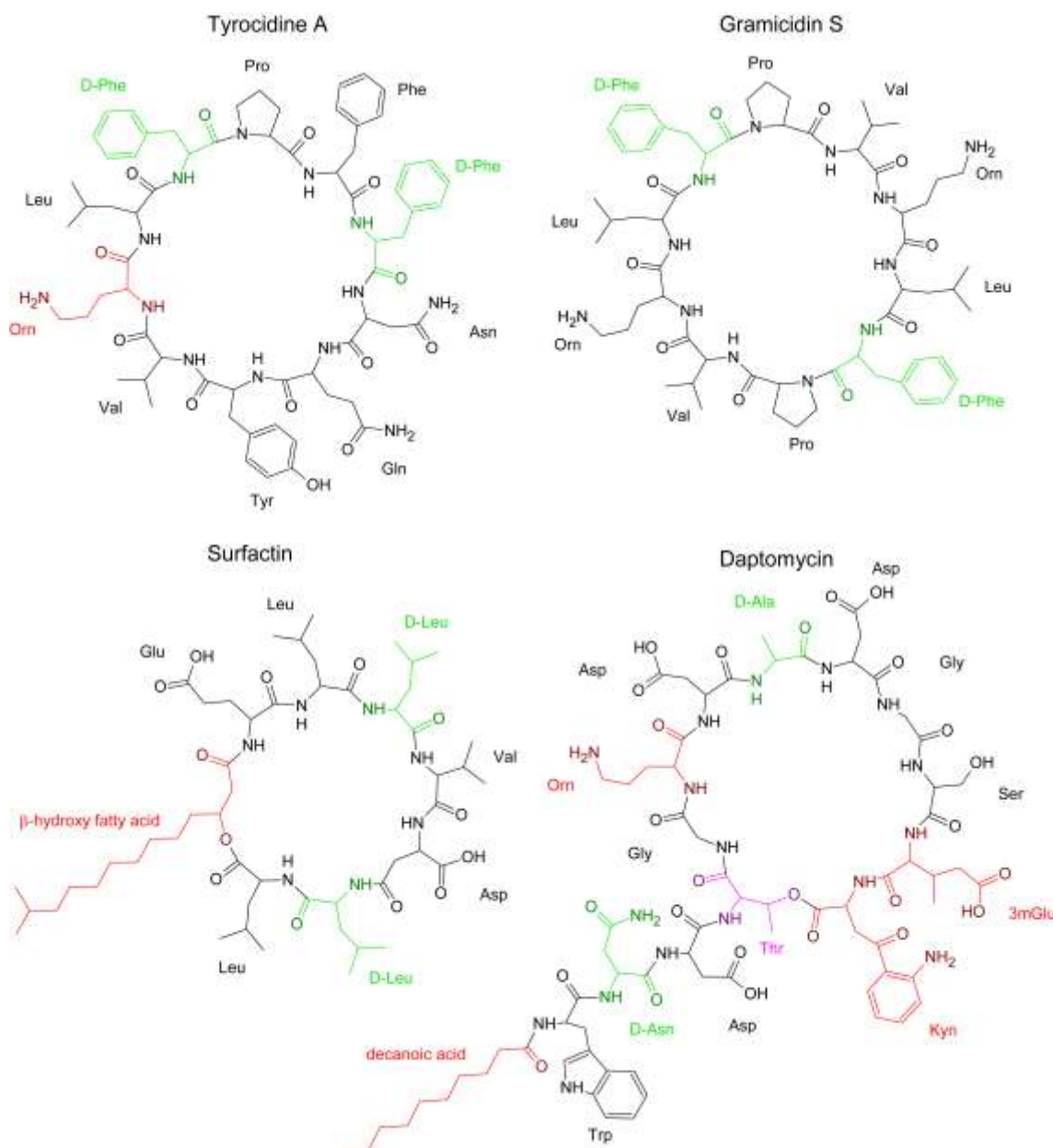


Figure 1-1. Examples of non-ribosomal peptides. Non-proteinogenic amino acids are highlighted in red, D-isomers in green, and an atypical peptide bond formed via a side-chain shown in purple. Orn, ornithine; 3mGlu, 3-methylglutamic acid; Kyn, kynurenine; standard three letter abbreviations are used for proteinogenic amino acids. Structures generated based on diagrams from Doekel *et al.* (2008) and Marahiel *et al.* (1997).

## 1.2 Multiple template model of non-ribosomal peptide synthesis

The accepted mode of non-ribosomal peptide synthesis is referred to as the multiple template model and was proposed by Stein et al. (Stein et al., 1994, 1996). A main characteristic of this model is that peptides are synthesised in a modular assembly line-like manner by non-ribosomal peptide synthetase (NRPS) enzymes. An NRPS pathway may be on one or multiple enzymes, and modules are classified as initiation, elongation and termination modules depending on their location. Modules are defined by their ability to recognise, activate and incorporate one monomer into the final peptide product (von Döhren et al., 1997; Marahiel et al., 1997).

In synthesis, an initiation module activates the first monomer and passes it to the immediately downstream elongation module. The elongation module activates a second monomer, condenses it with the monomer received from the upstream module and passes the resulting dipeptide to the next module downstream. In this way, each additional elongation module adds one additional residue. The final module is referred to as a termination module, and condenses one further monomer onto the peptide before releasing the final product (Hur et al., 2012; Sieber and Marahiel, 2005). In the modular system just described the modules are colinear with the peptide product, such that the module location dictates the position of the monomer in the final peptide (Marahiel et al., 1997). These co-linear NRPS pathways are referred to as Type A NRPS enzymes. Alternatively, iterative Type B pathways are similar, but differ in that the pathway is iterated and two to three independently-assembled smaller peptides are joined together to yield the final product. In contrast, non-linear Type C pathways are different as the order of modules in the NRPS complex does not represent the final product (Mootz et al., 2002).

Within each module of the NRPS assembly lines, discrete adenylation (A), thiolation (T) and condensation (C) domains function in the steps of substrate activation, transfer and condensation. These domains and their structural and functional independence from each other are central to this study, and are described in Figure 1-2 and the following sections.

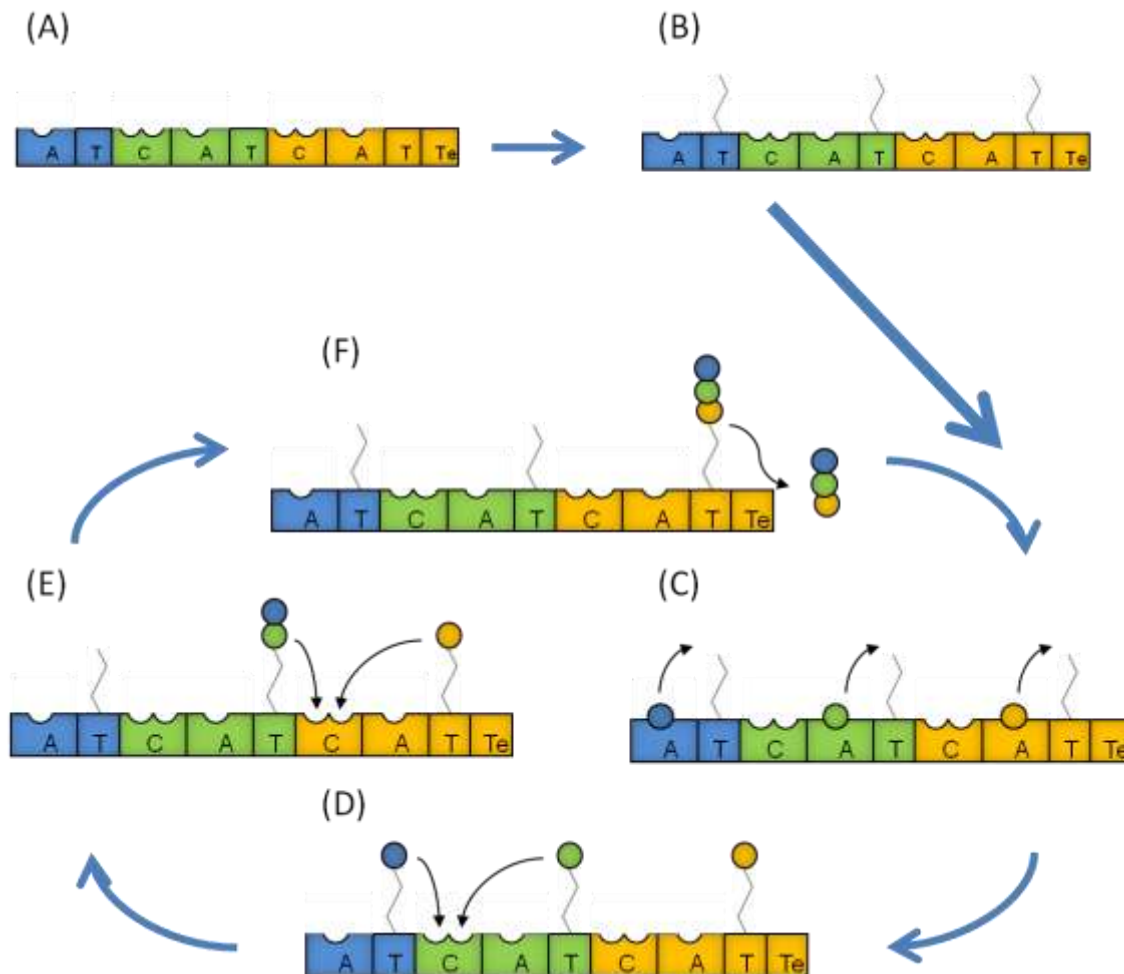


Figure 1-2. The domain architecture and mechanism of synthesis for a hypothetical NRPS. **(A)** The arrangement of domains within a hypothetical three module NRPS containing an initiation (blue), an elongation (green) and a termination (orange) module. **(B)** Activation of each module by post-translational attachment of a 4'-phosphopantetheine (4'-pp) cofactor to the T domain (sometimes referred to as a peptidyl carrier protein (PCP) domain). **(C)** The first domain to act within a module is an A domain that activates and tethers a specific monomer to the 4'-pp prosthetic group of the immediately downstream T domain. **(D)** The monomer attached to the T domain can then be passed between C domains. C domains are located upstream to the A domain, and C-A-T domains represent a basic elongation module. The C domain catalyses peptide bond formation between the carboxyl group of the donor substrate attached to the T domain of the upstream module and the amino group of the acceptor substrate of the downstream T domain. **(E)** Peptide bond formation breaks the upstream thioester bond, resulting in the peptide being attached to the downstream T domain, and this peptide serves as the donor substrate for the C domain of the next module. **(F)** After addition of the final monomer by the termination module, the peptide product is released by a thioesterase (TE) domain. After peptide release, the NRPS is returned to step (C) and peptide synthesis can repeat. NRPS domains are covered by several reviews (Felnagle et al., 2008; Hur et al., 2012; Marahiel et al., 1997; Sieber and Marahiel, 2005).

### 1.2.1 Adenylation domains

NRPS A domains function to adenylate a specific monomer and attach it to a T domain. Based on sequence similarity, they are part of a superfamily of adenylate-forming enzymes including acyl-CoA synthases and luciferases (Turgay et al., 1992). A domains catalyse an ATP-dependent two-step reaction. First an aminoacyl-adenylate is formed with concomitant release of pyrophosphate ( $\text{PP}_i$ ), followed by attachment of the aminoacyl intermediate to the 4'-pp thiol group of that module's T domain and the concomitant release of AMP (Figure 1-3) (Gevers et al., 1968, 1969; Stachelhaus and Marahiel, 1995; Stein et al., 1994). Although typically functioning as part of a multimodular enzyme, A domains are semi-autonomous and deletion experiments have shown the adenylation reaction (Figure 1-3A) can occur when A domains are expressed in isolation (Dieckmann et al., 1995; Haese et al., 1994; Stachelhaus and Marahiel, 1995). However, the thiolation step (Figure 1-3B) requires the presence of a T domain (Stachelhaus and Marahiel, 1995).

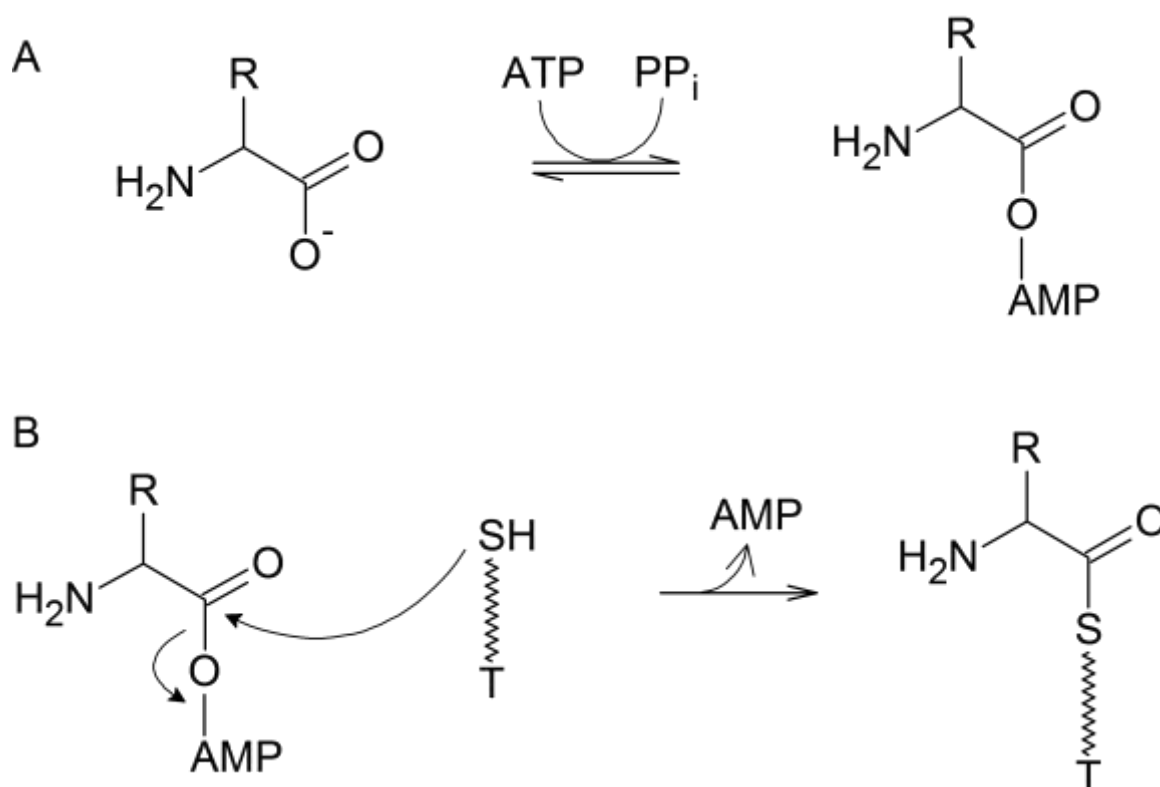


Figure 1-3. Two step adenylation domain reaction. (A) The A domain adenylates the amino acid substrate using ATP as a substrate. (B) The aminoacyl adenylate is then attacked by the terminal thiol of a 4'-pp group of the immediately downstream T domain to form a thioester. Image modified from Felnagle et al. (2008).

### 1.2.1.1 A domain structure and specificity

The first NRPS A domain structure to be determined was the X-ray crystal structure of PheA, the A domain from the NRPS GrsA (Conti et al., 1997). GrsA is the first NRPS involved in gramicidin S synthesis. It is a single module NRPS that activates Phe, epimerises it and passes it to the downstream enzyme GrsB. Together GrsA and GrsB synthesise a pentapeptide and two of these pentapeptides are connected head-to-tail to form the decapeptide gramicidin S. GrsA contains a single module consisting of an A, T and epimerisation (E) domain (The E domain epimerises L-Phe to a D-isomer following attachment to the T domain; Section 1.2.6). In this thesis, multiple domains of the same enzyme will be separated by hyphens, i.e. the GrsA domains will be referred to as A-T-E domains.

The X-ray crystal structure of PheA was solved to 1.9 Å and consisted of a large N-terminal domain (~450 amino acids) and a smaller C-terminal domain (~100 amino acids) connected by a small linker region (Conti et al., 1997). Despite sharing only 16 % amino acid sequence identity, PheA had high structural homology to the previously solved crystal structure of firefly luciferase (Conti et al., 1996, 1997). Sequence alignments demonstrate A domains have ten conserved motifs named A1 to A10. These are predicted to be involved in ATP binding, hydrolysis and adenylation of the substrate (Marahiel et al., 1997; Turgay et al., 1992). Consistent with their proposed catalytic roles these motifs are all located around the active site with the exception of motifs A1 and A2 (Marahiel et al., 1997).

An important feature of the PheA structure was that it was obtained in complex with Phe and AMP. The presence of Phe in the binding pocket of PheA allowed ten residues lining the amino acid binding pocket to be identified (Conti et al., 1997). Due to the large structural homology between PheA and firefly luciferase, and greater sequence similarity among A domains, it was assumed the structure of all A domains would be highly similar (Marahiel et al., 1997). Subsequently it was shown the substrate binding pocket residues can be identified by alignment of A domain peptide sequences, and these residues can be used to predict the substrate activated by A domains (Challis et al., 2000; Stachelhaus et al., 1999). Based on the binding pocket residues, online servers have been developed to provide substrate specificity predictions (Bachmann and Ravel, 2009). More recent algorithms have also been developed to take into account all residues within 8 Å of the

substrate (a distance that includes the substrate binding pocket residues) (Rausch et al., 2005; Rottig et al., 2011). These algorithms that take into account additional residues were found to have improved prediction ability (Rausch et al., 2005), suggesting residues close to the binding pocket also influence the substrate specificity.

Later A domain structures had an overall similar topography, but differed in the relative orientation of the two subdomains (May et al., 2002; Tanovic et al., 2008; Yonus et al., 2008). These changes in orientation were proposed to be essential for catalysis, and a mechanism of catalysis involving catalytic cycling of A domain structure was proposed (Yonus et al., 2008). In this model, an open conformation of the A domain allows ATP and the amino acid substrate to enter their respective binding pockets. Electrostatic interactions between ATP and the A domain cause a conformational shift of the small subdomain, creating a closed conformation that promotes formation of the aminoacyl adenylate. The formation of the aminoacyl adenylate and loss of pyrophosphate then causes further changes to the A domain structure, which push the 4'-pp prosthetic group of the T domain into the A domain binding pocket. This allows the formation of the aminoacyl thioester linkage to the 4'-pp prosthetic group with the concurrent release of AMP.

### 1.2.2 Thiolation domains

Located downstream to A domains are T domains, 80-100 residue domains to which activated amino acids and the growing peptide chain are attached (Sieber and Marahiel, 2005). The T domain is functionally and structurally related to other carrier proteins involved in polyketide and fatty acid synthesis (Mercer and Burkart, 2007). Following translation, T domains are modified by attachment of a 4'-pp prosthetic group derived from Coenzyme A (CoA). The attachment site is a conserved Ser residue within an [I/L]GG[D/H]SL motif (Schlumbohm et al., 1991; Stein et al., 1994), and attachment is mediated by phosphopantetheine transferases (PPTase) (Lambalot et al., 1996; Mootz et al., 2001). Three classes of PPTase exist; Sfp-type PPTases act on NRPS T domains and are named due to Sfp from surfactin biosynthesis being the archetype of this class (Beld et al., 2013; Hur et al., 2012).

The 4'-pp group attached to the T domain has a reach of about 18 Å and is thought to act as a swinging arm that transfers substrates between catalytic centres (von Döhren et al.,

1997; Hur et al., 2012). Although not catalysing any reactions itself, expression of the T domain from the first module of tyrocidine A synthesis as an isolated domain found it retains its ability to be activated by phosphopantetheine transferases and receive monomers *in trans* from A domains (Stachelhaus et al., 1996).

#### 1.2.2.1 T domain structure

An NMR solution structure of the T domain from the seventh module of tyrocidine A synthesis was the first T domain structure solved (Weber et al., 2000). Tyrocidine A is synthesised by three NRPS enzymes, TycA, TycB and TycC, which contain one, three and six modules respectively. Here, modules from tyrocidine synthetases are referred to by the enzyme they are located in (TycA, TycB or TycC) followed by a subscript of the module number within that enzyme, i.e. the seventh tyrocidine synthetase module is referred to as TycC<sub>3</sub>. The structure of the T domain from TycC<sub>3</sub> was shown to contain a bundle of four antiparallel helices named helix I-IV, with the 4'-pp attachment site located in a large loop between helices I and II; a similar topography to acyl carrier proteins (Weber et al., 2000). The TycC<sub>3</sub> T domain was further analysed by NMR spectroscopy and found to have three states: apo (A), holo (H) and A/H (Koglin et al., 2006). Prior to attachment of the 4'-pp cofactor, the T domain was found to slowly change between the A and A/H states; after attachment, the equilibrium was shifted to between the A/H and H states (Koglin et al., 2006). Mutation of the 4'-pp attachment Ser residue to an Ala locked the structure in the A state. The A-state contains helices I, II and IV, with the residues of helix III being stretched as a loop and embedded within the protein core. Within the A/H state, helix III reforms and is located outside the core of the protein. The H-state differs in that helix III is again unravelled. This unravelling of the helix III shifts the 4'-phosphopantetheine by 16 Å. Titration with other domains found the PPTase Sfp only interacted with the A-state, whereas SrfTEII (a type II thioesterase (TE<sub>II</sub>) functioning in the removal of aberrant CoA; see Section 1.2.5) interacted specifically with the H-state. This shows the conformational changes are important for the interactions with catalytic partners (Koglin et al., 2006). Additionally, it was suggested that the conformational shifting between states may also be involved in the movement of substrates between A and C domains (Koglin et al., 2006).

Substrates attached to T domains need to reach multiple catalytic partners. For example, in the context of an elongation module, a T domain needs to have a monomer attached to

it by an upstream A domain, and the tethered monomer then needs to cycle between both the upstream and downstream C domains. The role of conformational switching being important for activity was supported by later NMR and crystal structures of T domains solved in the presence of other NRPS domains (Frueh et al., 2008; Samel et al., 2007; Tanovic et al., 2008). These structures showed T domain conformational changes were necessary for the attached substrate to reach all catalytic partners, however, the precise conformational changes are unknown (Strieker et al., 2010; Weissman and Müller, 2008).

### 1.2.3 Condensation domains

The ~450 residue C domain is located upstream to the A-T domains and mediates peptide bond formation between substrates attached to the T domains of adjacent modules (Figure 1-4). C domains share significant amino acid identity with E domains, but do not appear to be closely related to any other protein. However, sequence alignments of C domains identified a conserved catalytic HHxxxDG motif, which is also found in chloramphenicol acetyltransferases (CAT) and dihydrolipoyl transacetylases (E2p) (De Crécy-Lagard et al., 1995). The second His residue of the HHxxxDG motif was confirmed to be essential for condensation by mutagenesis studies (Bergendahl et al., 2002; Roche and Walsh, 2003; Stachelhaus et al., 1998) and it was suggested this His, in a similar manner to that present in CAT and E2p, functions as a general base to deprotonate the acceptor substrate amine, enabling nucleophilic attack of the carbonyl group of the donor substrate (Bergendahl et al., 2002).

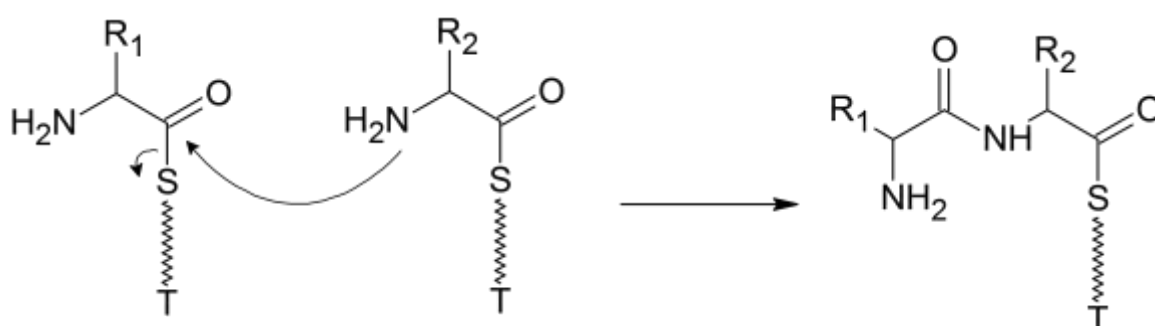


Figure 1-4. Peptide bond formation catalysed by the C domain reaction. Image modified from Felnagle et al. (2008).

### 1.2.3.1 C domain structure

The first structure of a C domain was of VibH solved at 2.55 Å by X-ray crystallography (Keating et al., 2002). VibH is an unusual C domain as it is a standalone domain (rather than part of a multi domain module), which catalyses condensation of 2,3-dihydroxybenzoic acid (DHB), tethered to the standalone T domain VibB, directly onto norspermidine. Nevertheless it was argued that the structure of VibH would resemble both C and E domains because of a high degree of primary sequence similarity (Keating et al., 2002). The structure of VibH contained two subdomains with the putative catalytic His placed in a V-shaped solvent channel running between the subdomains. There was an extended flexible loop spanning over the solvent channel from one subdomain to the other. The channel provided access by the donor and acceptor substrates from adjacent ends of the channel. Mutation of the catalytic His to either Ala or Glu did not hinder activity, suggesting that this residue does not act as a base for VibH (Keating et al., 2002). The His not acting as a base was hypothesised to be due to VibH binding norspermidine directly and not requiring a base for catalysis, or that VibH may function to align the residues close to one another and the reaction occurs spontaneously due to amide bonds being thermodynamically favoured over thioesters (Keating et al., 2002).

The second C domain crystal structure to be solved was of the T-C domains from the ninth and tenth modules of tyrocidine synthesis at a resolution of 1.8 Å (Samel et al., 2007). The C domain contained a similar V-shape solvent channel to VibH with a loop spanning the channel. The authors made an analogy of the loop spanning the channel to the flexible lid regions of NRPS TE domains. Comparing the structure to VibH found the subdomains were highly similar but the orientation of the subdomains to each other was rotated by 12°. Calculating the pK values of the putative catalytic His residue suggested the His residue is protonated under physiological conditions and may not act as a catalytic base as previously suggested (Bergendahl et al., 2002). Instead the His residue may function to electrostatically stabilise a tetrahedral reaction intermediate (Samel et al., 2007). Another feature of the tyrocidine synthetase T-C domain structure was the T domain facing away from the catalytic centre of the C domain at a distance of over 47 Å to the His residue. In this state the 4'-pp attachment site was at a distance too great for the T domain to pass its substrate to the C domain catalytic centre (~20 Å in length)

and may represent a state compatible for interactions with the upstream A or C domain (Samel et al., 2007).

A third C domain structure was part of a crystal structure solved at 2.6 Å of SrfA-C, the entire termination module from surfactin biosynthesis (Tanovic et al., 2008). SrfA-C contains C-A-T-TE domains and adds the final residue to surfactin prior to release by the TE domain. To obtain uniform crystals, it was necessary to mutate the Ser attachment site of the T domain to Ala (Strieker et al., 2010). The domains had similar folds to structures of individual domains solved in other studies. In the structure, there was a large surface of interactions at the C/A domain interface ( $\sim 1621 \text{ Å}^2$ ) and the linker region between these domains appeared inflexible as it was well defined and closely associated with the C and A domains (Tanovic et al., 2008). The interactions at the C/A interface suggest these domains form an invariant “workbench” and do not shift from one another during catalysis. Importantly, it was hypothesised this invariant workbench may limit the ability to fuse NRPS enzymes at the C/A domain junction (Tanovic et al., 2008). The active sites of the C and A domain were separated by greater than 63 Å – a distance not coverable by the 4'-pp cofactor of the T domain – supporting the assumption that large changes in T domain conformation would be needed during catalysis (Tanovic et al., 2008).

### 1.2.3.2 C domain substrate specificity

Although A domains govern the initial selection of monomers, C domains have also been shown to exhibit specificity. The specificity of C domains was first tested by loading aminoacyl-CoA directly onto the T domains of a two module NRPS system. The first module incorporates Phe and contains A-T-E domains, the second incorporates Pro and contains C-A-T domains (Belshaw et al., 1999). These two modules, referred to as PheATE and ProCAT, were previously shown to non-enzymatically release the dipeptide as a diketopiperazine (DKP) product following condensation (Figure 1-5) (Stachelhaus et al., 1998). Artificial loading of the T domain from PheATE with five different amino acids (L-Ala, D-Ala, L-Leu, L-Phe and D-Phe) and the T domain from ProCAT with Pro led to a DKP being formed in all cases. As the T domain of PheATE passes its substrate to the C domain donor site of ProCAT, it was concluded C domains have no stereo or side-chain specificity towards the monomer received. Next, the same five substrates were loaded onto the T domain of ProCAT whilst Phe was loaded onto PheATE. As Pro is needed for DKP formation, condensation in this case was measured by attachment of a

dipeptide to ProCAT. Out of the five monomers loaded onto ProCAT, only L-Ala allowed product formation. Based on only one amino acid acting as an acceptor substrate, it was concluded that the acceptor sites of C domains have high side-chain and stereo specificity (Belshaw et al., 1999).

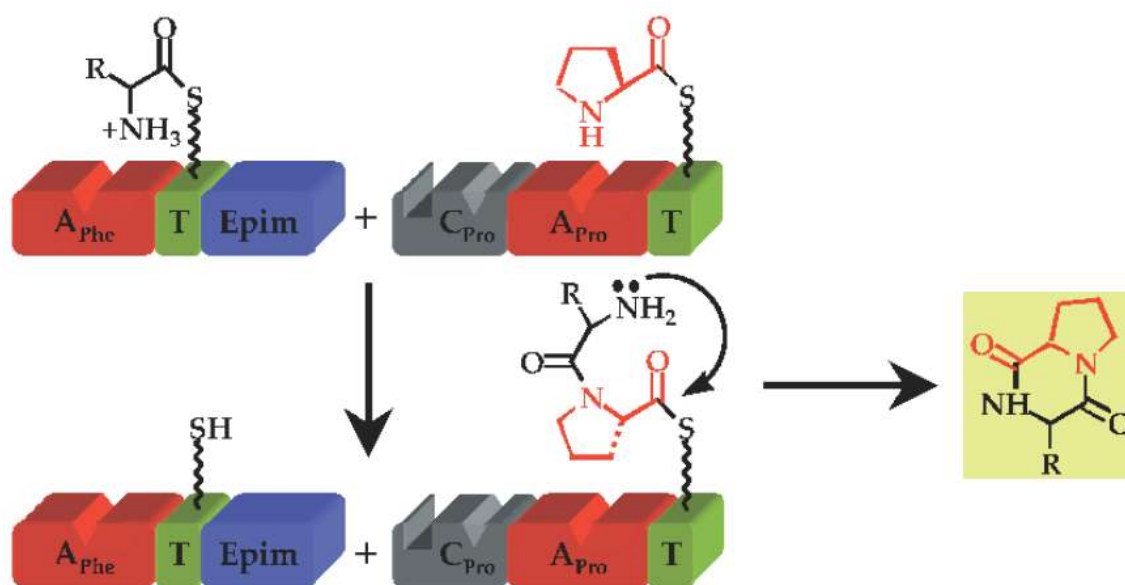


Figure 1-5. Formation of a DKP product using PheATE and ProCAT. Artificially loading an amino acid onto PheATE and Pro onto ProCAT leads to spontaneous release of a DKP product following condensation. R refers to the side-chain of amino acid loaded onto PheATE. Image modified from Belshaw et al. (1999).

A second substrate specificity study used aminoacyl-*N*-acetylcysteamine thioesters (aminoacyl-SNACs) as substrates for the C domains from ProCAT and the NRPS EntF of enterobactin synthesis (Ehmann et al., 2000). These aminoacyl-SNACs serve as mimics of substrates bound to T domains and pass substrates directly to the C domain. This was an advantage as in the previous study the E domain from PheATE may have isomerised the substrates, meaning it only tested amino acid side-chain specificity at the C domain donor site. This second study tested L-Phe- and D-Phe-SNACs as donor substrates for ProCAT and found only D-Phe could be condensed by the C domain. In addition, the acceptor site specificity of the C domain from EntF was tested with six non-cognate L-amino acids and the corresponding D-amino acids. This found the rate of product formation halved when using an L-Ala-SNAC, and was greatly reduced for all other aminoacyl-SNACs (including the D-Ser-SNAC) compared with the wild type L-Ser-SNAC. The authors concluded C domains have relaxed side-chain and strong

stereo specificity at the donor site, and strict side-chain and stereo specificity at the acceptor site (Ehmann et al., 2000).

A problem with the above substrate specificity studies is they only tested substrate specificity towards individual amino acids. With the exception of the initiation module, C domains generally receive a peptide as a donor substrate. A second problem is the small sample size, since the donor site side-chain specificity was only tested with Phe, Ala and Leu. A subsequent paper tested the donor site substrate specificity using the PheATE, ProCAT system with seven dipeptides and one tetrapeptide (Stein et al., 2005). The ProCAT module was fused with a TE domain, because an earlier study showed the TE domain releases a dipeptide quicker than the spontaneous DKP reaction (Schwarzer et al., 2001). The PheA module was loaded with the dipeptides and tetrapeptide, and ProCAT-TE was loaded with Pro. Released products were detected by mass spectrometry when using the donor tetrapeptide and 5/7 dipeptide substrates. The two dipeptides unable to be condensed by the C domain had a C-terminal Thr or Lys residue, whereas all the condensed products had either a C-terminal Phe, Leu or Ser residue. This result suggests C domains can sometimes discriminate based on the terminal amino acid of an incoming peptide (Stein et al., 2005).

### 1.2.3.3 Control of condensation timing

According to the multiple template model, non-ribosomal peptide synthesis starts at the initiation module and occurs by addition of monomers in the N-terminus to C-terminus direction. The directionality of synthesis was initially shown in studies of adding amino acids to NRPS extracts, whereby initiation of peptide synthesis required the first amino acid and absence of an amino acid would stop synthesis at a truncated peptide (Gevers et al., 1969; Kleinkauf et al., 1969; Lipmann et al., 1970). For example, synthesis of tyrocidine (with a peptide sequence of D-Phe-L-Pro-L-Phe-D-Phe-L-Asn-L-Gln-L-Phe-L-Val-L-Orn-L-Leu) by partially purified enzyme extracts required the presence of all the amino acids of tyrocidine, whereas absence of a single residue resulted in a truncated peptide attached to the NRPS, i.e. absence of only Asn from tyrocidine extracts would produce a Phe-Pro-Phe-Phe tetrapeptide and no other peptides such as Gln-Phe-Val-Orn-Leu (Lipmann et al., 1970). To explain why peptide synthesis does not initiate at elongation modules, it was proposed C domains contain a “waiting” position, which holds the aminoacyl thioester blocking it from acting as a substrate for other domains until after

condensation occurs (de Crecy-Lagard et al., 1997). Owing to the high acceptor site specificity of C domains, it was suggested that a C domain acceptor site binding pocket may function to hold the aminoacyl thioester until after peptide bond formation (Belshaw et al., 1999).

Based on the hypothesis of the C domain holding the aminoacyl thioester, an elongation module could be converted into an initiation module by deletion of the C domain or inability of the C domain to bind the acceptor substrate due to substrate specificity. This was tested by Linne and Marahiel (2000) by creating a series of truncations of the second and third modules of TycB. Both these elongation modules could initiate synthesis when the C domain was deleted but not otherwise (Linne and Marahiel, 2000). Furthermore, TycB<sub>3</sub> normally activates L-Phe and incorporates D-Phe into tyrocidine. Adding D-Phe instead of L-Phe as a substrate for TycB<sub>3</sub> found peptide synthesis was initiated even when TycB<sub>3</sub> contained a C domain. Presumably this was a result of the A domain activating D-Phe, but the C domain acceptor site not holding D-Phe due to stereo specificity. These results support that the C domain is important in timing of peptide bond formation and also suggest epimerisation by the E domain occurs after release from the C domain “waiting” position (Linne and Marahiel, 2000).

#### 1.2.3.4 C domain specialisation

As discussed above, C domains have been shown to have stereo specificity towards the donor substrate. In agreement with this functional diversity, phylogenetic analysis identifies subtypes of C domain according to their incoming donor substrate or additional functions (Rausch et al., 2007). These subtypes are referred to as: <sup>L</sup>C<sub>L</sub>- and <sup>D</sup>C<sub>L</sub> domains named according on whether the substrate received within the donor site of the C domain is an L-amino or D-amino, respectively. In addition there are starter C domains that receive a fatty acid in the donor site, and which are sometimes present in an initiation module; cyclisation (Cy) domains that can catalyse heterocyclisation of the substrate in addition to condensation; and dual E/C domains, which catalyse epimerisation and condensation (Rausch et al., 2007). These subtypes of C domain all condense the donor substrate onto an L-isomer of the acceptor substrate.

#### 1.2.4 Thioesterase domains

The termination module responsible for peptide release contains a TE domain immediately downstream to the T domain. The role of the TE domain in product release was confirmed by module rearrangement experiments, where moving the TE domain within the surfactin pathway to the C-terminus of earlier modules resulted in truncated products and deleting the TE domain abolished surfactin synthesis altogether (Ferra et al., 1997). In addition to product release, for Type B NRPSs, the peptide intermediates are thought to be held at the TE domain until synthesis of the entire peptide is complete (Mootz et al., 2002).

For product release by the TE domain, a Ser-Asp-His catalytic triad suggests TE domains catalyse their reactions via a similar mechanism to enzymes of the  $\alpha/\beta$ -hydrolase family (Ferra et al., 1997; Trauger et al., 2000). Mutagenesis studies found the Ser and His residues are essential, whereas the Asp residue is only sometimes required (Shaw-Reid et al., 1999; Tseng et al., 2002). The first step of catalysis is transfer of the peptide chain from the 4'-pp group of the T domain to the Ser residue of the TE domain catalytic triad via formation of an ester bond (Figure 1-6A), followed by release from the TE domain by hydrolysis or intramolecular cyclisation (Figure 1-6B). Intramolecular cyclisation can proceed via formation of amide or ester bonds. The amide bonds are formed with either an  $\alpha$ -amino or side-chain amino group, whereas ester bonds are formed with residues containing a side-chain hydroxyl group (Figure 1-6B) (Du and Lou, 2010; Felnagle et al., 2008).

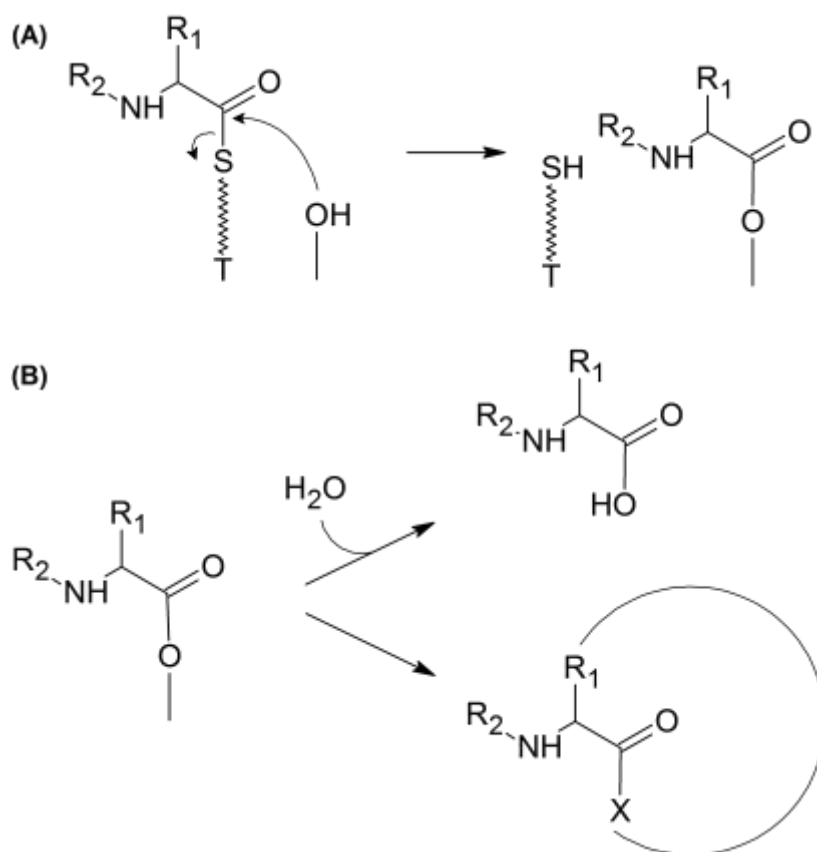


Figure 1-6. TE domain reactions. (A) Peptide transfer from the T domain to the Ser residue of the Ser-Asp-His catalytic triad of the TE domain. R1, amino acid side chain; R2, peptide chain (B) Product release occurs either by hydrolysis or intramolecular cyclisation. X can be either O or N. Image modified from Felnagle et al. (2008).

#### 1.2.4.1 Structure of TE domains

The crystal structure of SrfTE, the TE domain involved in surfactin biosynthesis, identified a similar fold to other members of the  $\alpha/\beta$  hydrolase superfamily with the active site within a cavity covered by a loop (Bruner et al., 2002). Although size exclusion chromatography showed SrfTE to be monomeric, the TE domain structure was a dimer; possibly an artefact caused by crystallisation. The loop in one monomer was open allowing access to the catalytic site, whereas it was closed in the second (Bruner et al., 2002). A second TE domain crystal structure was of FenTE from fengycin biosynthesis (Samel et al., 2006). This structure was similar to SrfTE, except solved in the monomeric state. The lid region was shorter than that of SrfTE, and molecular dynamic simulations suggested this loop does not cover the active site in FenTE (Samel et al., 2006). It was suggested that for cases like SrfTE the lid may function similarly to lids in lipases and sterically hinder access to the catalytic residue or occlude water during cyclisation (Samel

et al., 2006). However, sequence alignments of TE domains found that the size of the loops do not correlate to size of substrates released by TE domains. As an alternative, it was suggested the loop may aid in substrate binding as well as occlusion of other molecules (Samel et al., 2006).

### 1.2.5 Type II thioesterase enzymes

An additional enzyme often associated with NRPS pathways are TE<sub>II</sub> enzymes. It has been suggested that TE<sub>II</sub> enzymes may function to remove 4'-pp cofactors with aberrant thioesters (Schneider and Marahiel, 1998). This was supported by deletion experiments showing the absence of TE<sub>II</sub> enzymes greatly impairs non-ribosomal peptide production (Geoffroy et al., 2000; Schneider and Marahiel, 1998). In the *in vivo* setting, T domains are likely to be mis-primed with acetyl-CoA or by loading of the incorrect amino acid, resulting in peptide synthesis stalling (Hur et al., 2012). The ability of TE<sub>II</sub> domains to remove aberrant CoA from T domains was confirmed by *in vitro* assays (Schwarzer et al., 2002; Yeh et al., 2004). Based on increased specificity towards amino acid versus peptides substrates, it was suggested that C domain acceptor site specificity and editing of non-cognate residues creates a more energetically efficient product; releasing mis-activated substrates before incorporation rather than incorporating incorrect residues, requiring hydrolysis of the whole peptide chain (Yeh et al., 2004).

### 1.2.6 Accessory domains

In addition to the domains discussed above, accessory domains may be present to further increase the diversity of NRPS products. These can be independent from the NRPS complex, or located within the assembly line to act during synthesis. The internally integrated domains include: epimerisation (E) domains that convert L-amino acids into a D-amino acids, methyl transferase domains that methylate amino acids, and Cy domains that catalyse heterocyclisation of side chains of Cys, Ser and Thr residues (Walsh et al., 2001). The most relevant of the accessory domains to this study is the E domain.

Generally in non-ribosomal peptide synthesis, D-amino acids are incorporated by the A domain activating an L-amino acid, which is then racemised by the E domain after attachment to the T domain and, when present, release by the C domain. The E domain is located immediately downstream to that T domain (Marahiel et al., 1997; Sieber and Marahiel, 2005). E domains share the same HHxxxDG motif as C domains and, as noted

above, sequence alignments suggest similarity between C and E domains (De Crécy-Lagard et al., 1995). The donor substrate is epimerised rather than the acceptor substrate, meaning E domains act to epimerise aminoacyl-S-phosphopantetheine for initiation modules and peptidyl-S-phosphopantetheine for elongation and termination modules (Shiau et al., 1995; Stachelhaus and Marahiel, 1995; Stein et al., 1995; Stindl and Keller, 1994). As discussed in Section 1.2.3.3, this timing may be controlled by the aminoacyl thioester being held at the acceptor site of the C domain until after condensation occurs. The L- to D-aminoacyl thioester conversion by E domains is incomplete, for example the E domain from PheATE epimerises the Phe substrate to 66 % D-Phe and 34 % L-Phe (Stachelhaus and Walsh, 2000). However the DKP released from PheATE and ProCAT contains 98 % D-Phe, indicating the E domain racemises the substrate and the D-isomer is subsequently selected by the C domain (Stachelhaus and Walsh, 2000).

### 1.3 Engineering of novel NRPS enzymes

In the preceding sections, the general mechanism of synthesis by NRPS enzymes and the domains involved within each module were discussed. Being able to successfully alter the specificity of pathways would enable the creation of derivatives or entirely new pathways of non-ribosomal peptides. In nature, the diversity of non-ribosomal peptides is thought to have arisen from point mutation, substitution of domains and modules for alternative ones, and the deletion or insertion of modules (Fischbach et al., 2008). This diversity in peptides and semi-autonomous modular mode of synthesis makes NRPS enzymes an attractive target for alteration of substrate specificity. As A domains are the initial determinants of monomer selection, new products could be created by substituting A domains within a pathway for A domains that activate alternative substrates or by directly altering the substrate binding pocket of the A domain. Below A domain substitution and binding pocket alteration studies are discussed, followed by domain substitution experiments that treat C-A domains as inseparable partners. Left out of this discussion are previous domain substitution experiments in *P. aeruginosa* PAO1, which will be discussed in Section 1.4.2.

#### 1.3.1 Studies on A and A-T domain substitution

Due to their similarity, both A and A-T domain substitutions will also be considered in this section. The first efforts to alter the products of NRPS enzymes by domain

substitution replaced a Leu specifying A-T domain pair, from the final module of the surfactin NRPS SrfA-C, with alternative A-T domains using homologous recombination (Stachelhaus et al., 1995). A-T domains specifying five different substrates were substituted and new products detected from each variant using mass spectrometry (Stachelhaus et al., 1995). In a second study, Leu specifying A-T domains, from the second module of surfactin biosynthesis, were substituted with seven different A-T domains (Schneider et al., 1998). *In vitro* assays confirmed the A-T domain substitution constructs were functional in terms of adenylation and thioester formation. Despite this, the yield of all products was severely reduced *in vivo*, and the authors mentioned yields were similarly reduced in the first A-T domain substitution study (Schneider et al., 1998; Stachelhaus et al., 1995). In each of these studies, one substitution was for a Leu specifying A-T domain, i.e. the same substrate specified by the modified module. This implied the loss in activity was not solely due to issues of substrate specificity (Schneider et al., 1998). However, the later discovery that C domain exhibit strong acceptor site specificity (see Section 1.2.3.2) goes some way to explaining why the other ten A domain substitutions might have been non-functional.

A third study compared the function of A and A-T domain substitutions in TycA<sub>1</sub>, the first module of tyrocidine A synthesis, and revealed the importance of specialised T domains for interacting with downstream E domains (Linne et al., 2001). The TycA<sub>1</sub> module is an initiation module containing A-T-E domains. As this initiation module contains no upstream C domain, acceptor site specificity of the C domain would not be expected to limit the efficacy of domain substitutions. Eight variants were created where either just the A, or the A-T, domains from four modules were substituted for the corresponding domain(s) in TycA<sub>1</sub>. All A domain and A-T domain substitutions were active in *in vitro* adenylation assays, and the A domain substitution variants were further shown to epimerise the activated substrate. In contrast, three A-T domain substitutions that do not have an immediately downstream E domain in their native setting were unable to epimerise the activated substrate. Alignment of T domain primary sequences identified two residues adjacent to the Ser 4'-pp attachment site that are conserved differently for T domains upstream to E domains compared to those upstream to C domains. Mutagenesis confirmed each of these were essential for the amino acid to be passed from the T domain to the E domain (Linne et al., 2001).

A later domain substitution experiment suggested that disruption of the T/TE domain interface can also result in large decreases in activity. Domain substitutions were used to create dimodular hybrid NRPS enzymes in an effort to synthesise L-Asp-L-Phe, a precursor to the sweetener aspartame (Duerfahrt et al., 2003). The fifth module of surfactin synthesis (which specifies Asp) and modules one and three of tyrocidine synthesis (both of which specify Phe) were combined to create six dimodular enzymes encoding an L-Asp-L-Phe dipeptide. These six enzymes differed in the location of fusion sites, such that the activity of an A versus C-A domain substitution could be compared as well as a TE versus T-TE domain substitution. The most active domain substitution was a C-A domain substitution variant, which gave 5.8 times greater product yield than any variants in which the C/A or T/TE junctions had been disrupted. The loss of activity observed when the C/A domain junction had been disrupted was consistent with the previous observations that C domains exhibit acceptor site specificity, i.e. cannot tolerate alternative A domains immediately downstream; whereas the lack of activity observed whenever the T/TE domain boundary had been disrupted suggested that the role of a T domain upstream to a TE domain may be particularly specialised.

Outside of the pyoverdine system (the reasons from drawing this distinction will become clear shortly) there has been one further study examining an A domain substitution downstream to a C domain. For this substitution, the Ser specifying A domain from EntF, an NRPS involved in enterobactin synthesis, was replaced with an alternative Ser specifying A domain (Fischbach et al., 2007). This A domain substitution variant exhibited a 32-fold reduction in enterobactin synthesis compared to the wild type enzyme. This reduction was deemed to be likely due to insolubility, and directed evolution created a soluble variant with only a 3.8-fold reduction in activity (Fischbach et al., 2007).

In the A and A-T domain substitution studies described above, only three substitutions were made where the new A domain would be compatible with the acceptor site specificity of the C domain. That each of these three domain substitutions exhibited a substantial loss in activity relative to the wild type enzyme suggests that factors other than C domain acceptor site specificity are also important for the success of A domain substitutions. An alternative hypothesis put forward for why A domain substitutions tend to substantially impair activity is that they generally disrupt a structurally important C/A domain interface (Tanovic et al., 2008).

### 1.3.2 Alteration of A domain binding pockets

An alternative to A domain substitution is to directly alter the specificity of the A domain binding pocket. As the substrate specificity of A domains can be predicted from the binding pocket, it follows that the substrate specificity could potentially be altered by changes to just these residues. Studies that have attempted to alter A domain binding pockets have been separated here into two groups, depending on whether activity was measured *in vitro* or *in vivo*.

Studies testing the activity of modified A domain binding pockets *in vitro* have altered the binding pockets using multiple methods. These include: [i] changing the binding pocket to match other A domains (Eppelmann et al., 2002; Stachelhaus et al., 1999); [ii] making *in silico* predictions of binding pocket mutations that would maintain contact with a new substrate while having the least effect on structure (Chen et al., 2009; Stevens et al., 2006); and [iii] directed evolution where the residues lining the binding pocket are randomised and screening variants for activation of a new substrate (Villiers and Hollfelder, 2011; Zhang et al., 2013). In all these studies, the activity of mutated A domains was measured by the kinetics of the adenylation reaction. Alteration of the binding pocket ([i] above) yielded two successes where adenylation of new substrates occurred at similar rates to a wild type enzyme. However, specificity could not be changed for some A domains without a large reduction in activity (Eppelmann et al., 2002; Stachelhaus et al., 1999). From the other binding pocket modification studies, the best catalytic efficiency reached was 22 % of the catalytic efficiency for the wild type enzyme and substrate (Zhang et al., 2013). Interestingly, the study that reached 22 % catalytic efficiency also measured the rate of thioester formation. This revealed that the modified A domains were unable to pass the activated substrate to the associated carrier protein; a problem that was repaired by a point mutation (Zhang et al., 2013). This highlights that care needs to be taken when testing NRPS activity using kinetics alone, as there are subsequent reactions that must be able to occur to achieve effective non-ribosomal peptide synthesis.

*In vivo* studies of A domain specificity can be more complex, but have the advantage that detected products report on the combined outcomes of adenylation, thiolation and condensation reactions. Three *in vivo* studies have changed binding pockets to encode alternative substrates and successfully detected novel products by mass spectrometry

(Eppelmann et al., 2002; Thirlway et al., 2012; Uguru et al., 2004). However, yields were not quantified in any of these studies, making it difficult to accurately evaluate the success of substitutions. Another *in vivo* study randomised three residues of an A domain substrate binding pocket involved in andrimid synthesis to generate a library of 1404 different mutants and screened for new products by mass spectrometry (Evans et al., 2011). Four mutants were identified as being able to create new products, with two giving relatively high yields when fed excess substrate. A final *in vivo* study modified the third A domain in fusaricidin synthesis (Han et al., 2012). This A domain has promiscuous activity and can incorporate five different substrates into fusaricidin. Rather than add new activity, the authors of this study sought to improve the specificity of the A domain towards Phe, because fusaricidin containing Phe has the greatest antibiotic activity. Four residues of the binding pocket were mutated in various combinations to give a total of 6 mutants. This shifted the percent production of fusaricidin containing Phe from 9.4% to between 18.9 % and 26.4 % of the total for three of the mutants. The increase was not associated with a reduction in total fusaricidin synthesis, showing this method to be highly useful in altering the relative abundance of products.

Although in some cases successful, modification of A domain binding pockets often results in a large loss of activity. Despite trying thousands of combinations, studies that randomised the binding pockets have also had difficulty in producing new products. This suggests that residues outside the substrate binding pocket may have a larger influence than previously assumed. Supporting the importance of residues outside the binding pocket is the improved accuracy of updated algorithms for predicting A domain specificity based on all residues within a given distance of the A domain (Rausch et al., 2005).

### 1.3.3 C-A domain substitution

Both the issue of C domain acceptor site specificity and that of disrupting the C/A domain interface can be overcome by keeping native C-A domains together during domain substitution experiments. The first C-A domain substitution experiments used *in vitro* models like the PheATE/ProCAT system, rather than modify modules within the context of a pathway. Based on evidence of C domain acceptor site specificity, Mootz *et al.* (2000) generated new NRPS enzymes treating C and A domains as inseparable units. Using the PheATE and ProCAT system, entire modules were fused onto ProCAT to

create two trimodular NRPS pathways. The added modules were a Leu termination module and an Orn elongation module. To enable product release, a TE domain was fused onto the Orn elongation module. Both trimodule pathways were active in adenylation assays. However, consistent with the T/TE junction being important, the Orn variant synthesised its peptide at a rate 14 times slower than the termination module variant. A similar study identified flexible regions between T and C domains based on sequence alignment and used these to create three dimodular NRPS enzymes (Doekel and Marahiel, 2000). Two of these were made by fusing an A domain onto T-C-A-T-TE domains (T-C-A-T-TE domains represent a termination module plus the T domain from the upstream module) and a third was made by combining an A domain, a native T-C-A domain pairing, and a native T-TE domain pairing. All three variants produced the predicted dipeptides. Although focused on creating new enzymes and testing them *in vitro*, rather than modifying existing pathways, these two studies suggest that not disrupting the C/A domain junction and fusing enzymes in certain tolerated locations can enable the creation of new peptides.

The flexible region between T-C domains was further characterised using the daptomycin NRPS pathway (Doekel et al., 2008). Daptomycin is a 13 residue peptide that has an N-terminus decanoic acid group. It is synthesised by three NRPS enzymes named DptA, DptBC and DptD. When testing the flexibility of the T-C linker in the two module NRPS DptD, the location of restriction sites and addition or deletion of four residues had little effect on levels of daptomycin synthesis, demonstrating that this linker region is relatively tolerant of alteration (Doekel et al., 2008). CDA and A54145 are similar lipopeptides to daptomycin, having a similar structure but differing in some of the peptide sequence. Using the T-C linker splice site, the termination module of *dptD* was exchanged with the termination modules from CDA and A54145 synthesis. The predicted change in the final residue from Kyn to Trp and Ile/Val respectively were confirmed by mass spectrometry, and shown to have little effect on lipopeptide yield (Doekel et al., 2008). In addition, domain substitutions of Asn specifying C-A and C-A-T domains from the eleventh module of A54145 synthesis into the *dptD* gene were made. The C-A-T domain substitution disrupted the T/TE junction and resulted in no product, whereas the C-A domain substitution yielded lipopeptides at around 30 % to 50 % of WT levels (Doekel et al., 2008).

An earlier paper highlighted the ability to create multiple daptomycin analogues (Nguyen et al., 2006). First, modules 8 and 11 within the daptomycin pathway, which activate Ala and Ser respectively, were substituted for one another. These two domain substitutions yielded modified daptomycin at a level of about 16 % and 44 % compared to restriction site controls. Next a D-Asn module from the A54145 pathway was substituted into these modules as C-A-T and C-A-T-E domain substitutions. The substitutions produced modified daptomycin for the C-A-T and C-A-T-E domain substitutions respectively at 8 % and 3 % of WT levels in module 8, and 15 % and 8 % in module 11. A final test substituted four modules from LptC for four in DptB. The altered products were successfully produced and confirmed by mass spectrometry, albeit with a lipopeptide yield of only 0.4 % of WT levels. These alterations were combined with previous alterations of DptD at position 12 and 13 to create a combinatorial library of 30 variants. In total, 21 of these variants produced altered products (Nguyen et al., 2006).

A more recent example introduced domain substitutions in the A54145 biosynthetic pathway (Nguyen et al., 2010). As mentioned earlier, A54145 is a lipopeptide related to daptomycin. Similar to daptomycin, it is composed of a decanoic acid moiety attached to a 13 residue peptide. By making substitutions in multiple modules using the linkers between T-C domains, a total of eight variant pathways containing 1-5 changes to A54145 were created. Each substitution gave modified A54145 at between 0.3 % and 15.9 % of wild type levels (Nguyen et al., 2010). These examples of substitutions that have kept C-A domains together as inseparable partners are not exhaustive, but are selected to illustrate two ideas. First, the successes of C-A domain substitutions appear relatively high compared to modifications of the A-domain. Secondly, in spite of this increased success, there are still losses in activity caused by undetermined factors.

#### 1.3.4 Inter-enzyme communication

Another approach to generating modified products is by modifying interenzyme interactions. Many NRPS pathways are located on multiple enzymes, and the correct association between upstream and downstream enzymes is essential for correct peptide assembly. Short terminal communication-mediating (COM) domains were identified that are involved in mediating this interaction (Hahn and Stachelhaus, 2004). Upstream enzymes contain a C-terminus donor COM domain, which binds the N-terminus acceptor COM domain of the downstream NRPS enzyme (Hahn and Stachelhaus, 2004). These

COM domains can be substituted for new ones to create new products by changing the interactions of NRPS enzymes (Chiocchini et al., 2006; Hahn and Stachelhaus, 2006). Although the COM-domains can have a large effect on protein interactions, native NRPS partners can sometimes still interact despite COM domains being altered to non-cognate pairs (Chiocchini et al., 2006). As such, it was suggested factors outside the COM domains are likely to have an additional influence on protein interactions and this may be a further role of the T domain (Chiocchini et al., 2006). A recent review, highlighted that additional recognition elements are expected because COM domains are not always present; for example, the andrimid pathway contains many enzymes without any detectable docking domains (Koglin and Walsh, 2009).

### 1.3.5 Conclusions from previous domain substitutions efforts

An overall summary of previous domain substitution studies is that results can be highly variable from one system to the next, with few fully defined reasons for the losses in activity. The main reasons offered in reviews for why A domain substitutions lose activity relate to either the substrate specificity of the C domain or the C/A domain interface being essential (Baltz, 2012; Marahiel, 2009; Strieker et al., 2010). However, it is difficult to draw a conclusion on which of these is most likely based on the low number of A domain substitutions that have been published. The alternative of targeted A domain binding pocket modification would probably keep the C/A interface intact, however, it has been conclusively shown that the substrate specificity of some A domains cannot be efficiently changed just by altering the A domain binding pocket. Out of the studies discussed, substituting C-A domains together has produced the best results. In a recent review, rules for successful domain substitution were suggested. These rules were: (i) keep C-A domains together; (ii) C domains need to be located to receive the same type of substrate in their natural setting (i.e. an L-amino, D-amino or fatty acid); and (iii) the type and subtype of domain downstream to T domains needs to be kept constant (Baltz, 2012). Despite these rules, there is still frequently a large loss of activity in C-A domain substitutions. Further understanding for the reasons in loss of function for C-A domain substitutions is needed to better predict which substitutions are likely to be functional.

## 1.4 Pyoverdine as an alternative system to study domain substitution

Pyoverdine is a yellow-green siderophore used by fluorescent pseudomonads to sequester iron from the environment. Due to the poor solubility of ferric iron, passive uptake of iron from the environment is generally unable to support bacterial growth. To increase iron uptake bacteria often synthesise siderophores that, when secreted, chelate iron and are then taken into the cell (Hider and Kong, 2010). Pyoverdine is the primary siderophore used by fluorescent pseudomonads (Cornelis, 2010; Visca et al., 2007). In addition to pyoverdine, fluorescent Pseudomonads usually contain lower affinity secondary siderophores, which may function when iron levels are slightly limited or have alternative roles (Cornelis, 2010). The structure of pyoverdine consists of a conserved fluorescent dihydroquinoline chromophore with both an acyl chain and a variable peptide attached to it. The chromophore is conserved in all pyoverdines, whereas the peptide chain varies greatly between strains. The structure of pyoverdine synthesised by *P. aeruginosa* PAO1 is shown in Figure 1-7 (Briskot et al., 1989; Demange et al., 1990).

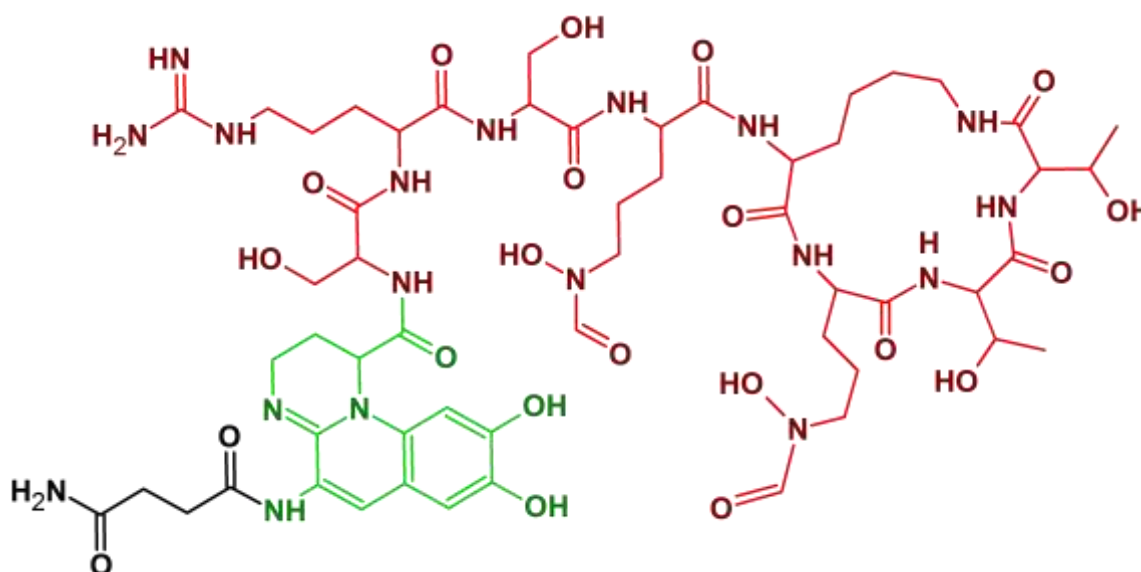


Figure 1-7. Pyoverdine from *P. aeruginosa* PAO1 consists of an acyl chain (black), chromophore (green) and peptide chain (red). The acyl chain shown here is succinamide, but can also be succinate, or  $\alpha$ -ketoglutarate (Briskot et al., 1989; Demange et al., 1990).

In *P. aeruginosa* PAO1, pyoverdine is synthesised by four NRPS enzymes (Figure 1-8, panels A & B), which are located in the cytoplasm. The first of these, PvdL, contains four

modules that assemble the residues later processed into the chromophore. Due to its role in chromophore assembly, PvdL is conserved among all fluorescent *Pseudomonads* (Mossialos et al., 2002). The first module of PvdL is unusual as it consists of an acyl CoA ligase-like domain and a T domain. The acyl CoA ligase domain incorporates a myristic or myristoleic acid group as the first residue of pyoverdine (Drake and Gulick, 2011; Hannauer et al., 2012). The three other modules of PvdL add an L-Glu, D-Tyr and L-diaminobutyrate (L-Dab) residue before the peptide is passed to the NRPS enzymes required for synthesis of the pyoverdine side-chain. In *P. aeruginosa* PAO1 these enzymes are PvdI, PvdJ and finally PvdD. PvdD governs incorporation of the two L-Thr residues at the C-terminus of the linear pyoverdine peptide (Ackerley et al., 2003), after which, the TE domain releases pyoverdine via intramolecular cyclisation. The released pyoverdine (Figure 1-8C) is transported into the periplasm. In the periplasm, the four residues added by PvdL are processed into the mature chromophore, in which the myristate and L-Glu are replaced by an acyl group such as succinamide, and the other two residues are modified. Mature pyoverdine is secreted into the extracellular environment by the ATP-dependent efflux pump PvdRT-OpmQ (reviewed in Schalk and Guillon (2013) and Visca et al. (2007)).

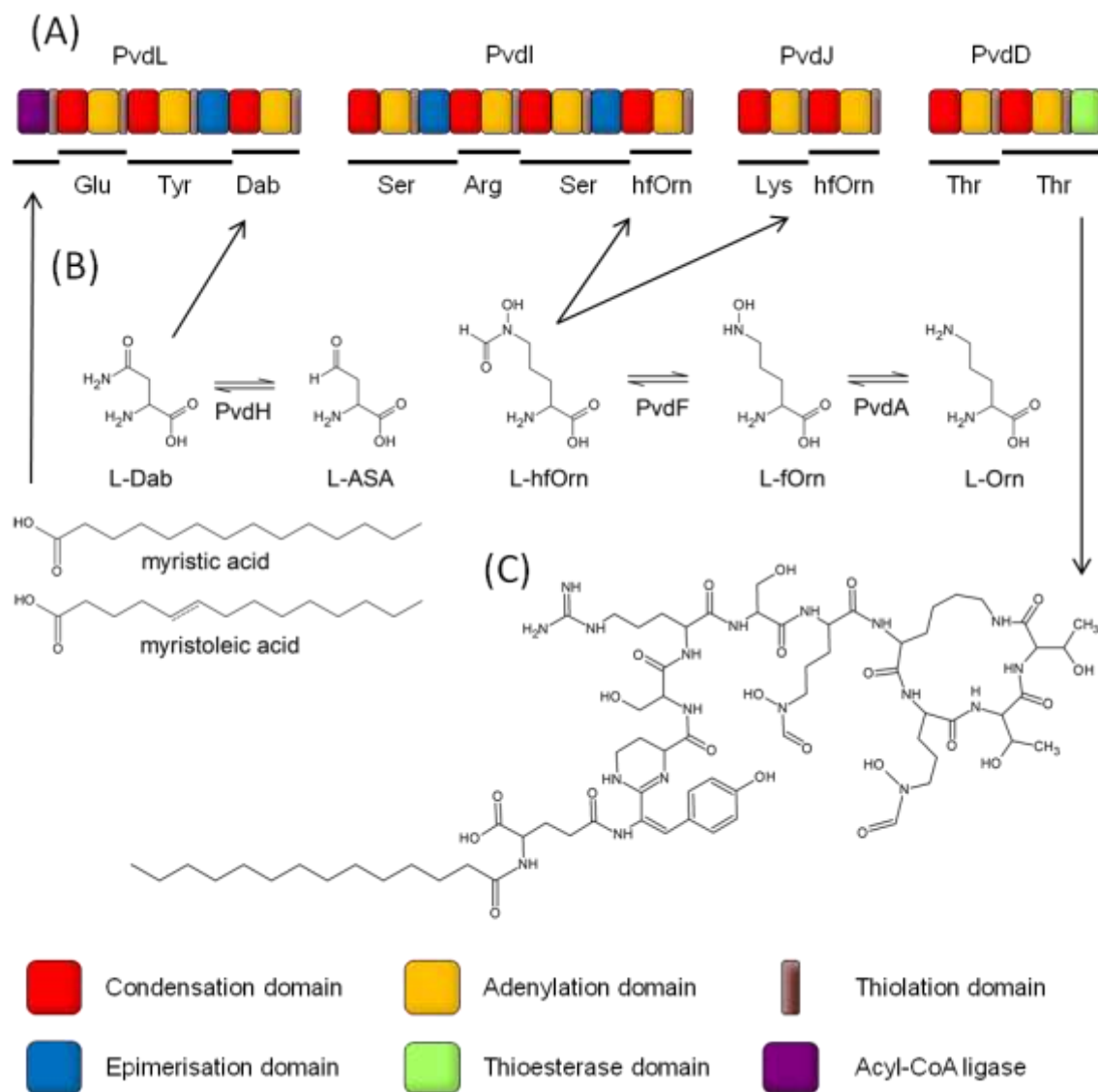


Figure 1-8. Synthesis of the pyoverdine precursor. (A) Pyoverdine is synthesised by the four NRPS enzymes PvdL, PvdI, PvdJ and PvdD. The domain architecture of each enzyme is indicated by the key at the bottom of the figure. The module boundaries are delineated by lines under each enzyme, and the substrate specified by the A domain for each module is listed below these lines. (B) Non-proteinogenic residues incorporated into pyoverdine. The first module of PvdL adds a myristic or myristoleic acid residue. PvdH converts L-Asp  $\beta$ -semialdehyde (L-ASA) into L-Dab that is incorporated into pyoverdine. PvdA and PvdF convert L-Orn into L- $N^5$ -formyl- $N^5$ -hydroxyornithine (L-hfOrn) that is incorporated into pyoverdine. (C) The pyoverdine precursor is released by the TE domain of PvdD. Figure adapted from Schalk and Guillon (2013) and Visca et al. (2007).

Once secreted, pyoverdine can bind iron and is then taken up by a membrane bound receptor, FpvA. In total, over 100 pyoverdines have been identified (Meyer et al., 2008; Schalk, 2008) with the sequence determined by mass spectrometry for more than 60 different pyoverdines (Meyer et al., 2008). The pyoverdines contain the same

chromophore but differ greatly in their peptide chains, which have diverged through genetic recombination and duplication events (Ravel and Cornelis, 2003; Smith et al., 2005). The ability to take up pyoverdines can be highly strain specific, such that a strain may not be able to use pyoverdine synthesised by other strains (Buyer and Leong, 1986; Meyer et al., 1997, 2002, 2008). A reason suggested for the specificity is that pyoverdine has diversified to block other strains from using the siderophore (Schalk and Guillon, 2013; Visca et al., 2007). Despite the specificity, heterologous uptake can occur with pyoverdines containing similar peptide chains (Schalk, 2008). In particular, *P. aeruginosa* PAO1 has been shown to exhibit promiscuity (Meyer et al., 1997, 1998, 1999).

#### 1.4.1 Reasons to use pyoverdine as a model for domain substitution

Many domain substitution studies detect products using mass spectrometry without quantification of products. This is because products can often be difficult to detect and/or purify, especially when synthesised at low levels. The main feature of pyoverdine that makes it useful for domain substitution studies is its ease of detection and ability to be quantified, based on absorbance and fluorescence of the chromophore. The large variation in peptide chains of pyoverdines indicates that changes to the variable peptide are unlikely to interfere with the absorbance and fluorescence of the invariant chromophore. Although specificity towards pyoverdines has been previously observed, *P. aeruginosa* PAO1 is useful for this study due to its ability to take up heterologous pyoverdines. In particular, it has been suggested the two terminal Thr residues are unlikely to be involved in recognition by FpvA in *P. aeruginosa* PAO1 (Meyer et al., 1999). The diversity of pyoverdines not only indicates that nature has repeatedly surmounted existing constraints to domain recombination, but it also provides a rich source of related domains for use in domain substitution experiments.

In this thesis, NRPS domain substitution variants were expressed *in vivo* and pyoverdine production was monitored. There are several advantages to measuring *in vivo* production of pyoverdine rather than the *in vitro* kinetics. Kinetic assays of NRPS domains generally measure a single reaction in isolation. However modifications can affect reactions not measured (Zhang et al., 2013) or modified NRPS enzymes may function *in vivo* despite loss of function for *in vitro* assays (Roche and Walsh, 2003). Moreover, purification of pyoverdine NRPS domains, previously attempted by Dr Owen, found them to become insoluble when purifying (Owen, 2010). Finally, presumably due to the size of NRPS

pathways, only small model systems have been used for *in vitro* studies. We are most interested in being able to modify modules in the context of entire pathways, and as such *in vivo* production would appear to be the most important metric.

#### 1.4.2 Previous work on the pyoverdine system

This study improves and extends early work on A domain substitutions in the pyoverdine synthesising NRPS enzyme PvdD (Ackerley and Lamont, 2004). PvdD contains two Thr specifying modules that incorporate the final two (C-terminal) residues into pyoverdine (Ackerley et al., 2003). The previous work using PvdD successfully generated attenuated pyoverdines by in-frame deletion of the first of the two modules of *pvdD* (Ackerley and Lamont, 2004). However, attempts to generate modified pyoverdine by domain substitution experiments were unsuccessful. When five different A domains were substituted into the first Thr specifying module of PvdD, the two Thr specifying A domain substitutions yielded functional recombinant PvdD variants, whereas the three enzymes containing non-Thr specifying A domains were all inactive. This result was highly successful in terms of domain substitution and is consistent with the A domain substitutions being limited by C domain acceptor site proof-reading constraints. In contrast, when the corresponding C-A domains were substituted, none of the substitutions were functional (Ackerley and Lamont, 2004).

The reasons for C-A domain substitutions being non-functional in PvdD are unclear. One possibility is the C-A domain substituted constructs from this previous study were inactive due to the inability of the introduced domains to communicate appropriately with the remainder of the pyoverdine biosynthetic machinery. Another possible explanation raised to explain this lack of activity was that C domains may sometimes exhibit donor site proofreading after all, such that the introduced C-domain might prove incompatible with the incoming residue supplied by the upstream NRPS module, or that a newly introduced amino acid substrate might not be recognised at the donor site of the C domain immediately downstream (Ackerley and Lamont, 2004). There is evidence of some donor site specificity towards the terminal residue of peptides (Stein et al., 2005), and additional evidence shows starter C domains (which join fatty acid residues to the first amino acid in a non-ribosomal peptide) can exhibit strong donor site specificity toward fatty acid substrates (Kraas et al., 2012). To eliminate the confounding possibility that introduction of a novel amino acid into a growing peptide chain might cause donor-site incompatibility

with the C domain immediately downstream, in this study we chose to focus on domain substitutions into the second module of PvdD, which (as the terminal pyoverdine synthetase; Figure 1-8A) has no additional C domain located downstream. An additional advantage is that the second module of PvdD condenses substrates *in cis*, i.e. attached to T-domains within the same enzyme, and we reasoned that C-A domain substitutions into this module would therefore be less likely to disrupt critical interactions between PvdD and the upstream NRPS, PvdJ.

## 1.5 Aims of this study

The overall aim of this thesis was to rigorously define the genetic constraints on domain substitution in PvdD.

This was first addressed in Chapter 3 by testing whether A domain substitutions are highly functional in pyoverdine NRPS enzymes, but limited by acceptor site substrate specificity. Due to the low number of previously published A domain substitution studies downstream to C domains, it was unclear whether the primary factor restricting the functionality of A domain substituted NRPS constructs is C domain acceptor site proof-reading or disruption of critical C/A domain interactions.

We then sought to determine, using the same genetic templates, whether C-A domain substitutions could produce novel pyoverdines. Previous attempts to change residues in pyoverdine by domain substitution were not successful. Chapter 4 examines whether new pyoverdines can be produced by domain substitution and found that new pyoverdines could sometimes be produced. The function of some C-A domain substitutions is important, as it enables PvdD to be used to understand why other C-A domain substitutions become non-functional.

An emerging view is that conformational changes of the T domain and its interactions with other domains are vital for NRPS function. Chapter 5 explores whether keeping the T/C domain junction intact improves activity of domain substitutions, and examines the general portability of T domains in PvdD. This work found T domains are highly portable in PvdD when certain constraints are maintained.

Due to no C domain structure having been solved in the presence of a substrate, little is known about the residues involved in the substrate specificity within the acceptor site of

the C domain. Chapter 6 aimed to isolate the region of the C domain vital for acceptor site specificity, and developed a plasmid useful for the directed evolution of NRPS enzymes in *P. aeruginosa* PAO1.



## Chapter 2: **Methods**

### 2.1 **Chemical, reagents, media and enzymes**

All chemicals, reagents and media were obtained from Sigma-Aldrich (St Louis, MO, USA) and Thermo Fisher Scientific (Waltham, MA, USA), unless otherwise noted. Isopropyl  $\beta$ -D-1-thiogalactoside (IPTG) was obtained from Bioline (London, UK). EDDHA (ethylenediamine-N,N'-bis(2-hydroxyphenylacetic acid)) was supplied by Professor Ian Lamont from University of Otago (Dunedin, NZ).

Enzymes were sourced from several companies. Phusion™ high-fidelity DNA polymerase was purchased from Finnzymes (Espoo, Finland) and New England Biolabs (Ipswich, MA, USA). Biomix™ red was obtained from Bioline. Restriction enzymes were supplied by New England Biolabs. T4 DNA ligase was supplied by Bioline and Invitrogen (Carlsbad, CA, USA).

Unless otherwise stated, all solutions were made in ultrapure water. Ultrapure water was prepared by distillation followed by deionisation to a minimum conductivity of 18 M $\Omega$ ·cm at 25 °C using a Barnstead™ Easypure™ II deioniser (Thermo Fisher Scientific).

### 2.2 **Oligonucleotide primers**

DNA oligonucleotide primers were supplied by Integrated DNA Technologies (Coralville, IA, USA) as lyophilised powders. Stock solutions were created by dissolution of primers in TE buffer (10 mM Tris-Cl pH 8.0, 1 mM ethylenediaminetetraacetic acid [EDTA]) to an oligonucleotide concentration of 100  $\mu$ M. Working stocks of primers were made by diluting the stock solution to 10  $\mu$ M in ultrapure water. All primers were stored at -20 °C. The sequences of primers used in this study are shown in Table 2-1.

Table 2-1. Oligonucleotide primers used in this study

Function Name	Sequence (5' → 3'; restriction sites underlined)
Primers for pSMC	
CATfwd	CCCCGAATTCCATATGCAAGCACTCATAGAGAAGGTG
CATrev	CCCCACTAGTCAATCCCTGGGCGAACGC
TTefwd	CGCCCAGGGATTGTCTAGAGCGGCCGCTCGCAGCAGGCCTATCGAGCGC
TTerev	CCCCGAGCTCTCAGCGCCCGGCACGCTCCAGG
Primers for pSMA	
Thr-WT_Cafwd	GGGGTCTAGAACGACGGATGCGGTCTCGACGA
CATCrev	ATAGGCCTGCTGCGACGCGGCCGCTCTAGAATCAGCGCAACCGCCTGT
Forward primers for C-A domain substitutions	
Thr-A_Cafwd	CCCCTCTAGAGGTGTCAATCTCTTCGAG
Ser-B_Cafwd	CCCTCTAGAGCGGCGAGCGCTGCCCCTGTGCT
Lys-C_Cafwd	CCCTCTAGATTTCGCCCCCTGCCGATCCCGC
Thr-D_Cafwd	CCCCTCTAGAACGACGGATGCGGTCTCGACGATACCGCTTGCCGATCGG CAGGAAGATATCTACCCGCTGTGCG
Thr-E_Cafwd	CCCCTCTAGAAATCTCTACGGGGTCACACCGA
Ser-F_Cafwd	CCCCTCTAGAGGGCAGGGCAATGCTGCG
Asp-G_Cafwd	CCCTCTAGAGATTTTCGCGCTGTTGCCGATCGCG
Gly-H_Cafwd	CCTCTAGAAACCTCTACGGGGTGACGCGCAT
Ser-I_Cafwd	CCCTCTAGAGACTTTTCCCGGTTTCCGATTCC
hfOrn-J_Cafwd	CCCCTCTAGACAGGCCCGAGGCGCG
Forward primers for A domain substitutions	
Thr-WT_Afwd	CCCCTCTAGACGAACAACGGTTGAGCTA
Thr-A_Afwd	CCCCTCTAGACGAACAACGGTTGAGCTA
Ser-B_Afwd	CCCCTCTAGAGGAACGCCTGGACTACGCCGAG
Lys-C_Afwd	GGTCTAGAGGGGCAGGCCTTGAGCTACGCC
Thr-D_Afwd	CCCCTCTAGATGGCGAGCAATTGAGATA
Thr-E_Afwd	CCCCTCTAGATGGCGAGCAATTGAGCTA
Ser-F_Afwd	CCCCTCTAGAGGCTGAACAACCTGAGCTA
Asp-G_Afwd	CCCTCTAGACGATGGCTCGCTCAGTTACGGC
Gly-H_Afwd	CCCTCTAGACGAGCAGACCTTGAGCTACGCCG
Ser-I_Afwd	CCCTCTAGATGCAACCACGCTGACCTACGCCC
Reverse primers for C-A and A domain substitutions	
Thr-WT_Rev	CCCCGCGGCCGCATCCGGTTGCGGCAACGCCTGC
Thr-A_Rev	CCCCGCGGCCGCATCCGGTTGCGGCAACGCCTGC
Ser-B_Rev	CCCCGCGGCCGCTTGCGGTGCGGCAGCGCCTTG
Lys-C_Rev	CCCCGCGGCCGCATCCGGCTTCGGCAGCGCCCGA
Thr-D/E_Rev	CCCCGCGGCCGCATCCGGCGCGGGCAGCGCCTTG
Ser-F_Rev	CCCCGCGGCCGCGTCTGGTGTGCGCAGGGC
Asp-G_Rev	CCCCGCGGCCGCGTCCGGCAGCGGCAATCTGGCC
Gly-H_Rev	CCCCGCGGCCGCGTCAGGCGCCGGCAGCGCCTTG
Ser-I_Rev	CCCCGCGGCCGCATCCGGCAATGGCAAGGCCTTG
hfOrn-J-Rev	GCGACGCGGCCGCATCCGGGCGCGGCAGG
T2 domain substitutions	
TE_Fwd	CCCCGCGGCCGCCCCACTAGTAATCTCTACGGGGTCACACCGA
TTerev	CCCCGAGCTCTCAGCGCCCGGCACGCTCCAGG
T2_Fwd	CAACCGGATGCGGCCGCGTCGCAACAGGCCTATCGAGCGCCC
T2_Rev	CATCCGTCGTGCTAGCCAATCCCTGGGCGAACGC
T1_Fwd	CAACCGGATGCGGCCGCGTCGCGAGCAGGCCTATCGAGCGC
T1_Rev	CATCCGTCGTGCTAGCGATGCATGCCGCTGTTCCGC
T-C-A domain substitutions	
TCASer-B_Fwd	CCCCGCGGCCGCCCCGCGCCAGGCCGGAGAGCCT
TCASer-F_Fwd	CCCCGCGGCCGCGCTTGCGAGGCGGCCTACATCGCT
TCAhfOrn-J_Fwd	CGGATGCGGCCGCGGTGCAACAGGCCTGGCAG
T1 domain substitutions	
CATTE_Fwd	CCCCGCGGCCGCCCCACTAGTACGACGGATGCGGTCTCGACGA
C1F	CCCCGCGGCCGCCCCGCGCCAGGCCGGAGAGCCT

C1R	CCCCGCTAGCCAGGGAAGCGGCGAAGTCTGCC
C2F	CCCCGCGGCCGCGTTGCGAGGCGGCCTACATCGCT
C2R	CCCCGCTAGCCGCGAGCCTGGGCGAAAGCTGCC
C3F	CCCCGCGGCCGCGCCGCGCCAGCGCCCGTACCAGGCG
C3R	CCCCGCTAGCCACCCGCTCGGCGAAGCCCGCC
C4F	CCCCGCGGCCGCGGCGCGCCAGGCCTAC
C4R	CCCCGCTAGCGCAAGCGGCTGCGAACG
C5F	CCCCGCGGCCGCGTTGCGCACAGGTCATGTGG
C5R	CCCCGCTAGCCAGGGCCTGCACGAAGTCAC
C6F	CCCCGCGGCCGCGGTGCAACAGGCCTGGCAG
C6R	CCCCGCTAGCCAGTGCGGCGACAAACC
E1F	CCCCGCGGCCGCGGCGGGGCGAGACGCATGT
E1R	CCCCGCTAGCGGCGACTCGCGCCAG
E2F	CCCCGCGGCCGCGGTCAAACAGCGTTACACCGC
E2R	CCCCGCTAGCTGCCACCGAGGCCAGCTG
E3F	CCCCGCGGCCGCGGCGGCAAGGCTTACGTA
E3R	CCCCGCTAGCGGCCACCTGGGCCAAGGC
E4F	CCCCGCGGCCGCGCCACAGCAGGCCTTTGTGCT
E4R	CCCCGCTAGCCGCCACCCGCGCCAG
Ct1F	CCCCGCGGCCGCGTTGCGAGCAGGTCTACGTGGCG
Ct1R	CCCCGCTAGCGGCAAGCTCGGCATAGGCTTGC
Ct2F	CCCCGCGGCCGCGTTGCGAGCAGGTCTACGTGGCG
Ct2R	CCCCGCTAGCGACGCCGTGGCAGAACTCACCC
Ct3F	CCCCGCGGCCGCGTTGCGAGGCGGCCTACATCGCT
Ct3R	CCCCGCTAGCCAGCGTGGCGCAGAAGTCCGCC
Ct4F	CCCCGCGGCCGCGCTCAGCAGCGCTACCAGGCA
Ct4R	CCCCGCTAGCGACGGCCTCGCTGAACGCTCCC
TE1F	<i>same as T2_Fwd</i>
TE1R	<i>same as T2_Rev</i>
TE2F	CCCCGCGGCCGCGTCCAGCAAGGCCACGTC
TE2R	CCCCGCTAGCCATGAAATCGGCCAGCTCACC
TE3F	CCCCGCGGCCGCGGCGTTGCGTGAACATGTGGCAC
TE3R	CCCCGCTAGCCGCCACGGCGGTGAGCTC
TE4F	CCCCGCGGCCGCGGGCGCGGACGCTTATCAGG
TE4R	CCCCGCTAGCCAGGCAGTTGACCTGGCCGCGT
Creation of pDEC and variant C domains	
Thr-WT_Afwd2	CCCCACTAGTCCCCGTCGACGAACAACGGTTGAGCTACG
Thr-WT_Rev2	<i>same as Thr-WT_Rev</i>
Thr-WT_Cfwd	<i>same as Thr-WT_CAfwd</i>
Thr-WT_Crev	CCCCCCTCGAGGATCAGCGCAACCGCCTG
Lys-C_Cfwd	<i>same as Lys-C_CAfwd</i>
Lys-C_Crev	CCCCCCTCGAGCGCCAGTGCTTGCGCCTG
Lys-C_C <sub>12</sub> Rev	CCCCGTCGACCCCCACTAGTTGCTCGAACGGCAG
Sild2CSpeI fwd	CCCCACTAGTGGAAGCCTTGCAGC
DE_OL_1F	CGATCCAGTACGCGGACTAC
DE_OL_1R	GTAGTCCGCGTACTGGATCG
DE_OL_2F	CAGGACCTGCCGTTTCGAG
DE_OL_2R	CTCGAACGGCAGGTCCTG
Thr-WT_C <sub>12</sub> Rev	CCCCACTAGTCCCCGTCGACGAACAACGGTTGAGCTACG
SilJC1SpeI Fwd	CCCCACTAGTGAGGCTTTGCAGC
Hom_LyslpFwd	GTGCTGTTCAACCATCAGGCCGACTCTCGTTCGG
Hom_LyslpRev	ATCCAGCGTCAGGTCGAAGGCCACGGAGCTGCTC

---

## 2.3 Plasmids

Table 2-2. Plasmids used in this study

Plasmid	Short description	Source
pSW196	Integration proficient vector; P <sub>BAD</sub> promoter.	(Baynham et al., 2006)
pSMC	pSW196 based staging plasmid allowing the substitution of CA domains into <i>pvdD</i> .	This study
pSMA	pSW196 based staging plasmid allowing the substitution of A domains into <i>pvdD</i> .	This study
pST2	pSW196 based staging plasmid allowing the substitution of T domains into the second module of <i>pvdD</i> .	This study
pTCA	pSW196 based staging plasmid allowing the substitution of T-C-A domains into the second module of <i>pvdD</i> .	This study
pST1	pSW196 based staging plasmid allowing the substitution of T domains into the second module of <i>pvdD</i> .	This study
pUCP22	Non-integrative plasmid for expression in <i>P. aeruginosa</i> PAO1	(West et al., 1994)
pUCBAD	New vector created by adding the P <sub>BAD</sub> promoter from pSW196 into pUCP22	This study
pDEC	C domain substitution plasmid constructed from pUCBAD	This study
pTRN	Substitution plasmid for changing the third variable region of the second module of <i>pvdD</i>	This study

## 2.4 Bacterial strains

Table 2-3. *E. coli* strains used in this study

Strain	Relevant characteristics *	Source
DH5α λpir	<i>supE</i> 44 <i>DlacU</i> 169 (Ø80 <i>lacZ</i> DM5) <i>hsdR</i> 17 λ(pir)	Invitrogen
BL21	F <sup>-</sup> <i>ompT</i> <i>gal</i> <i>dcm</i> <i>lon</i> <i>hsdS</i> <sub>B</sub> (r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) λ(DE3)	Novagen
BL21 Δ <i>entD</i>	<i>entD</i> deletion strain	(Owen et al., 2012)

\* A list of standard abbreviations can be found at

[http://openwetware.org/wiki/E.\\_coli\\_genotypes#Nomenclature\\_.26\\_Abbreviations](http://openwetware.org/wiki/E._coli_genotypes#Nomenclature_.26_Abbreviations)

Table 2-4. *Pseudomonas* strains used in this study\*

Strain	Relevant characteristics	Source
<i>P. aeruginosa</i> PAO1	Wild type	Laboratory stocks
<i>P. aeruginosa</i> PAO1 Δ <i>pvdD</i>	<i>pvdD</i> deletion mutant, Pvd <sup>-</sup>	(Ackerley and Lamont, 2004)
<i>P. syringae</i> pv. <i>phaseolicola</i> 1448A	Wild type	Laboratory stocks
<i>P. putida</i> KT2440	Wild type	Laboratory stocks
<i>P. fluorescens</i> SBW25	Wild type	Prof. Paul Rainey

\* Domain substitution strains were created using *P. aeruginosa* PAO1 Δ*pvdD* and are not listed in this table for clarity.

## 2.5 Bacterial growth and maintenance

Media were prepared in ultrapure water and autoclaved at 121 °C for 30 min. Storage of media was at room temperature.

### 2.5.1 Liquid media

#### **Luria broth (LB)\***

Tryptone	10	g·L <sup>-1</sup>
Yeast extract	5	g·L <sup>-1</sup>
NaCl	5	g·L <sup>-1</sup>

\* Supplied as a premixed powder

#### **King's B (KB)**

Bacto-peptone	20	g·L <sup>-1</sup>
K <sub>2</sub> HPO <sub>4</sub>	1.5	g·L <sup>-1</sup>
Glycerol	1	% (v/v)
MgSO <sub>4</sub>	6.1	mM*

\* Added post-autoclave from a filter sterilised 1 M solution

#### **Succinate media**

K <sub>2</sub> HPO <sub>4</sub>	6	g·L <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	3	g·L <sup>-1</sup>
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1	g·L <sup>-1</sup>
Succinate	4	g·L <sup>-1</sup>
MgSO <sub>4</sub>	6.1	mM*

\* Added post-autoclave from a filter sterilised 1 M solution

#### **M9 succinate\*\*\***

Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O	6.78	g·L <sup>-1</sup>
KH <sub>2</sub> PO <sub>4</sub>	3	g·L <sup>-1</sup>
NH <sub>4</sub> Cl	1	g·L <sup>-1</sup>
NaCl	0.5	g·L <sup>-1</sup>
Succinate	4	g·L <sup>-1</sup>
MgSO <sub>4</sub>	6.1	mM***

\* M9 salts provided as a premixed powder

\*\* Adjusted to pH 7.0 using NaOH prior to autoclaving

\*\*\* Added post-autoclave from a filter sterilised 1 M solution

### 2.5.2 Solid media

To solidify media, 1.5 % (w/v) agar was added prior to autoclaving. Media was allowed to cool to less than 50 °C before addition of supplements. After supplements were added, media was poured aseptically into 90 mm Petri dishes and incubated at room temperature until set. Agar plates were stored at 4 °C.

### 2.5.3 Media supplements

Antibiotic stock solutions were prepared at 1000x the concentration used in media. Solutions were filter sterilised and stored at -20 °C. Table 2-5 shows the media concentrations of antibiotics used in this study. Other supplements were IPTG and EDDHA. The concentration of IPTG in the stock solution was 100 mg·mL<sup>-1</sup>. EDDHA was prepared to a final stock solution concentration of 200 mg·mL<sup>-1</sup> with NaOH added to facilitate solubility.

Table 2-5. Concentrations of antibiotics in media

Antibiotic	<i>E. coli</i> (µg·mL <sup>-1</sup> )	<i>P. aeruginosa</i> PAO1 (µg·mL <sup>-1</sup> )
Ampicillin	100	-
Kanamycin	50	-
Spectinomycin	50	-
Gentamycin	-	50
Tetracycline	15	100

### 2.5.4 Growth and storage

Unless otherwise stated, bacterial strains were grown in LB medium. Cultures were incubated at 37 °C for *E. coli* and *P. aeruginosa*, and 30 °C for *P. putida*, *P. syringae* and *P. fluorescens*. Liquid cultures were aerated in an orbital shaker at 200 - 300 rpm. Strains were stored for short periods of time on agar plates at 4 °C. For long term storage, a liquid culture was grown overnight, mixed in a 1:1 ratio with 80 % (v/v) glycerol and stored at -80 °C.

## 2.6 General molecular biology methods

### 2.6.1 Isolation and purification of DNA

Genomic, plasmid and linearised DNA were isolated and purified using the kits in Table 2-6 according to the manufacturer's instructions. DNA purified using a Zymoclean™ Gel DNA Recovery kit (Zymo Research; Irvine, CA, USA) was subsequently cleaned with a DNA Clean & Concentrator™ kit (Zymo Research) to remove traces of ethidium bromide. This increased the efficiency in downstream reactions. All DNA was stored at -20 °C in ultrapure water.

Table 2-6. Kits used for DNA purification

DNA	Name of kit	Supplier
Genomic Plasmid*	Qiagen DNeasy™ Blood & Tissue	Qiagen; GmbH, Germany
	Zyppy™ Plasmid Miniprep	Zymo Research; Irvine, CA, USA
	Geneaid High-Speed Plasmid Mini	Geneaid Biotech; Sijhih City, Taiwan
Linear**	DNA Clean & Concentrator™	Zymo Research
	Zymoclean™ Gel DNA Recovery	Zymo Research

\* Multiple kits were used for the same purpose

\*\* For the purposes of this chart, linear refers to DNA linearised by restriction digest or PCR amplification

## 2.6.2 Confirmation of DNA products

### 2.6.2.1 Agarose gel electrophoresis

To qualitatively determine the size of DNA fragments, samples were electrophoresed on 1 % (w/v) agarose gels (1 % (w/v) agarose, 1x TAE buffer [40 mM Tris pH 8.0, 20 mM acetic acid, 1 mM EDTA], 1  $\mu\text{g}\cdot\text{mL}^{-1}$  ethidium bromide). Prior to loading, DNA samples were mixed with 5x DNA loading buffer (Bioline). HyperLadder™ 1 (Bioline) was added to an adjacent well for size determination. Gels were electrophoresed at 100 - 150 V for 30 - 45 min, and then viewed using an ultra-violet (UV) transilluminator.

### 2.6.2.2 DNA quantification

Following purification, DNA samples were quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

### 2.6.2.3 DNA Sequencing

Plasmids created in this study were sequenced by Macrogen Inc. (Seoul, South Korea).

## 2.6.3 PCR protocols

### 2.6.3.1 Standard reactions

For plasmid construction, DNA was amplified with Phusion™ high-fidelity polymerase. For colony screening and diagnostic purposes Biomix Red™ was used. The standard PCR protocols for Phusion™ and Biomix Red™ are shown in Table 2-7 and Table 2-8, respectively. These PCR protocols were optimised experimentally to maximise yield and specificity.

Table 2-7. PCR protocol used for Phusion™ polymerase

Temperature	Time	
98 °C	1 min	
98 °C	15 s	30 cycles
58 °C	30 s	
72 °C	30 s·kb <sup>-1</sup> *	
72 °C	5 min	

\* kb = kilobase

Table 2-8. PCR protocol used for Biomix Red™ polymerase

Temperature	Time	
95 °C	5 min	
95 °C	20 s	15 cycles
56 °C	30 s	
72 °C	1 min·kb <sup>-1</sup>	
95 °C	20 s	15 cycles
52 °C	30 s	
72 °C	1 min·kb <sup>-1</sup>	
72 °C	5 min	

### 2.6.3.2 Overlap Extension PCR

Overlap Extension PCR is a two-step process to anneal DNA fragments without using restriction sites. Firstly, two DNA fragments were amplified by Phusion™ high-fidelity polymerase. The fragments were designed such that the 3' end of the upstream fragment and 5' end of the downstream fragment contained 12 - 20 basepair (bp) of homology. This region of homology served as a primer for a subsequent reaction, such that the downstream fragment would bind to the upstream fragment and act as a reverse primer and vice versa. The second reaction used Biomix Red™ and proceeded for five cycles with only the two fragments to enable the full length product to form. Then the forward and reverse primers for the full length product were added to allow exponential amplification of the annealed fragments for 30 cycles. The protocol for the second reaction is shown below in Table 2-9.

Table 2-9. Overlap extension protocol used for Biomix Red™ polymerase

Temperature	Time	
95 °C	5 min	
95 °C	20 s	5 cycles
56 °C	30 s	
72 °C	1 min·kb <sup>-1</sup>	
Primers added		
95 °C	20 s	30 cycles
60 °C	30 s	
72 °C	1 min·kb <sup>-1</sup>	
72 °C	5 min	

### 2.6.3.3 Site-directed mutagenesis

Site-directed mutagenesis was used to alter the sequence of a gene already within a plasmid. The protocol was similar to the QuikChange™ Site-Directed Mutagenesis Kit (Agilent Technologies; Santa Clara, CA, USA), except an amplicon was used instead of primers due to the large region to be changed. In brief, a short piece of DNA was amplified that had homology on each end to where it was to be inserted into a plasmid. This amplicon was then used as the primer in a PCR reaction to amplify the plasmid. The resulting product was cleaned and restriction digested overnight using DpnI to remove any of the original plasmid. Following restriction digest, the PCR amplified plasmid was transformed into *E. coli*.

### 2.6.3.4 STaggered Elongation Process PCR

For STaggered Elongation Process (STEP) PCR based recombination (Zhao et al., 1998), two fragments of DNA containing moderate levels homology were amplified using Phusion™. These were purified and eluted in ultrapure water. A second reaction used both purified fragments as template and the forward primer for one fragment and the reverse primer for the other fragment. In the PCR protocol for the second reaction (Table 2-10), a low annealing temperature and short annealing/extension step was used to cause DNA extension to halt before complete extension. This meant that, in homologous

regions, incomplete products could bind either sequence and act as a primer. The result of this template switching was amplicons consisting of sequence from both gene fragments.

Table 2-10. Protocol for StEP PCR recombination

Temperature	Time	
98 °C	1 min	
98 °C	20 s	80 cycles
56 °C	5 s	
72 °C	5 min	

#### 2.6.4 Restriction enzyme digest

Restriction digests were carried out according to the manufacturer's protocol. After restriction digest, linearised DNA was cleaned (Section 2.6.1) and stored at -20 °C. For difficult to digest DNA, digests were incubated overnight. The next day, additional enzyme, buffers and ultrapure water were added, doubling the volume. This was left to digest for at least 4 h before purification of digested DNA.

#### 2.6.5 Ligation

T4 DNA Ligase was used for ligation of DNA according to manufacturer instructions. Ligations typically used between a 1:1 and 3:1 molar insert to vector ratio and were either transformed as is or purified with a DNA Clean & Concentrator™ kit (Section 2.6.1) before transformation.

### 2.7 Bacterial transformation

#### 2.7.1 *E. coli* transformation by the calcium chloride method

To make *E. coli* competent for transformation, a culture of DH5α was grown in liquid LB until stationary-phase. The culture was diluted to an OD<sub>600</sub> of 0.05 in LB supplemented with 10 mM MgCl<sub>2</sub> and grown until an OD<sub>600</sub> of 0.4 was reached. From this point the cells were kept on ice as much as possible. The culture was centrifuged for 20 minutes at 4,000 rpm, 4 °C to pellet the bacteria. The pellet was resuspended in 1 vol of TFB I (50 mM MnCl<sub>2</sub>, 10 mM CaCl<sub>2</sub>, 30 mM CH<sub>3</sub>CO<sub>2</sub>K, 15 % (v/v) glycerol) and incubated on ice for 2 h. The bacteria was centrifuged, the supernatant discarded, and the pellet

resuspended in 0.1 vol of TFB II (10 mM NaMOPS pH 7, 75 mM CaCl<sub>2</sub>, 10 mM KCl, 15 % (v/v) glycerol). Following resuspension, 100 µL volumes were aliquoted into Eppendorf tubes, snap-frozen and stored at -80 °C.

For each transformation, a 100 µL aliquot of competent cells was defrosted on ice. DNA was added in a volume of less than 10 µL and the cells were incubated on ice for 20-30 min. The sample was heat shocked at 42 °C for 2 min, then incubated on ice for 2 min. A volume of 1 mL of LB was added and the cells recovered at 37 °C, 250 rpm for 40 min. Recovered cells were spread on solid media supplemented with the relevant antibiotics and incubated at 37 °C overnight.

Colonies with correct plasmid sequences were screened for using PCR based methods and confirmed with restriction digest analysis and DNA sequencing.

### 2.7.2 *P. aeruginosa* transformation by electroporation

For the transformation of *P. aeruginosa*, a culture was grown in liquid media overnight until late stationary-phase. For each transformation, a total of 3 mL of culture was aliquoted into two eppendorf tubes. These were centrifuged at 4000 rpm for 20 min. The supernatant was discarded and the pellet resuspended in 1 mL of 300 mM sucrose. The centrifugation and resuspension step was repeated. Next, the cells were centrifuged and both pellets resuspended together in 100 µL of 300 mM sucrose. Competent *P. aeruginosa* cells were prepared at room temperature and used immediately for transformation.

For each plasmid to be transformed, a volume of less than 5 µL of purified plasmid was added to a 100 µL aliquot of competent cells. This step was modified for transforming integration plasmids such as pSW196, where approximately 10 µg of plasmid DNA was required. To achieve this amount of DNA, when competent cells were being prepared, 50 µL of 600 mM sucrose was mixed with 50 µL of purified plasmid and used for the final resuspension. The cells containing added plasmid DNA were next added to a 2 mm electroporation cuvette. Cells were electroporated using a GenePulser Xcell™ electroporation system (Bio-Rad Laboratories, Inc.; Hercules, CA, USA) with the default settings for *P. aeruginosa*. Immediately following electroporation, 1 mL of LB was added to the electroporation cuvette. The culture was poured into a 15 mL falcon tube and recovered at 200 rpm, 37 °C for between 2 to 3 h. Recovered transformations were spread

on solid media plates containing the necessary antibiotics and incubated at 37 °C overnight.

Integrating plasmids were constructed in *E. coli* and confirmed by sequencing prior to transformation. Non-integrative plasmids were screened for using PCR, restriction digest analysis and DNA sequencing. As extraction of plasmid DNA from *P. aeruginosa* resulted in low yields, plasmids were transformed into and extracted from *E. coli* to use for restriction digest analysis and DNA sequencing.

## 2.8 Creation of the domain substitution plasmids

### 2.8.1 Creation of the plasmids pSMC and pSMA

To create pSMC, the first module of *pvdD* containing C-A-T domains was PCR amplified using the primers CATfwd and CATrev. The amplicon was restriction digested and ligated into pSW196 using the 5' EcoRI and 3' XbaI restriction sites added during PCR by the primers. The location of the XbaI restriction site was chosen based on studies showing the highly variable stretch of amino acids between the T domain and C domain is tolerant of change (Doekel et al., 2008; Mootz et al., 2000). The location is 37 residues downstream to the conserved Ser residue of the T domain from the first module (Figure 3-2). The restriction site altered the coding sequence from ER→TR. Due to being within the linker region it was predicted this change would have minimal affect on enzyme function. After confirming integration of the first module, the T-TE domains from the second module of *pvdD* were amplified with primers TTeFwd and TTeRev, and then ligated downstream to the first module using the introduced 5' NotI and 3' SacI restriction sites. The location of the NotI site was chosen based on domain substitutions being previously successful in the first module of *pvdD* (Ackerley and Lamont, 2004). The NotI restriction site was located by aligning the first and second modules of *pvdD* and selecting the region in the second module equivalent to where the restriction site was added previously. The amino acid sequence in this location was identical between the first and second modules. At this point the plasmid pSMC was constructed and contained the C-A-T domains from the first module and T-TE domains from the second module of *pvdD* separated by a SpeI and NotI restriction site. New C-A domains could be ligated into pSMC using the restriction sites XbaI and NotI. The restriction enzyme recognition

site for XbaI differs from SpeI, but creates the same overhang 5' overhang, meaning ligation of these sites creates a sequence unable to be digested by either XbaI or SpeI.

The A domain substitution plasmid, pSMA, was created by ligating the C domain from the second module of *pvdD* into pSMC using a 5' XbaI and a 3' NotI restriction site added during PCR amplification by the primers Thr-WT\_CAfwd and CATCrev. Ligation of the C domain into pSMC using XbaI destroyed the SpeI restriction site of pSMC. A new SpeI restriction site was added downstream to the introduced C domain by the reverse primer. Like the NotI restriction site, the location of the new SpeI restriction site was selected by sequence alignment with the first module of *pvdD* and is based on previous *pvdD* A domain substitutions (Ackerley and Lamont, 2004). This enabled A domains to be ligated into the plasmid using the restriction sites XbaI and NotI.

## 2.8.2 Creation of domain substitution plasmids for Chapter 5

### 2.8.2.1 Creation of the plasmid pST2

To construct the T2 domain substitution plasmid, the terminal T-TE domains from the plasmid pSMC:Thr-WT (pSMC containing the C-A domains from the second module of *pvdD*) were excised by digestion with NotI and SalI, and then replaced with the *pvdD* TE domain amplified using the primers TE\_Fwd and TTeRev (Table 2-1) and digested with NotI and SalI. The primer TE\_Fwd added a new SpeI restriction site downstream to the NotI restriction site to enable the subsequent insertion of new T domains. The resulting plasmid named pST2 contained a copy of *pvdD* lacking the T domain from the second module and containing SpeI and NotI restriction sites to insert new T domains into the second module of *pvdD*.

### 2.8.2.2 Creation of the plasmid pTCA

The T-C-A domain substitution plasmid pTCA was created in a similar manner to pSMC. The difference was that for the first step, C-A domains were ligated into pSW196 instead of C-A-T domains. These C-A domains from the first module were amplified using primers CATfwd and Thr-A\_Rev (Table 2-1) and ligated into pSW196 using EcoRI and NotI restriction sites. Next the T-TE domains were ligated downstream to the C-A domains using the 5' NotI and 3' SacI restriction sites introduced by the primers during PCR. The resulting plasmid contained the C-A domains from the first module of

*pvdD* and the T-TE domains from the second module. A NotI restriction site between these allowed insertion of T-C-A domains from other modules.

### 2.8.2.3 Creation of pST1

The domain substitution plasmid pST1 was created to substitute new T domains into the first module of *pvdD*. It was created by PCR amplification of the entire second module of *pvdD* using primers CATTE\_Fwd and TTeRev (Table 2-1), and ligation using NotI and SalI restriction sites into the intermediate plasmid from the creation of pTCA that contained the C-A domains from the first module. The primer CATTE\_Fwd added NotI and SpeI restriction sites between the domains from the first module and the second module of *pvdD*, which allowed the subsequent insertion of T domains using these restriction sites.

## 2.8.3 Creation of the directed evolution plasmid and loop substitution for Chapter 6

### 2.8.3.1 Creation of the plasmid pDEC

To create a plasmid with tighter regulation of gene expression for substitution of C domains, the P<sub>BAD</sub> promoter was excised from pSW196 using the restriction sites NsiI and SacI and ligated into pUCP22 using the restriction sites PstI and SacI. The resulting plasmid was named pUCBAD. Next, the *pvdD* gene lacking the C-A domains from the second module was excised from the plasmid pSMC using NheI and SacI restriction sites and annealed into the pUCBAD vector using the restriction sites NheI and SacI. The A domain from the second module was then PCR amplified using primers Thr-WT\_Afwd2 and Thr-WT\_Rev, and ligated into this vector using SpeI and NotI restriction sites. The resulting plasmid containing *pvdD* that was lacking the second C domain was named pDEC. The forward primer Thr-WT\_Afwd2, for amplifying the A domain, added a SalI restriction site downstream to the SpeI restriction site to allow directional cloning of the C domain. The C domain was inserted with XbaI and XhoI restriction sites added by the primers Thr-WT\_CAfwd and Thr-WT\_Crev. The XhoI/SalI fusion site was silent following ligation, i.e. did not alter the primary sequence of *pvdD*.

### 2.8.3.2 Creation of the chimeric C domain constructs

Three chimeric C domains were created in the pDEC plasmid. The first of these was J12-D3. This plasmid was created by PCR amplifying the first two variable regions of JC<sub>Lys</sub> using the primers Lys-C\_CAfwd and Lys-C\_C12Rev, and ligating these into pDEC using XbaI and SalI restriction sites. The reverse primer added a SpeI restriction site so that the third variable region of DC<sub>Thr</sub> could be amplified and ligated into this new plasmid. This plasmid was created this way so that the third variable region could easily be modified. The J1-D23 chimeric C domain was created using overlap PCR whereby the J1 fragment was amplified using the primers Lys-C\_CAfwd and DE\_OL\_1R and the D23 fragment was amplified using the primers DE\_OL\_1F and Thr-WT\_Crev. These were then used in an overlap PCR reaction according to methods in Section 2.6.3.2. The J1-D2-J3 chimera was created in a similar way, in which the J1-D2 fragment was amplified from the J1-D23 chimeric C domain using the primers Lys-C\_CAfwd and DE\_OL\_2R and the J3 fragment was amplified using the primers DE\_OL\_2F and Lys-C\_Crev. The resulting chimeric C domain from overlap PCR of these fragments was ligated into the pDEC plasmid.

### 2.8.3.3 Creation of the loop substitution construct

To replace the loop covering the active site for the second C domain of *pvdD*, a site-directed mutagenesis technique was used. In brief, the loop from the *pvdJ* C domain was amplified with primers Hom\_LyslFwd and Hom\_LyslRev. These primers added homology to the *pvdD* gene at each end of the amplicon. The amplified loop was then used to amplify a PCR2.1 plasmid containing the C domain from *pvdD* according to methods in Section 2.6.3.3. The resulting amplified plasmid had the loop from *pvdD* replaced with that from *pvdJ*. The C domain from this plasmid was subsequently PCR amplified then cloned into pDEC using the restriction sites XbaI and XhoI for the C domain and SpeI and SalI for the vector.

### 2.8.3.4 Construction of the plasmid pTRN

A new plasmid was needed to insert the recombined third variable regions from the second C domain of *pvdD* and the first C domain from *pvdJ* into the *pvdD* C domain. For this, the first two variable regions of the *pvdD* C domain were PCR amplified using the primers Thr-WT\_CAfwd and Thr-WT\_C12Rev. This was ligated into the pDEC vector

using a 5' SpeI/XbaI and a 3' SalI/SalI ligation. The resulting plasmid named pTRN contained SpeI and SalI restriction sites to insert the recombined third variable region.

## **2.9 Analysis of pyoverdine production by recombinant and WT *P. aeruginosa* PAO1 strains**

### **2.9.1 Assessment of strains grown on KB agar plates**

Single colonies were picked from an LB agar plate using a pipette tip and the pipette tip stabbed into a KB agar plate supplemented with the relevant antibiotics and L-arabinose. For assessing growth under severe iron limitation, the KB agar plates were supplemented with 200 µg·mL EDDHA. KB agar plates were incubated at 37 °C for 24 - 36 h, then viewed and photographed in a UV-transilluminator.

### **2.9.2 Pyoverdine synthesis in liquid media**

Liquid media assays of pyoverdine production were carried out in 96-well plates. Strains were grown in triplicate for 24 h in 200 µL LB supplemented with the appropriate antibiotics. Each well was diluted 20x in a new 96-well plate and to a total volume of 200 µL using M9 succinate media supplemented with antibiotics and 0.2 % (w/v) L-arabinose. The concentration of L-arabinose was selected based on initial tests as part of my honours' project. The cultures in the second plate were grown for 24 h at 37 °C under aeration. The culture from each well was then transferred to a microcentrifuge tube and centrifuged to pellet cells. A 100 µL volume of supernatant was transferred to a clean 96-well plate and 100 µL of fresh M9 media added. The absorbance (400 nm) and fluorescence (emission 440 nm, excitation 400 nm) was measured using a microplate reader.

### **2.9.3 MALDI-TOF/TOF mass spectrometry**

α-Cyano-4-hydroxycinnamic acid (CHCA) matrix (Beavis et al., 1992) was prepared by combining 500 µL acetonitrile, 500 µL ultrapure water, 1 µL trifluoroacetic acid and 10 µg of CHCA. The solution was vortexed, followed by centrifugation for 5 min to separate insoluble CHCA. The supernatant was transferred to a clean tube and used in mass spectrometry.

Samples for mass spectrometry were taken from cultures grown in M9 media expressing pyoverdine as per Section 2.9.2. A 1  $\mu$ L sample of supernatant was mixed with CHCA matrix solution in a sample to matrix ratio of 1:20 (v/v). The sample was mixed by pipetting, then 1  $\mu$ L spotted onto an Opti-TOF® 384 well MALDI plate (Applied Biosystems, Foster City, CA) and allowed to dry at room temperature. From Chapter 5 onwards, the sample to matrix ratio was adjusted to 1:5 (v/v). This allowed easier detection of peaks with low abundance. The new ratio altered the size of the peak corresponding to loss of H<sub>2</sub>O for all pyoverdines and increased the abundance of a non-annotated peak for the truncated pyoverdine product. Nonetheless, this was clearly the same product. Samples were analysed using a MALDI TOF/TOF 5800 mass spectrometer (Applied Biosystems) in positive ion mode. Each spot was externally calibrated using cal2 calibration mixture (Applied Biosystems).

#### **2.9.4 Screening the StEP library**

After transformation of the StEP ligations, the bacteria were recovered before spreading on LB agar plates supplemented with gentimycin. Plates were then incubated at 37 °C for 24 hours. Colonies were then picked with a toothpick and inoculated into a 96-well plate containing 100  $\mu$ L of M9 media supplemented with 0.1 % L-arabinose. The cultures in the 96-well plate were grown for 24 h at 37 °C under aeration. After 24 h, 100  $\mu$ L of fresh M9 media was added to each well, and then the fluorescence (excitation 400 nm, emission 440 nm) was measured using a microplate reader.

### **2.10 Bioinformatics**

#### **2.10.1 Prediction of NRPS substrate specificity**

The substrate specificity of A domains was predicted using the NRPS/PKS analysis web server (<http://nrps.igs.umaryland.edu/nrps/>), which makes predictions based on the eight residues lining the substrate binding pocket (Bachmann and Ravel, 2009).

#### **2.10.2 Multiple sequence alignments and phylogenies**

Multiple sequence alignments of amino acids were created using the ClustalW algorithm. Phylogenetic trees were created from alignments using the maximum likelihood method and Jones-Taylor-Thornton model (Jones et al., 1992). Data sets were bootstrapped using

1000 resampling events (Felsenstein, 1985). Evolutionary analysis was performed using MEGA 5.2.1 (Tamura et al., 2011).

### 2.10.3 Pairwise sequence alignment

Pairwise sequence alignment was performed using the Needle pairwise alignment tool available at [http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/). This generates alignments based on the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970)

### 2.10.4 Structural model creation and selection

Structural models of the C domain from the second module of PvdD were created using multiple methods. The best structural models created from each method were submitted to the QMEAN server to obtain QMEAN6 and QMEANclust scores (Benkert et al., 2009a).

#### 2.10.4.1 Automated modelling servers

The sequence of the condensation domain from the second module of PvdD was submitted to the servers listed in Table 2-11. These all create models from a protein sequence and required no further user input.

#### 2.10.4.2 Model creation using SWISS-MODEL

The structure of TycC (PDB code: 2JGP) was used for modelling as it was predicted to be the most suitable by the Swiss-Model server (<http://swissmodel.expasy.org/>) (Arnold et al., 2006). The sequence of the C domain from Thr-WT was aligned to TycC in Swiss PDB Viewer (Guex and Peitsch, 1996, 1997) and several alignments were submitted to the Swiss-Model server. The best model was selected by QMEAN6.

#### 2.10.4.3 Model creation using Modeller 9.11

Models were created in Modeller 9.11 (Eswar et al., 2001) using modified Python 2 (Python Software Corporation; Wilmington, DE, USA) scripts available in tutorials found at <http://salilab.org/modeller/tutorial/>. Models were built using basic alignment and alignment to multiple templates. At least 50 models were created by each method and the best models selected by QMEAN6. The loop between residues 363 and 373 of the best model was optimised because this was a region of high error in the models.

Table 2-11. Automated structural modelling servers.

Server	URL	Reference
3D-Jigsaw	<a href="http://bmm.cancerresearchuk.org/~3djigsaw/">http://bmm.cancerresearchuk.org/~3djigsaw/</a>	(Bates et al., 2001)
CPHmodels 3.2	<a href="http://www.cbs.dtu.dk/services/CPHmodels/">http://www.cbs.dtu.dk/services/CPHmodels/</a>	(Nielsen et al., 2010)
ESyPred3D	<a href="http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/">http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/</a>	(Lambert et al., 2002)
SPARKSX	<a href="http://sparks.informatics.iupui.edu/sparks-x/">http://sparks.informatics.iupui.edu/sparks-x/</a>	(Yang et al., 2011)
nFOLD	<a href="http://www.reading.ac.uk/bioinf/nFOLD/">http://www.reading.ac.uk/bioinf/nFOLD/</a>	(Jones et al., 2005)
3D-Jigsaw Populus	<a href="http://bmm.cancerresearchuk.org/~populus/">http://bmm.cancerresearchuk.org/~populus/</a>	(Offman et al., 2006)
RaptorX	<a href="http://raptorx.uchicago.edu/">http://raptorx.uchicago.edu/</a>	(Källberg et al., 2012)
FUGUE	<a href="http://tardis.nibio.go.jp/fugue/prfsearch.html">http://tardis.nibio.go.jp/fugue/prfsearch.html</a>	(Shi et al., 2001)
M4T	<a href="http://manaslu.aecom.yu.edu/M4T/">http://manaslu.aecom.yu.edu/M4T/</a>	(Rykunov et al., 2009)
I-TASSER	<a href="http://zhanglab.ccmb.med.umich.edu/I-TASSER/">http://zhanglab.ccmb.med.umich.edu/I-TASSER/</a>	(Zhang, 2008)
Chunk-TASSER	<a href="http://cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html">http://cssb.biology.gatech.edu/skolnick/webservice/chunk-TASSER/index.html</a>	(Zhou and Skolnick, 2007)
MULTICOM	<a href="http://casp.rnet.missouri.edu/multicom_3d.html">http://casp.rnet.missouri.edu/multicom_3d.html</a>	(Wang et al., 2010)
Phyre2	<a href="http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index">http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index</a>	(Kelley and Sternberg, 2009)
MUFOLD	<a href="http://mufold.org/prediction.php">http://mufold.org/prediction.php</a>	(Zhang et al., 2010)
TASSER-VMT	<a href="http://cssb.biology.gatech.edu/skolnick/webservice/TASSER-VMT/index.html">http://cssb.biology.gatech.edu/skolnick/webservice/TASSER-VMT/index.html</a>	(Zhou and Skolnick, 2011)
Pro-sp3-TASSER	<a href="http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html">http://cssb.biology.gatech.edu/skolnick/webservice/pro-sp3-TASSER/index.html</a>	(Zhou and Skolnick, 2009)

### 2.10.5 Viewing of homology and structural models

Homology and structural models were viewed in Swiss-PdbViewer and PyMol (Version 1.5.0.4; Schrödinger, LLC). All structural pictures were rendered in PyMol.

## Chapter 3: Construction and characterisation of *pvdD* A domain substitution variants

### 3.1 Introduction

Owing to their role in the selection and activation of monomers, A domains are sometimes referred to as the gatekeepers of non-ribosomal peptide synthesis (Keating and Walsh, 1999; May et al., 2002; Villiers and Hollfelder, 2011). Despite this role, early attempts to create new products by substitution of A domains (as discussed in Section 1.3.1) met limited success. The fact that such experiments generally gave drastically reduced yields, or no product at all, suggests that additional barriers to recombination exist. Two main hypotheses to explain why constructs bearing substituted A domains are frequently non-functional are: (i) that substrate specificity within the acceptor site of the C domain blocks addition of non-cognate substrates activated by substituted A domains (Belshaw et al., 1999; Ehmann et al., 2000); and (ii) introducing new A domains disrupts essential interactions between the C and A domain (Tanovic et al., 2008). These hypotheses led to a shift from A domain substitutions towards: (i) substituting C-A domains as catalytic partners; or (ii) directly altering the substrate binding pocket of A domains. However, despite a relatively large number of studies focusing on these alternatives, the outcome has generally been a large reduction in enzyme activity. On the other hand limited information is available on A domain substitutions. As a consequence it is unknown which of the two hypotheses above is the primary cause for loss of function in A domain substitutions; or, as mentioned in a recent review, whether A domains can be functionally substituted at all (Baltz, 2012).

This chapter extends earlier work by Associate Professor David Ackerley on A domain substitutions in the pyoverdine synthesising NRPS enzyme, PvdD. PvdD contains two Thr specifying modules that incorporate the final two residues into pyoverdine (Ackerley et al., 2003). In previous work, five A domains were substituted into the first module of *pvdD* (Ackerley and Lamont, 2004). The two PvdD variants that were substituted with Thr specifying A domains were found to synthesise pyoverdine, whereas no pyoverdine synthesis was detected from the three non-Thr specifying A domain variants. These results support the hypothesis of A domain substitutions primarily being limited by C domain acceptor site specificity, but are limited by the small sample size. Moreover,

because this earlier study focused on the first module of *pvdD*, a possible confounding effect was a potential inability of the downstream C domain to tolerate alternative amino acid residues in its donor site. In the work described in this chapter the decision was made to instead focus on the second module of PvdD, which has no C domain downstream, thereby eliminating the possibility of a downstream C domain being unable to receive alternative residues at its donor site.

The aim of the research described in this chapter was to assess whether the main barrier to successful A domain substitutions is acceptor site substrate specificity of the C domain or disruption of the C/A domain junction. To achieve this, a staging plasmid was built for substituting A domains into the second module of *pvdD*. A total of nine modules, three of which specify Thr, were selected as a source of alternative A domains. These modules were sourced from the pyoverdine NRPS genes of four fluorescent pseudomonads. The genetic regions defining the alternative A domains were amplified by PCR and ligated into the A domain substitution plasmid to create variants of *pvdD*. These variants were transformed into a *pvdD* deletion mutant and assessed for pyoverdine production using the fluorescent and siderophore properties of pyoverdine as initial screens, followed by a more detailed mass spectrometry analysis.

### 3.2 Construction of the A domain substitution staging plasmid and A domain substitution variants

To construct variants of *pvdD* containing alternative A domains, a staging plasmid (pSMA; Figure 3-1A) was built, in which alternative A domains could be substituted into *pvdD*. To construct pSMA, a C-A domain substitution plasmid (pSMC; Figure 3-1B) was first constructed, and then modified by introducing the C domain from the second module of *pvdD*. The C-A domain substitution plasmid would later be required for the research described in Chapter 4, and this strategy of construction allowed both plasmids to be built consistently, and in the fewest steps. The C-A domain substitution plasmid, pSMC, was designed in conjunction with former Ackerley laboratory PhD student, Dr Owen. The substitution constructs were built in the integrative plasmid pSW196 (Figure 3-1C) (Baynham et al., 2006). pSW196 was originally created from mini-CTX1; a plasmid that integrates into the *P. aeruginosa* chromosome using the integrase  $\phi$ CTX (Hoang et al., 2000). The main difference between mini-CTX1 and pSW196 is the addition of the

tightly controlled L-arabinose inducible  $P_{BAD}$  promoter derived from pBAD30 (Guzman et al., 1995). The features of chromosomal integration and the tight regulation of expression were considered to be important, since high-level overexpression of PvdD was found to lower the overall levels of pyoverdine synthesis in earlier work (Ackerley and Lamont, 2004).

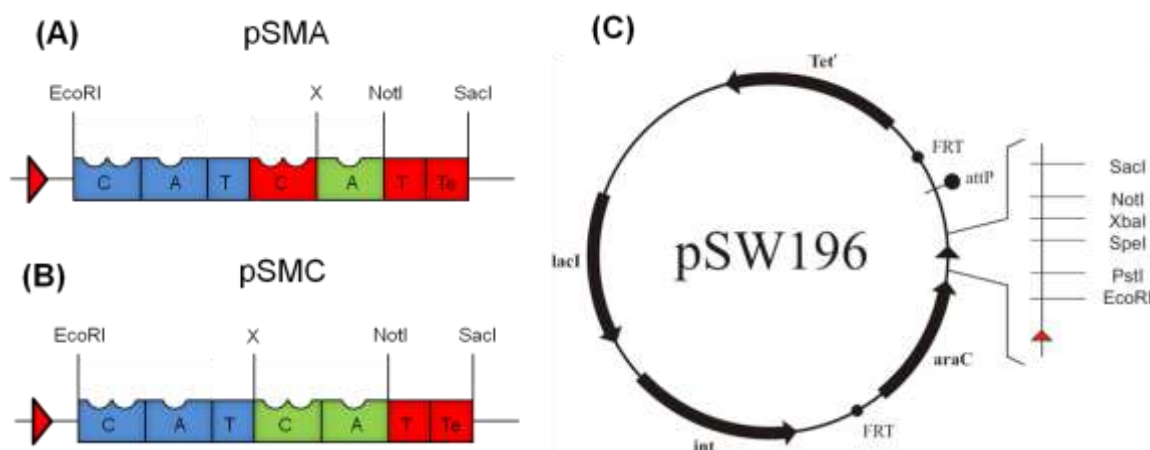


Figure 3-1. Domain arrangement within substitution plasmids, and plasmid map of pSW196. (A) The domain architecture of the A domain substitution plasmid constructed and used in this chapter. Domains coloured blue and red were derived from the first and second module of *pvdD*, respectively. Domains in green were absent in the substitution plasmids and indicate where alternative domains would be added. X refers to a *SpeI* restriction site that was destroyed when ligating C-A and A domains into the plasmids via a compatible *XbaI* sticky end. (B) The domains and restriction sites within the staging plasmid pSMC created for use in Chapter 4. See (A) for colour coding and definition of X. (C) Plasmid map of pSW196, used as a base for the construction of staging plasmids.

Full details of the construction of plasmids pSMC and pSMA are given in Section 2.8.1. For both plasmids, the location of the *NotI* site to be used as a fusion site between the A and T domains of the second module was chosen based on the equivalent location having previously been shown to be permissive of A domain substitutions in the first module of *pvdD* (Ackerley and Lamont, 2004). Introduction of this *NotI* site did not result in any change to the primary sequence of the protein. For pSMC, the location of the intended *SpeI/XbaI* fusion site (“X” in Figure 3-1B) was chosen based on studies showing the highly variable stretch of amino acids between the T and C domain is tolerant of change (Doekel et al., 2008; Mootz et al., 2000). At a protein level, this location is 37 residues downstream of the conserved Ser residue of the T domain from the first module of PvdD, and introduction of the restriction site altered the translated sequence from ER to TR (Figure 3-2). Due to its location within the linker region, it was

predicted this change would have minimal effect on enzyme function. Finally, to create pSMA, ligation of the *pvdD* C domain into pSMC using XbaI destroyed the SpeI restriction site present in pSMC and changed the coding sequence from ER to TR. This enabled a new SpeI restriction site to be added downstream of the introduced C domain by the reverse primer. As with the NotI restriction site, the location of the new SpeI restriction site (“X” in Figure 3-1A) was based on previously successful A domain substitutions into the first module of *pvdD* (Ackerley and Lamont, 2004).

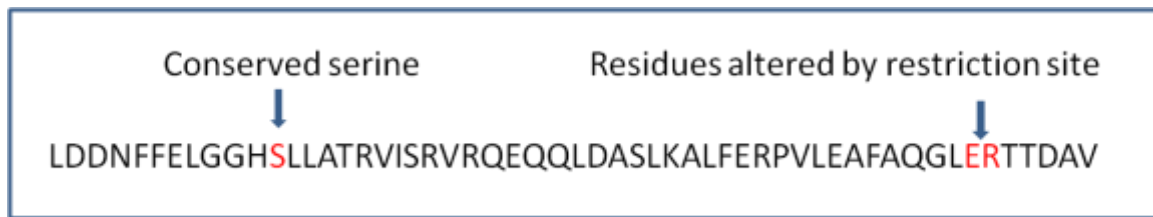


Figure 3-2. Location the of upstream restriction site for pSMC. Protein sequence showing the conserved Ser residue of the T domain from the first module of PvdD in relation to residues altered by introduction of the restriction site for C-A domain substitution. The location was selected to be 37 residues downstream to the conserved Ser residue of the T domain.

Having created pSMA, nine A domains were amplified by PCR and ligated into the plasmid via a XbaI/SpeI fusion at the 5' end, and a NotI/NotI fusion at the 3' end. The various modules the A domains were sourced from are shown in Figure 3-3, labelled A to I. They specify a range of substrates and were sourced from four fluorescent *Pseudomonas* strains. The primers used to amplify each A domain are listed in Table 2-1. Once constructed, the A domain substitution variants were sequenced to confirm no mutations had been introduced during PCR or cloning. The sequences matched the database sequences, except for a silent point mutation in module F, and a C→G substitution at position 2455 of module C that changed the encoded amino acid from Pro to Ala. This mutation in module C was confirmed to be within the genomic sequence of the *P. aeruginosa* PAO1 strain and is located in an NRPS module involved in pyoverdine synthesis. Since *P. aeruginosa* PAO1 synthesises pyoverdine, this mutation was deemed unlikely to inhibit production in the context of the A domain substitution. This domain was later confirmed as functional in the context of a C-A domain substitution (see Section 4.3).

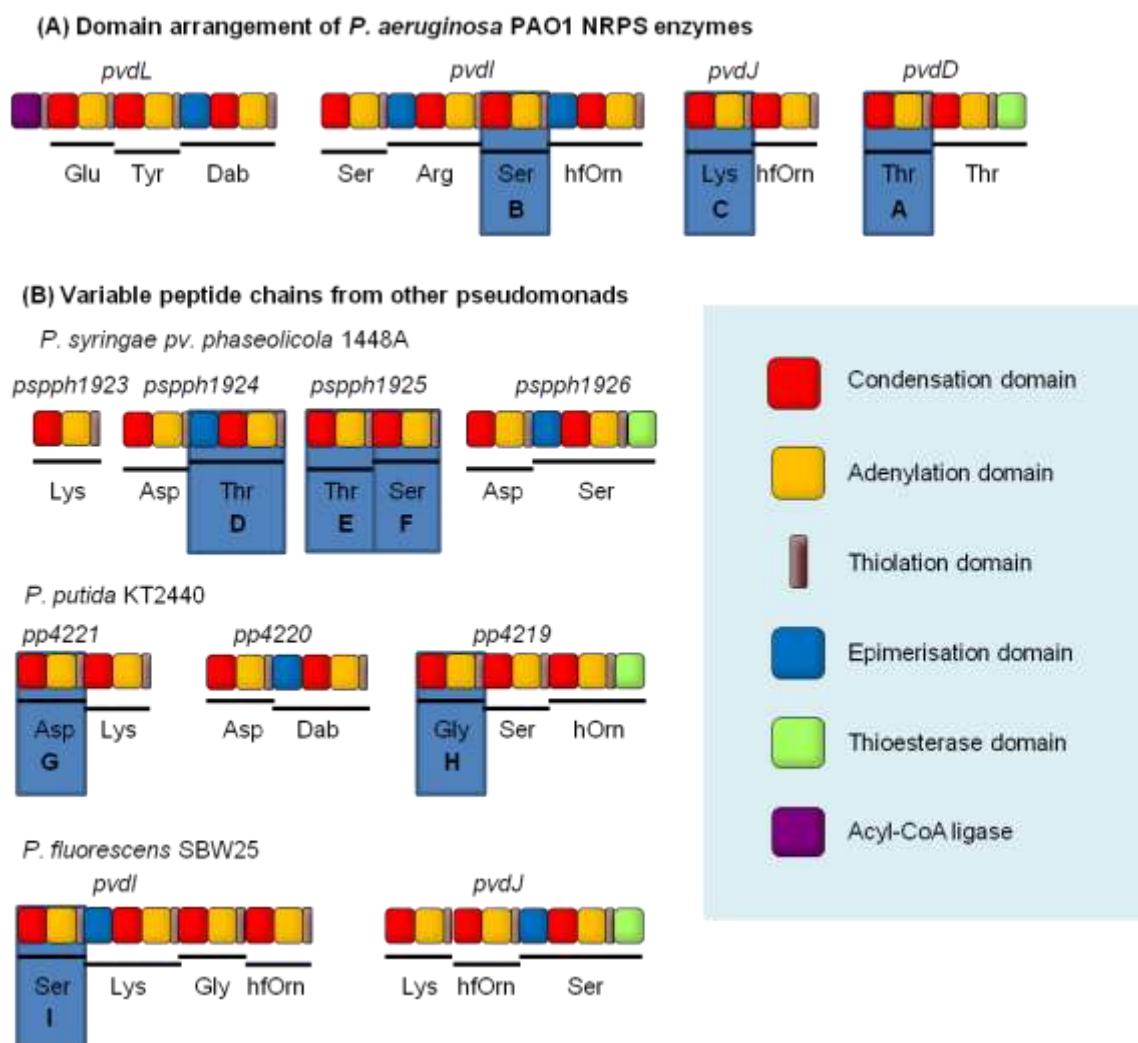


Figure 3-3. NRPS genes with modules used for domain substitution highlighted and labelled from A to I. (A) Domain architecture of the four NRPS genes involved in *P. aeruginosa* PAO1 pyoverdine synthesis. (B) Domain architecture of genes involved in the synthesis of pyoverdine variable peptide chains from three other *Pseudomonas* strains. Accession numbers and substrate predictions of genes used for domain substitution are shown in Appendix A. Dab, 2,4-diaminobutyric acid; hfOrn, L- $N^5$ -formyl- $N^5$ -hydroxyornithine; hOrn,  $N^5$ -hydroxyornithine.

The A domain substitution variants were transformed into a *pvdD* deletion mutant previously created and characterised as being unable to synthesise pyoverdine (Ackerley and Lamont, 2004). The resulting strains, expressing variants of *pvdD* that contain alternative A domains, are referred to in this chapter by the substrate specified by the new A domain followed by the letter of the module in Figure 3-3; for example, the strain expressing the variant containing the A domain from the first module of *pvdD* is referred to as strain Thr-A. In addition to the A domain substitutions, a positive control containing the restriction sites was created by PCR amplification of the A domain from the second module of *pvdD*, using the primers Thr-WT\_Afwd and Thr-WT\_Rev (Table 2-1), and

ligation of the product into pSMA. This wild type restriction site control is referred to as Thr-WT.

### 3.3 Evaluation of pyoverdine synthesis by A domain substitution strains

#### 3.3.1 Growth of A domain substitution strains on solid KB media

Previous work conducting A domain substitutions in *pvdD* assessed their function based on the fluorescence and siderophore activity when grown on solid KB media, both in the absence and presence of the iron chelator EDDHA (Ackerley and Lamont, 2004). KB media was designed as a low iron medium for the induction of pyoverdine expression (King et al., 1954) and the presence of pyoverdine can be observed under UV light due to its fluorescence. In the presence of EDDHA, mutants deficient in pyoverdine production are unable to grow due to the inability to passively uptake iron (McMorran et al., 2001).

To test the function of the nine A domain substitution strains, each strain, in addition to a *pvdD* deletion mutant negative control (*pvdD* deletion strain containing the empty pSW196 plasmid from which pSMA was derived) and Thr-WT positive control, was inoculated onto solid KB media containing 0.2 % L-arabinose to induce *pvdD* expression. This was repeated both in the absence and presence of EDDHA. These plates were then incubated for 24 h at 37 °C, after which photographs were taken on a UV transilluminator (Figure 3-4). No differences were observed between the Thr-WT restriction site control and wild type strain under either of these conditions (data not shown).

The three Thr specifying A domain substitution strains, namely Thr-A, Thr-D and Thr-E, had increased levels of fluorescence compared to the negative control, and were also viable in the presence of EDDHA. The strain Thr-E had noticeably reduced fluorescence and viability relative to strains Thr-A, Thr-D and Thr-WT, indicating reduced pyoverdine synthesis; however, it was still clearly viable. In contrast, none of the non-Thr A domain substitution strains were viable in the presence of EDDHA and all appeared to have levels of fluorescence similar to the negative control. Due to the subjective nature of this test, it was difficult to ascertain whether there might have been small increases in fluorescence. In addition, for growth in the presence of EDDHA it was unknown whether substitution of the terminal Thr of pyoverdine for an alternative residue might interfere with iron

binding or uptake and thereby cause A domain substitution strains to become non-viable. However, there is evidence *P. aeruginosa* PAO1 can use pyoverdines from some other strains (Meyer et al., 1997, 1998, 1999), and it has been suggested that the two terminal Thr residues of *P. aeruginosa* PAO1 pyoverdine are not involved in specificity by FpvA (Meyer et al., 1999). To better address these issues, a more quantitative analysis was required.

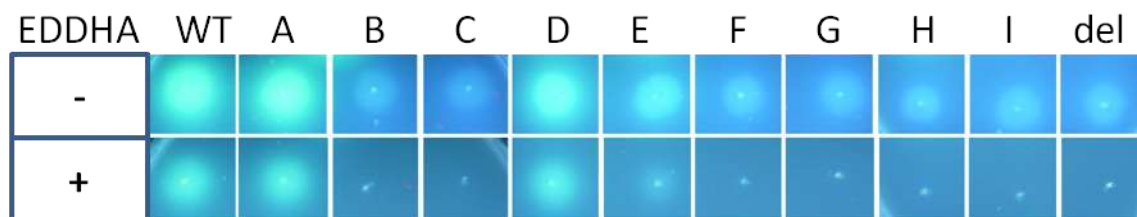


Figure 3-4. Fluorescence of A domain substitutions strains on KB agar media in the absence (-) or presence (+) of EDDHA. Plates were inoculated with Thr-WT, A domain substitution strains A to I, and the *pvdD* deletion mutant (del) then incubated for 24 h at 37 °C. Photographs were taken under UV light.

### 3.3.2 Pyoverdine production by A domain substitution strains grown in liquid media

To more accurately quantify levels of pyoverdine, assays were performed in liquid media. Pyoverdine can be detected by spectrophotometry using absorbance (400 nm) and fluorescence (excitation 400 nm, emission 450 nm) (Hannauer et al., 2010; Yeterian et al., 2009). Initial fluorescence tests found emission at 440 nm to give a stronger signal than 450 nm, and as such this wavelength was used here. It was also found absorbance was linear across a wide range, whereas fluorescence was rapidly saturated (Appendix B). This meant absorbance was more accurate in terms of quantifying pyoverdine levels than fluorescence. However, fluorescence was more sensitive and able to detect pyoverdine at a much lower levels. As a consequence, both absorbance and fluorescence were used on different occasions throughout this research to most appropriately assess levels of pyoverdine synthesis.

The A domain substitution strains were grown for 24 h in M9 media and both the absorbance and fluorescence of the supernatant were measured (Figure 3-5). The wild type strain had 78 % and 90 % the levels of absorbance and fluorescence of the Thr-WT restriction site control strain, respectively (data not shown). This indicated

complementation of *pvdD* by a single-gene copy expressed from the genome-integrated plasmid pSW196 was not reducing pyoverdine production as had been found in previous work using a high copy number plasmid (Ackerley and Lamont, 2004). The three Thr specifying A domain substitution strains exhibited increased absorbance and fluorescence compared to all other strains. The two substitution strains Thr-A and Thr-D, which exhibited the highest levels of growth in the presence of EDDHA, had similar levels of fluorescence and absorbance to the Thr-WT strain. The strain Thr-E had absorbance levels approximately 29 % of Thr-WT, confirming lower levels of pyoverdine synthesis. In contrast, the remaining (non-Thr specifying) A domain substitution strains all had levels of absorbance similar to the *pvdD* deletion mutant, with possibly a small increase in absorbance observed for strain Ser-F. Interestingly, each of the non-Thr specifying A domain substitution strains appeared to have a small increase in fluorescence compared to the *pvdD* deletion strain. This small increase in fluorescence suggested the non-Thr specifying A domain substitution strains may have been producing, albeit at very low levels, an unknown pyoverdine-like product.

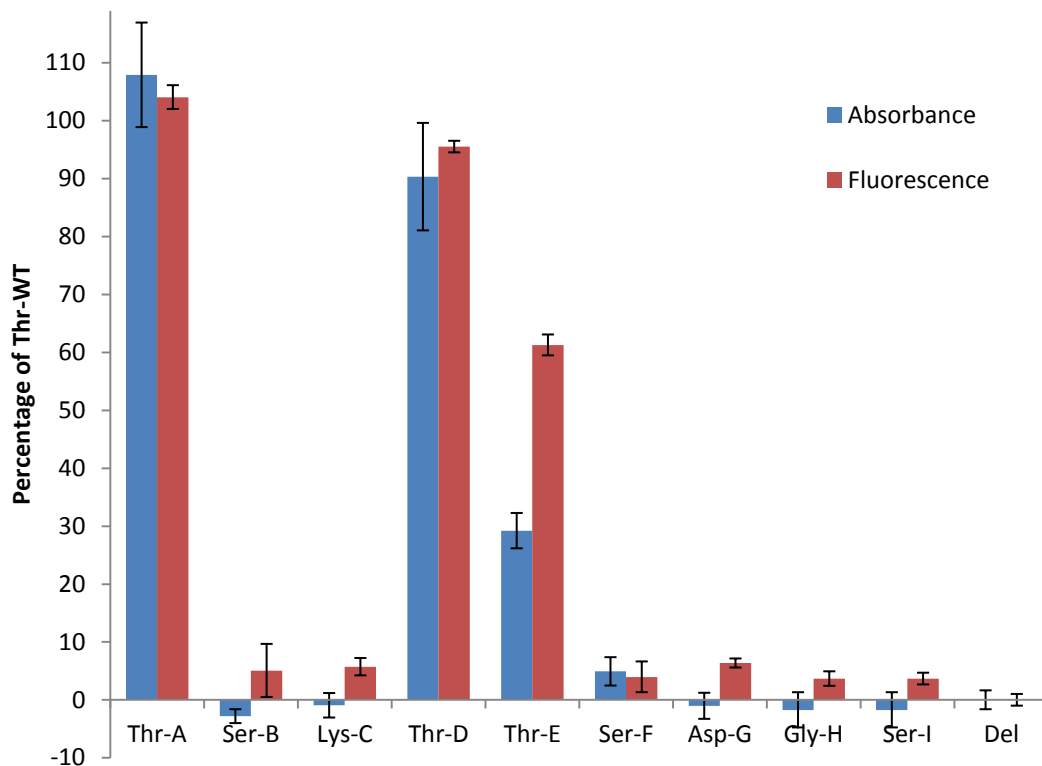


Figure 3-5. Percentage of pyoverdine production from A domain substitution strains grown in liquid media. Values are expressed as a percentage relative to the absorbance (400 nm) or fluorescence (ex. 400 nm/ em. 440 nm) measured for the Thr-WT strain, having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation.

### 3.4 MALDI-TOF MS/MS confirmation of pyoverdine production

#### 3.4.1 Pyoverdine produced by Thr specifying A domain substitutions

The synthesis of wild type pyoverdine by Thr specifying A domain substitution strains was next confirmed by MALDI-TOF MS/MS. To detect pyoverdine, strains were grown in liquid M9 media as in Section 3.3.2 and the supernatant used directly for analysis. The most dominant acyl chain on pyoverdine from *P. aeruginosa* PAO1 is a succinamide moiety (Demange et al., 1990). This pyoverdine has an m/z of 1333.6 (Hannauer et al., 2012) and, consistent with production of wild type pyoverdine, an ion with a m/z ratio within 0.05 % of 1333.6 was detected in the supernatant from each of the Thr specifying A domain substitution strains (Figure 3-6). A second major ion detected at 1031.5 m/z corresponds to the peptide chain attached to two carbons of the chromophore following a *retro*-Diels-Alder (RDA) process opening the tetrahydropyrimidine ring (Demange et al., 1990).

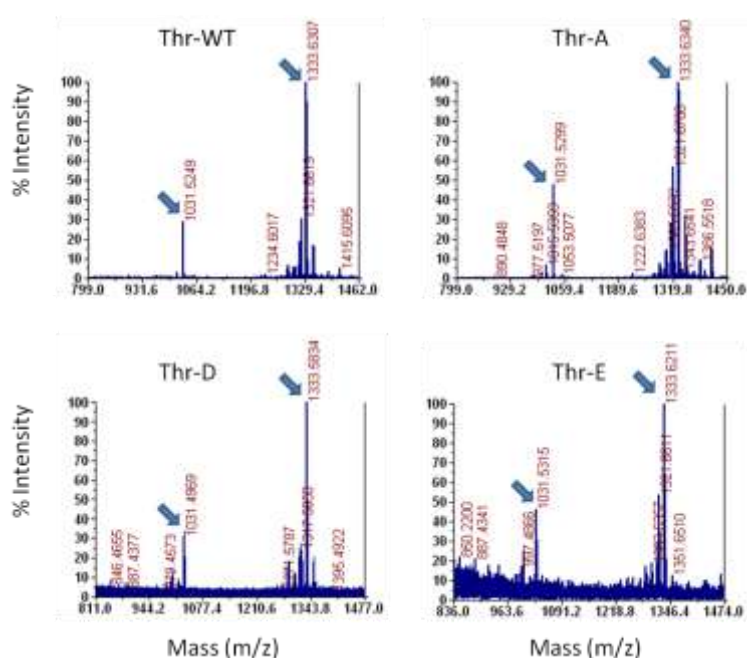


Figure 3-6. Mass spectra obtained from the supernatant of each of the Thr specifying A domain substitution strains and Thr-WT. Arrows indicate peaks at 1031.5 and 1333.6 m/z, corresponding to the peptide chain released by the RDA reaction and wild type pyoverdine, respectively.

To analyse the pyoverdine products further and increase the sensitivity of detection, collision-induced dissociation (CID) was performed on the 1333.6 m/z ion (Figure 3-7,

Table 3-1). The five most abundant ions detected after CID from all four Thr specifying A domain substitution strains were of  $m/z$  890.5, 1031.5, 1234.6, 1291.6 and 1316.6 (Figure 3-7, indicated by arrows). Four of these, namely 890.5, 1031.5, 1234.6 and 1316.6  $m/z$ , correspond to C-terminal ions (losing part of the pyoverdine at the chromophore terminus). The first two of these correspond to the  $Y_7''$ -ion and the peptide chain released during RDA fragmentation of the chromophore (Demange et al., 1990). The nomenclature for pyoverdine Y-ions, in addition to A- and B-ions (discussed below), is illustrated in Figure 3-8. The last two ions are consistent with pyoverdine losing: (i) the succinamide acyl group; and (ii) the  $NH_3$  group from the chromophore (Fuchs and Budzikiewicz, 2000). Smaller C-terminal ions are likely absent, because the chromophore stabilises the positive charge, causing  $Y''$ -ions to be rare (Fuchs and Budzikiewicz, 2000). An exception to the stabilisation is when the fragment contains a basic residue such as arginine (Fuchs and Budzikiewicz, 2000), as seen here by the presence of the  $Y_7''$ -ion. The other highly abundant ion has an  $m/z$  of 1291.6 and is consistent with loss of  $CH_2N_2$  from the arginine residue (Hannauer et al., 2012). These five ions are emphasised here because containing the residues added by *pvdD* and their high abundance proved useful in verifying pyoverdine synthesised at low levels by other strains.

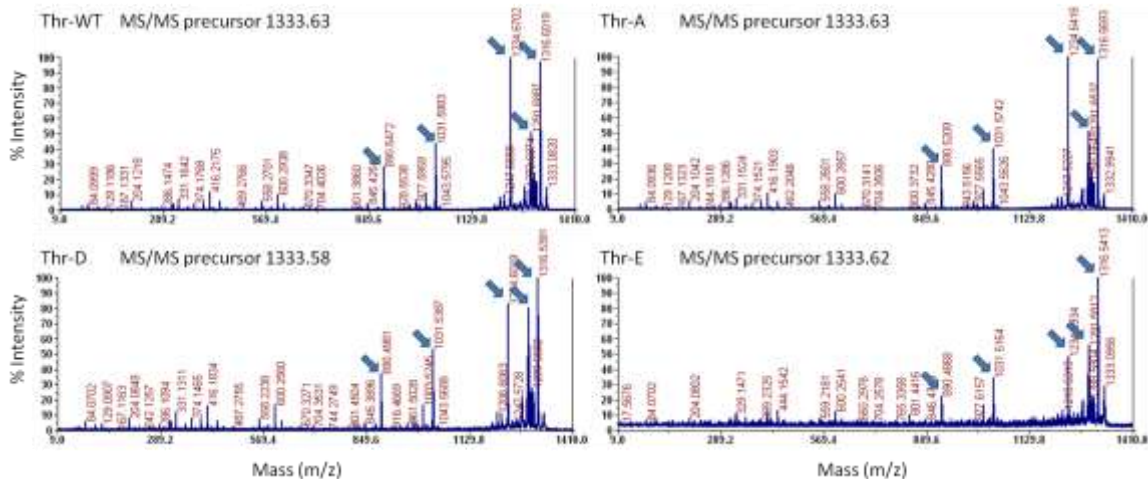


Figure 3-7. CID spectra of pyoverdine synthesised by Thr-WT and the three Thr specifying A domain substitution strains. The five most abundant ions are indicated by arrows.

Table 3-1. Ions detected by CID of pyoverdine from Thr specifying A domain substitution strains.

	Thr-WT	Thr-A	Thr-D	Thr-E	Calculated	
<b>Ion</b>	<b>m/z</b>				<b>m/z</b>	<b>Composition</b>
[RDA + H - Suca] <sup>+</sup>	204.12	204.10	204.08	204.08	204.08	C <sub>10</sub> H <sub>10</sub> N <sub>3</sub> O <sub>2</sub>
A <sub>1</sub> - Suca	317.20	317.18	317.18	ND	317.12	C <sub>15</sub> H <sub>17</sub> N <sub>4</sub> O <sub>4</sub>
A <sub>1</sub> - NH <sub>3</sub>	399.22	399.21	399.19	399.23	399.13	C <sub>19</sub> H <sub>19</sub> N <sub>4</sub> O <sub>6</sub>
A <sub>1</sub>	416.22	416.19	416.18	416.21	416.16	C <sub>19</sub> H <sub>22</sub> N <sub>5</sub> O <sub>6</sub>
B <sub>1</sub>	444.19	444.17	444.16	444.15	444.15	C <sub>20</sub> H <sub>22</sub> N <sub>5</sub> O <sub>7</sub>
B <sub>2</sub>	600.29	600.29	600.26	600.25	600.25	C <sub>26</sub> H <sub>34</sub> N <sub>9</sub> O <sub>8</sub>
B <sub>3</sub>	687.33	687.31	687.35	ND	687.29	C <sub>29</sub> H <sub>39</sub> N <sub>10</sub> O <sub>10</sub>
B <sub>4</sub>	845.43	845.43	845.39	845.41	845.35	C <sub>35</sub> H <sub>49</sub> N <sub>12</sub> O <sub>13</sub>
Y <sub>7</sub> <sup>''</sup>	890.55	890.52	890.50	890.49	890.47	C <sub>35</sub> H <sub>64</sub> N <sub>13</sub> O <sub>14</sub>
[RDAb + H] <sup>+</sup>	1031.60	1031.57	1031.54	1031.52	1031.51	C <sub>41</sub> H <sub>71</sub> N <sub>14</sub> O <sub>17</sub>
[M + H - Suca] <sup>+</sup>	1234.67	1234.64	1234.60	1234.59	1234.58	C <sub>51</sub> H <sub>80</sub> N <sub>17</sub> O <sub>19</sub>
[M + H - CH <sub>2</sub> N <sub>2</sub> ] <sup>+</sup>	1291.69	1291.65	1291.60	1291.58	1291.59	C <sub>54</sub> H <sub>83</sub> N <sub>16</sub> O <sub>21</sub>
[M + H - NH <sub>3</sub> ] <sup>+</sup>	1316.60	1316.57	1316.54	1316.54	1316.59	C <sub>55</sub> H <sub>82</sub> N <sub>17</sub> O <sub>21</sub>
MS/MS precursor	1333.63	1333.63	1333.58	1333.62	1333.61	C <sub>55</sub> H <sub>85</sub> N <sub>18</sub> O <sub>21</sub>

Shaded cells indicate ions expected to contain the residue added by the second module of PvdD. Suca, succinamide; RDA, chromophoric fragment from the RDA process; RDAb, peptide chain fragment released by the RDA process; ND, not detected.

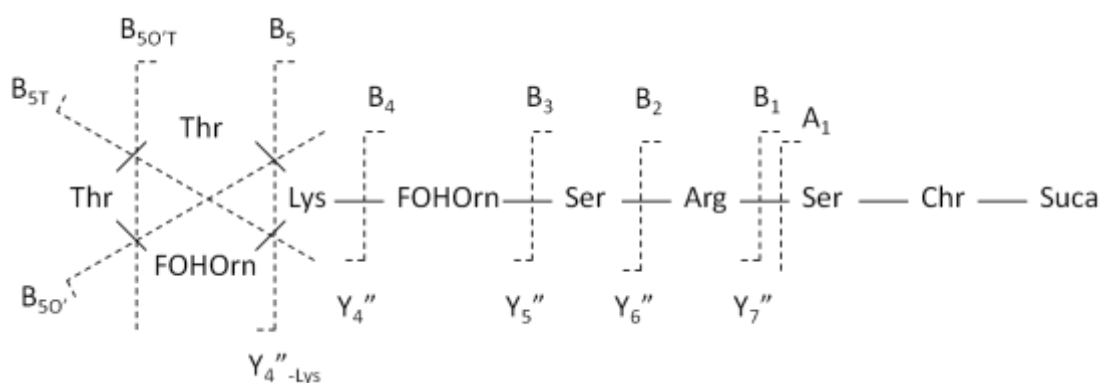


Figure 3-8. Nomenclature for pyoverdine ions formed by CID of pyoverdine as suggested by Fuchs and Budzikiewicz (Fuchs and Budzikiewicz, 2000). Dashed lines indicate location of fragmentation. Contrary to the common nomenclature where the N-terminus is depicted to the left (Roepstorff and Fohlman, 1984), the nomenclature for pyoverdines depict with N-terminus to the right (Fuchs and Budzikiewicz, 2000).

Additional annotated ions are associated with the chromophore and B-ions. The fragmentation patterns from CID of multiple pyoverdine chromophores with different acyl chains and different first amino acids in the variable peptide chain have been analysed in depth (Budzikiewicz et al., 2007). Characteristic of a pyoverdine chromophore containing a succinamide acyl chain and linked to Ser as the first residue of

the variable peptide chain, ions of  $m/z$  204.1, 317.1, 399.1 and 416.1 were detected by CID of pyoverdine from strains Thr-WT, Thr-A and Thr-D (Budzikiewicz et al., 2007). The ion of  $m/z$  317.1 was not detected from strain Thr-E but the other three were; because the ion of  $m/z$  317.1 is of low abundance, its absence seems likely to be due to increased noise rather than the existence of fundamentally different pyoverdine species. For strains Thr-WT, Thr-A and Thr-D, the N-terminal ions  $B_1$  to  $B_4$  were detected at 444.2, 600.3, 687.3 and 845.4  $m/z$ , confirming the sequence of pyoverdine up until the cyclic portion. The same ions were observed for strain Thr-E with the exception of the ion at 687.3  $m/z$ . Again, this was likely obscured by the increased background signal. Larger B-ions were not detected, consistent with the cyclic portion of the pyoverdine peptide chain not fragmenting during CID of  $[M + H]^+$  ions (Fuchs and Budzikiewicz, 2000). These results confirmed that wild type pyoverdine was being synthesised by the Thr specifying A domain substitution strains and were consistent with ions reported from CID of pyoverdine containing a succinate acyl group (Mossialos et al., 2002).

#### 3.4.2 Pyoverdine produced by non-Thr specifying A domain substitutions

In Section 3.3.2, low levels of increased fluorescence compared to the *pvdD* deletion strain were detected in the supernatant of non-Thr A domain substitution strains. This suggested that these A domain substitution strains may be producing low levels of pyoverdine. The supernatant from these strains was therefore analysed by mass spectrometry to confirm whether the increase in fluorescence was indeed due to pyoverdine synthesis and, if so, to identify which substrates were being incorporated into the terminal position of pyoverdine.

Due to the low levels of pyoverdine present, there was greater noise in the mass spectrometry results than had been observed for the Thr specifying A domain substitution strains. Despite this, from all strains a product of 1333.6  $m/z$  was detected and no peaks corresponding to the addition of new residues was present above the 20 signal/noise ratio set for CID (Figure 3-9). Additionally, the 1333.6  $m/z$  product was not detected from the negative control *pvdD* deletion strain. Analysis of the 1333.6  $m/z$  species by CID detected the five most abundant ions characteristic of pyoverdine fragmentation were present in spectra from each strain except for the  $Y_7''$  ion from strain Ser-F (Figure 3-10, Table 3-2). Other peaks were mostly absent, which seemed likely to be due to increased levels of

noise in these spectra caused by the low abundance of the precursor ion. Nonetheless, these results confirmed that the non-Thr specifying A domain substitution strains were synthesising low levels of WT pyoverdine.

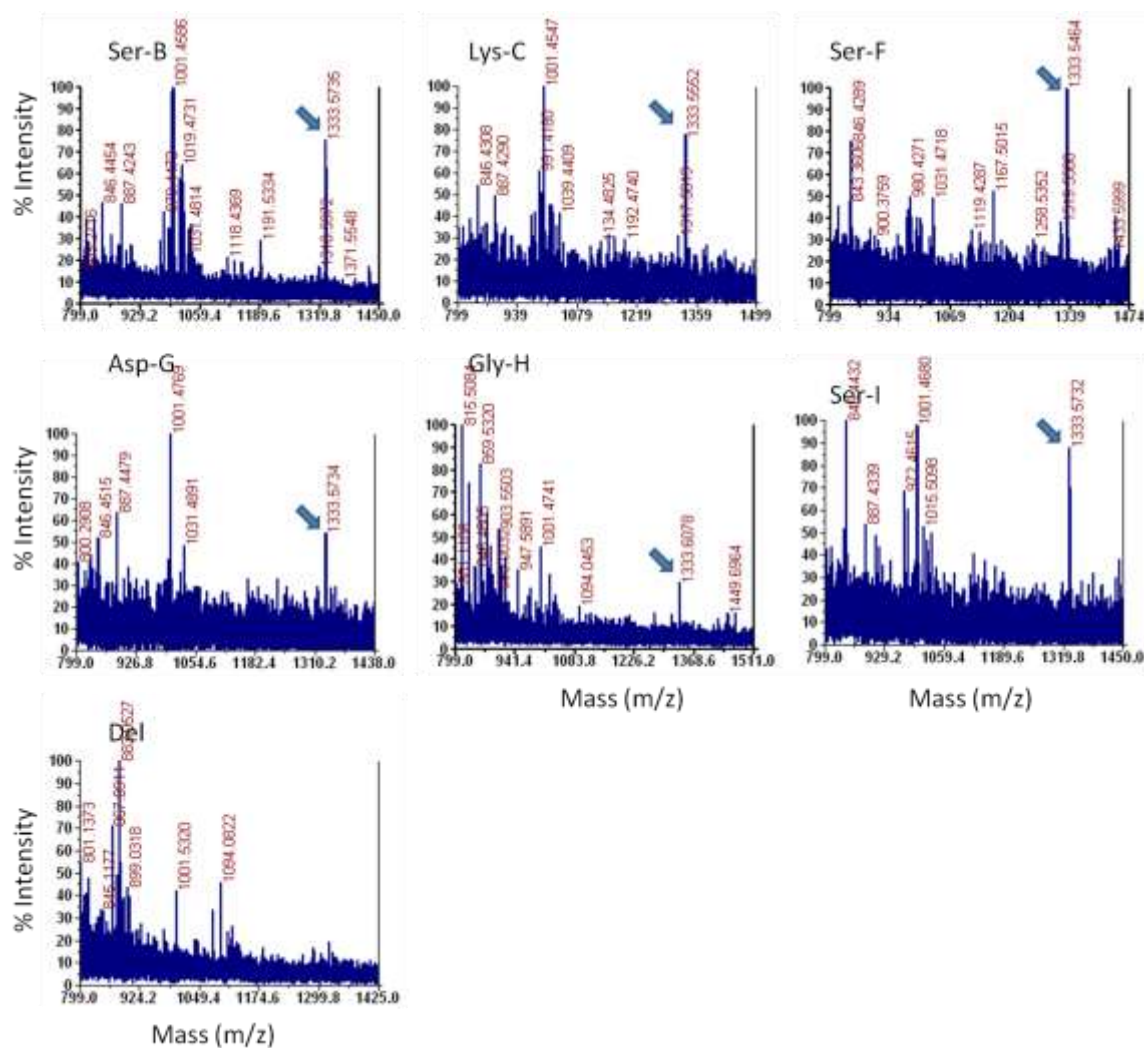


Figure 3-9. MALDI-TOF spectra obtained using the supernatant of non-Thr specifying A domain substitution strains. Arrows indicate location of peak at 1333.6 m/z.

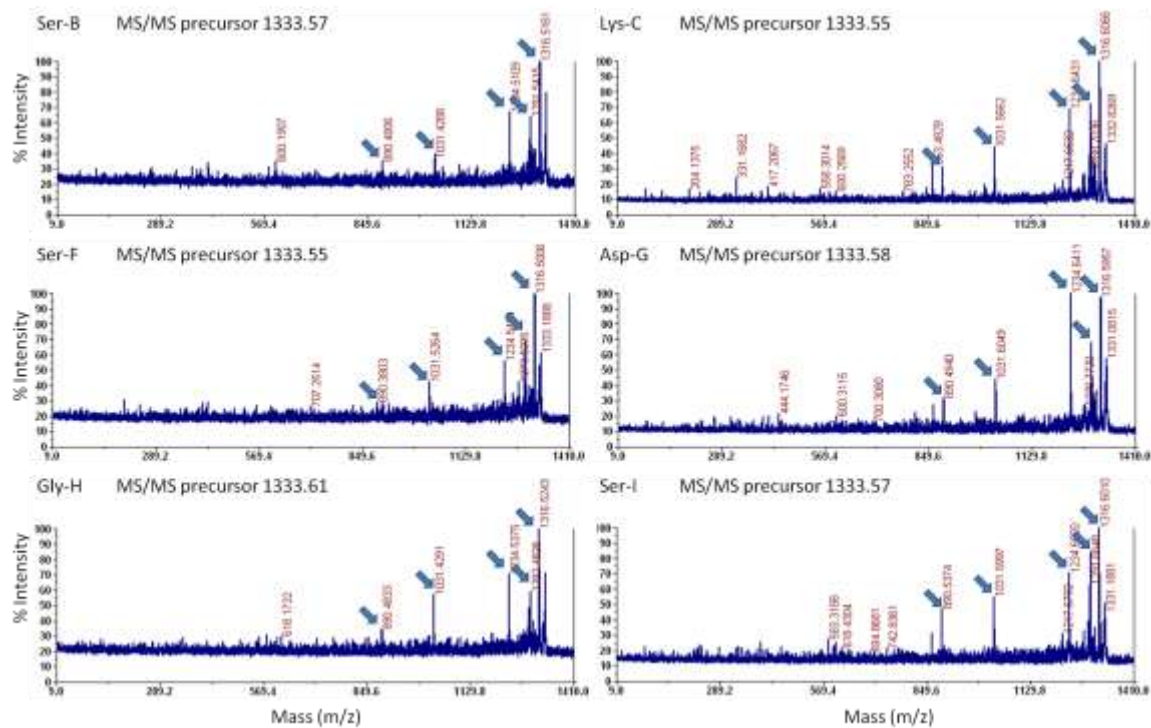


Figure 3-10. CID spectra of pyoverdine synthesised by non-Thr specifying A domain substitution strains. Arrows indicate the abundant ions characteristic of pyoverdine fragmentation.

Table 3-2. Ions detected by CID of pyoverdine from non-Thr specifying A domain substitution strains.

	Ser-B	Lys-C	Ser-F	Asp-G	Gly-H	Ser-I
<b>Ion</b>	<b>m/z</b>					
[RDA + H - Suca] <sup>+</sup>	ND	204.14	ND	ND	ND	ND
A <sub>1</sub> - Suca	ND	ND	ND	ND	ND	ND
A <sub>1</sub> - NH <sub>3</sub>	ND	ND	ND	ND	ND	ND
A <sub>1</sub>	ND	ND	ND	416.19	ND	ND
B <sub>1</sub>	ND	ND	ND	444.17	ND	ND
B <sub>2</sub>	600.19	600.26	600.32	600.31	600.19	ND
B <sub>3</sub>	ND	ND	ND	ND	ND	ND
B <sub>4</sub>	ND	ND	ND	ND	ND	ND
Y <sub>7</sub> <sup>''</sup>	890.48	890.53	ND	890.49	890.48	890.54
[RDAb + H] <sup>+</sup>	1031.43	1031.57	1031.61	1031.60	1031.43	1031.60
[M + H - Suca] <sup>+</sup>	1234.51	1234.64	1234.68	1234.64	1234.54	1234.65
[M + H - CH <sub>2</sub> N <sub>2</sub> ] <sup>+</sup>	1291.54	1291.67	1291.71	1291.66	1291.59	1291.66
[M + H - NH <sub>3</sub> ] <sup>+</sup>	1316.52	1316.61	1316.64	1316.60	1316.52	1316.60
MS/MS precursor	1333.57	1333.55	1333.55	1333.58	1333.61	1333.57

Shaded cells indicate ions expected to contain the residue added by the second module of PvdD. Suca, succinamide; RDA, chromophoric fragment from the RDA process; RDAb, peptide chain fragment released by the RDA process; ND, not detected.

### 3.5 Discussion

In this chapter an A domain substitution vector was created, targeting the second A domain of PvdD for change, and tested via substitution of nine alternative A domains. The Thr specifying A domain substitution strains all produced wild type pyoverdine, although a two-third decrease in pyoverdine levels relative to Thr-WT was observed for strain Thr-E. In contrast, non-Thr specifying A domain substitution strains made only trace amounts of pyoverdine, which was also shown to be wild type pyoverdine.

One possible reason for the high success of Thr specifying A domain substitutions is a closer similarity of these A domains to the Thr-WT A domain. In fact, the Thr specifying A domains had an average of 76.5 % amino acid identity to the Thr-WT A domain, compared to only 47.5 % for the A domains that specify alternative residues (Table 3-3). These numbers are slightly skewed by Thr-A sharing 99.6 % amino acid sequence identity to the Thr-WT A domain, the two modules of *pvdD* having almost certainly arisen due to a gene duplication event (Merriman et al., 1995). Nonetheless, the other two Thr specifying A domain substitutions still had an average of 17.3 % greater sequence identity to the Thr-WT A domain than did the remaining A domains tested in this study, and on this basis could perhaps be expected to have a greater likelihood of being active in the PvdD context. Even so, the observation that 3/3 Thr specifying A domains were active, while 6/6 non-Thr specifying A domains were inactive provides strong evidence that substrate specificity plays a major role in determining whether an A domain substituted construct will be active. These observations are consistent with the previous smaller scale study conducted in the first module of *pvdD* (Ackerley and Lamont, 2004). Moreover, in the previous study, the introduced Thr specifying A domains shared no greater amino acid identity with the PvdD A domains than the alternative A domains introduced in this work (Table 3-3). Collectively, these results strongly indicate that Thr specifying A domain substitutions are highly successful and that the loss in function seen with the non-Thr specifying A domain substitutions was unlikely to be a consequence of overall reduction in sequence identity.

Table 3-3. Amino acid sequence identity of introduced A domains with the module 2 PvdD A domain that was substituted.

Substitution*	Identity	Similarity	Gaps
Thr-A	99.6	99.8	0.0
Ser-B	50.6	64.1	7.3
Lys-C	52.3	64.9	5.9
Thr-D	64.7	77.4	2.3
Thr-E	65.1	77.3	3.5
Ser-F	47.6	61.5	5.7
Asp-G	40.7	56.3	5.3
Gly-H	47.1	62.3	6.8
Ser-I	46.8	61.7	6.9
<b>Thr-<i>snbC</i> m1</b>	<b>53.2</b>	<b>64.5</b>	<b>4.5</b>
<b>Thr-<i>syrB</i> m9</b>	<b>44.1</b>	<b>60.2</b>	<b>5.4</b>
<b>Cys-<i>acvA</i> m2</b>	<b>36.2</b>	<b>54.0</b>	<b>11.8</b>
<b>Ser-<i>pvdI</i> m1</b>	<b>51.4</b>	<b>64.8</b>	<b>5.7</b>
<b>Val-<i>acvA</i> m3</b>	<b>36.4</b>	<b>54.7</b>	<b>10.8</b>

\* Substitutions created in this study are aligned to the A domain from Thr-WT. Substitutions from previous work (highlighted in bold) were aligned to the A domain from the first module. Nomenclature for A domain substitutions from previous work is the substrate activated by the module, followed by the gene the module was sourced and the module within the named gene (Ackerley and Lamont, 2004).

The two main hypothesised reasons for loss of function for A domain substitutions are that they are limited by C domain acceptor site substrate specificity and/or disruption of the C-A junction. The high success rate of Thr specifying A domain substitutions in this work, together with the previous work in *pvdD* (Ackerley and Lamont, 2004) implies disruption of the C/A junction in *pvdD* is not a major cause for loss in function. Instead, this data fits better with the view that PvdD A domain substitutions are limited mainly by C domain acceptor site substrate specificity. Further support for this hypothesis is the production of low levels of wild type pyoverdine and absence of modified pyoverdine by non-Thr specifying A domain substitutions. An explanation for this result is low-level ‘leaky’ activation of Thr by A domains that specify other substrates. This is supported by a large number of A domains having been observed to exhibit low-level promiscuous activities in *in vitro* kinetic studies (Doekel and Marahiel, 2000; Mootz and Marahiel, 1997; Stevens et al., 2006; Takahashi et al., 2007; Villiers and Hollfelder, 2009). Selection of Thr at the acceptor site of the C domain would leave other substrates attached to the T domain, where they would likely be released by TE<sub>II</sub> enzymes due to decreased catalytic turnover (Yeh et al., 2004).

The production of wild type pyoverdine was particularly interesting in that it allowed a direct comparison of the relative strengths of the substrate specificities of the A domain

and the acceptor site of the C domain. If these specificities were approximately equivalent, then both modified and wild type pyoverdines should be produced. As no modified pyoverdine species were observed, it can be concluded that the C domain proof-reading is more stringent in PvdD. Given that some modules with C domains have been identified that can incorporate alternative substrates (often similar substrates and at lowered levels) into the peptide (Han et al., 2012; Konz et al., 1999; Lawen and Traber, 1993; Peypoux et al., 1991), this may not be the case for all C domains.

In conclusion, these results indicate that A domain substitutions have potential to be highly successful in PvdD but are limited by the substrate specificity within the C domain acceptor site. One implication of this result is that A domain substitution could potentially be a viable approach to domain substitutions if the C domain acceptor site specificity could be altered or relaxed. Currently little is known of the mechanism of acceptor site substrate specificity and residues involved. Out of the C domain structures discussed in Section 1.2.3.1, none were solved in complex with substrate and although Keating et al. (2002) suggested residues that may be involved in substrate specificity, these have not been confirmed. Moreover, phylogenetic studies did not identify any specificity conferring residues (Rausch et al., 2007; Roongsawang et al., 2005). As an alternative to bypass C domain acceptor site specificity, the following chapters examine whether C-A domain substitutions can be functional in PvdD and factors that may limit the success of C-A domain substitutions.



## Chapter 4: Substitution of C-A domains in *pvdD*

### 4.1 Introduction

The previous chapter suggests A domain substitutions are limited mainly by the acceptor site specificity of the C domain. A way to bypass the limitation of C domain acceptor site substrate specificity is to substitute C-A domains as functional pairs. The relaxed donor site specificity of C domains (Section 1.2.3.2) implies that substituting C-A domains to a new location would not interfere with C domain specificity. A second advantage of substituting C-A domains together is it keeps the C/A domain interface intact, which may still have a minor influence on the activity of PvdD variants. Nevertheless as discussed in Section 1.3.3, C-A domain substitutions often lose activity and result in yields of the modified peptide being greatly diminished relative to that produced by the wild type NRPS system. In practical terms this issue limits the number of substitutions possible within a single biosynthetic chain, and hence the diversity of modified products that can potentially be made. For predictable outcomes, there is a need for greater understanding of the factors limiting C-A domain substitution.

Previous C-A domain substitutions in the first module of *pvdD* were all non-functional (Ackerley and Lamont, 2004). This could have been due to issues caused by using distantly related C-A domains with low homology or by disrupting the interaction with the upstream enzyme PvdJ. Another explanation raised for this lack of activity was that C domains may sometimes exhibit donor site proofreading after all, such that the introduced C domain might prove incompatible with the incoming residue supplied by the upstream NRPS module, or that a newly introduced amino acid substrate might not be recognised at the donor site of the C domain immediately downstream (Ackerley and Lamont, 2004). While the previous *in vitro* observations of Belshaw *et al* (1999) and Ehmann *et al* (2000) failed to identify any C domains exhibiting strong amino acid side-chain specificity at the donor position, there is some evidence of C domains with donor site specificity towards peptides (Stein *et al.*, 2005) and fatty acids (Kraas *et al.*, 2012). As mentioned previously, using the second module of PvdD eliminates any problems with potential downstream donor site specificity as the terminal pyoverdine synthetase has no C domain located downstream. A further advantage is that the second module of PvdD condenses substrates *in cis*, i.e. attached to T domains within the same

enzyme. Thus C-A domain substitutions in this module are less likely to disrupt interenzyme interactions.

This chapter aims to determine whether C-A domain substitutions can be successful within *pvdD* and whether new products can be incorporated into pyoverdine. A total of nine C-A domain substitution variants were created using the vector pSMC (construction described in Section 3.2). The domains were sourced from the same modules as the A domain substitutions. These were tested in a similar manner to the A domain substitution strains and it was found that three C-A domain substitutions created functional PvdD enzymes. Interestingly, the other C-A domain substitutions displayed low levels of increased fluorescence relative to the *pvdD* deletion strain. Mass spectrometry was used to identify the products that were being synthesised by the C-A domain substitution strains.

## 4.2 Creation of C-A domain substitution strains

The plasmid pSMC made in Section 3.2 was used to create C-A domain substitutions. This plasmid contained a copy of *pvdD* lacking the C-A domains from the second module and containing a SpeI and NotI restriction site in their place. Nine C-A domains from modules A to I, also used for A domain substitution in Chapter 3, were amplified by PCR, restriction digested, and ligated into pSMC using the restriction sites XbaI and NotI. Plasmid sequences were confirmed by sequencing, and sequences matched the database sequences except for the point mutations within two A domains discussed in Section 3.2. The plasmids were transformed into the *P. aeruginosa* PAO1 *pvdD* deletion strain. The resulting strains were named according to the convention used for the A domain substitution strains but differentiated A domain substitutions by starting with CA (for example, the variant containing the C-A domains from the first module of *pvdD* is referred to as strain CAThr-A). The positive control strain containing the restriction sites and C-A domains from the second module of PvdD was named CAThr-WT and showed no reduction in pyoverdine synthesis compared to wild type *P. aeruginosa* PAO1 (not shown).

### 4.3 Evaluating levels of pyoverdine synthesis

#### 4.3.1 Fluorescence of C-A domain substitution strains on solidified KB media

When grown on solid media, only the C-A domain substitution strains containing C-A domains derived from modules Thr-A and Lys-C (Figure 3-3) had levels of fluorescence that were visibly higher than the background of the *pvdD* mutant control (Figure 4-1). In the presence of EDDHA, strains CAThr-A, CALys-C, and CASer-F were the only strains that remained viable (Figure 4-1). These results suggest novel residues can be incorporated into pyoverdine, and that these variant pyoverdines are then able to transport iron into *P. aeruginosa* PAO1. Interestingly, in contrast to the A domain substitutions, only one C-A domain substitution that activates Thr was functional and the other two were non-functional. Of these two non-functional Thr C-A domain substitutions, strain CAThr-D contains a C domain that naturally receives a D-isomer from the upstream module and was expected to be non-functional due to the stereospecificity of C domain donor sites (Belshaw et al., 1999; Ehmann et al., 2000). More surprisingly strain CAThr-E, which not only contains an A domain that was functional in the A domain substitution experiments but also a C domain that “expects” in its native context to receive an L-Thr residue in its donor site, had no detectable levels of fluorescence. This observation, together with the apparent activity of strains CAThr-A and CALys-C (which contain C domains that do not receive L-Thr in their native contexts), strongly suggests that C domain proof-reading of amino acid identity at the donor sites is not the main factor limiting the success of C-A domain substitution experiments.

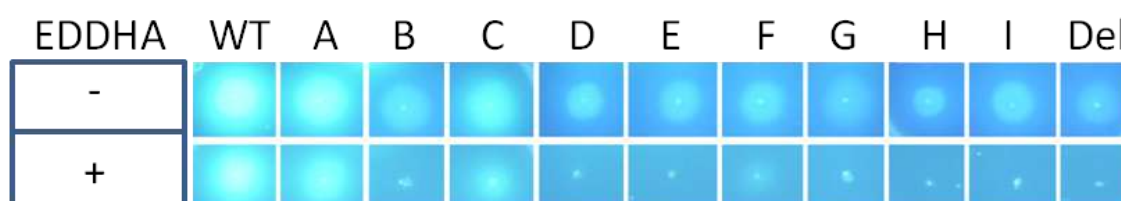


Figure 4-1. Fluorescence of C-A domain substitution strains on KB agar plates in the absence (-) or presence (+) of EDDHA. Plates were inoculated with the positive control strain CAThr-WT (WT), C-A domain substitution strains A to I (as per letter designations in Figure 1), and the *pvdD* deletion pyoverdine negative control (Del), then incubated for 24 h at 37 °C. Photographs were taken under UV light.

#### 4.3.2 Pyoverdine production of C-A domain substitution strains in liquid media

When absorbance of the supernatant of C-A domain substitution strains grown in liquid media was measured (Figure 4-2, blue bars), strains CAThr-A and CALys-C had high levels of absorbance at 83 % and 76 % of the CAThr-WT positive control. Strain CASer-F, which exhibited low levels of growth on solidified King's B media containing EDDHA (Figure 3A), exhibited lower levels of absorbance at 18 % of CAThr-WT levels. In contrast, the other C-A domain substitution strains all had much lower levels of absorbance at levels similar to the *pvdD* deletion strain. Nonetheless, when fluorescence of the supernatant was measured (Figure 4-2, red bars), all C-A domain substitutions had increased levels of fluorescence relative to the *pvdD* deletion strain, except for strain CAAsp-G. Although strain CAAsp-G had increased absorbance, it seems most likely this was due to an artefact because fluorescence is more sensitive in detection of pyoverdine at lower levels. These results suggested that, similar to A domain substitutions, most C-A domain substitution strains still synthesised low levels of a pyoverdine product.

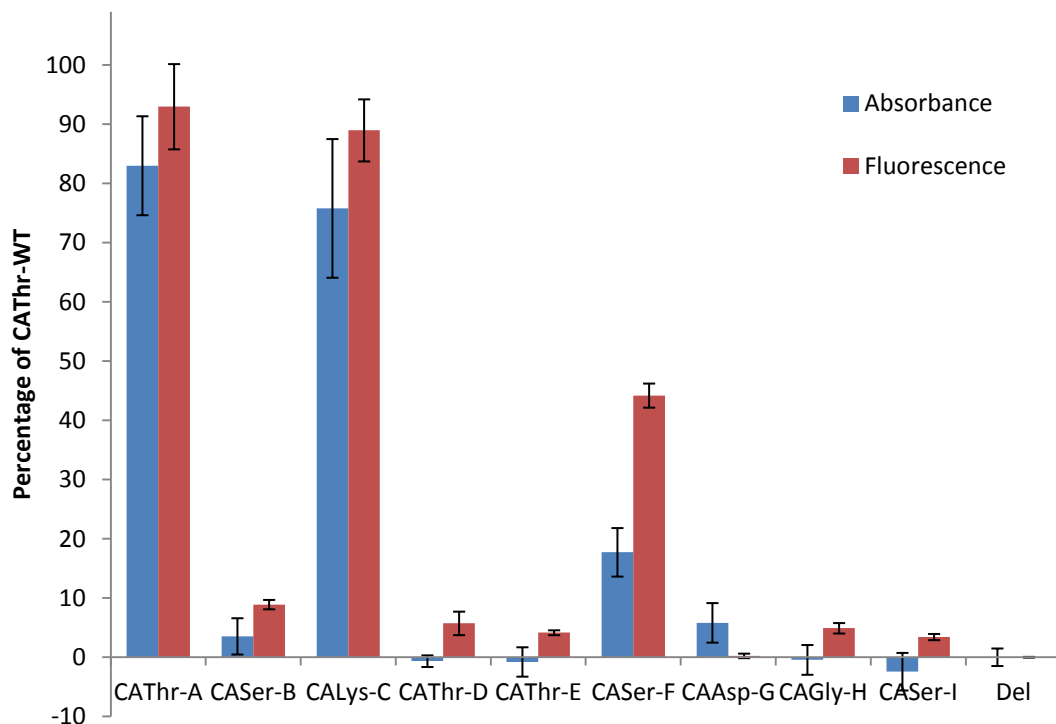


Figure 4-2. Liquid media assays of relative levels of pyoverdine production by C-A domain substitution strains. Values are expressed as a percentage of absorbance (400 nm) or fluorescence (ex. 400 nm/ em. 440 nm) relative to the CAThr-WT strain, having been zeroed against the background levels recorded for the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation.

## 4.4 MALDI-TOF MS/MS confirmation of pyoverdine production

### 4.4.1 Pyoverdine produced by highly functional C-A domain substitutions

The synthesis of wild type pyoverdine by strain CAThr-A and incorporation of Lys and Ser into pyoverdine by strains CALys-C and CASer-F was confirmed by mass spectrometry. Substitution of the terminal residue of pyoverdine for Lys and Ser was calculated to change the  $m/z$  of pyoverdine  $[M + H]^+$  ions to 1360.7 ( $C_{57}H_{90}N_{19}O_{20}$ ) and 1319.6 ( $C_{54}H_{83}N_{18}O_{21}$ ), respectively. Mass spectrum analysis of the supernatants from the C-A domain substitution strains (Figure 4-3) detected products of  $m/z$  1333.6 from CAThr-A, 1360.7 from CALys-C and 1319.6 from CASer-F, as well as the corresponding peptide fragments caused by RDA process degradation of the chromophore. From the CID spectra of the  $m/z$  1360.7 and 1319.6 pyoverdine species, the five abundant ions expected to contain the new Lys and Ser residues (Figure 4-4, indicated by arrows; Table 4-1, shaded cells) had changes in mass consistent with the altered residue. These results confirmed that the CALys-C and CASer-F domain substitution strains were incorporating new residues into pyoverdine.

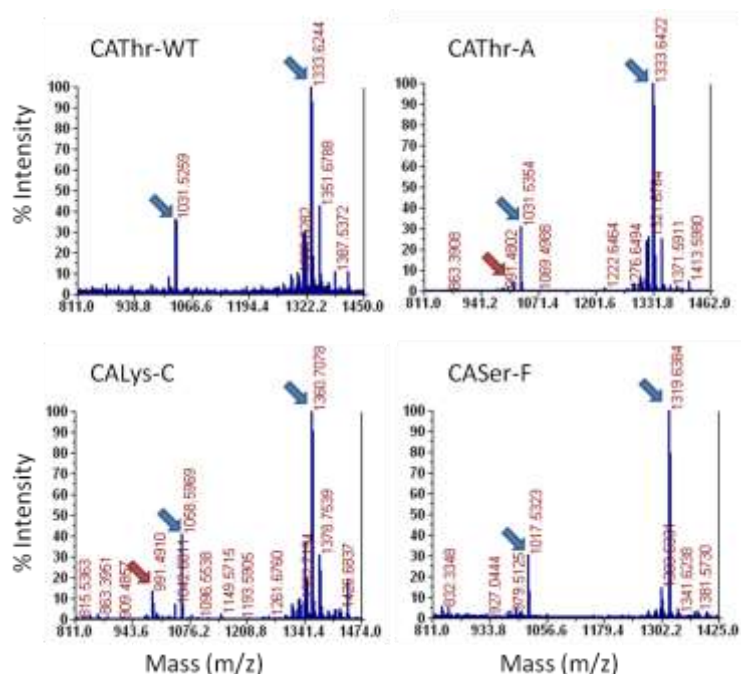


Figure 4-3. MALDI-TOF spectra obtained using the supernatant from functional C-A domain substitution variants. Blue arrows highlight the peaks consistent with the RDA peptide and pyoverdine. Red arrow indicates a truncated pyoverdine product that will be discussed in Section 4.4.2.

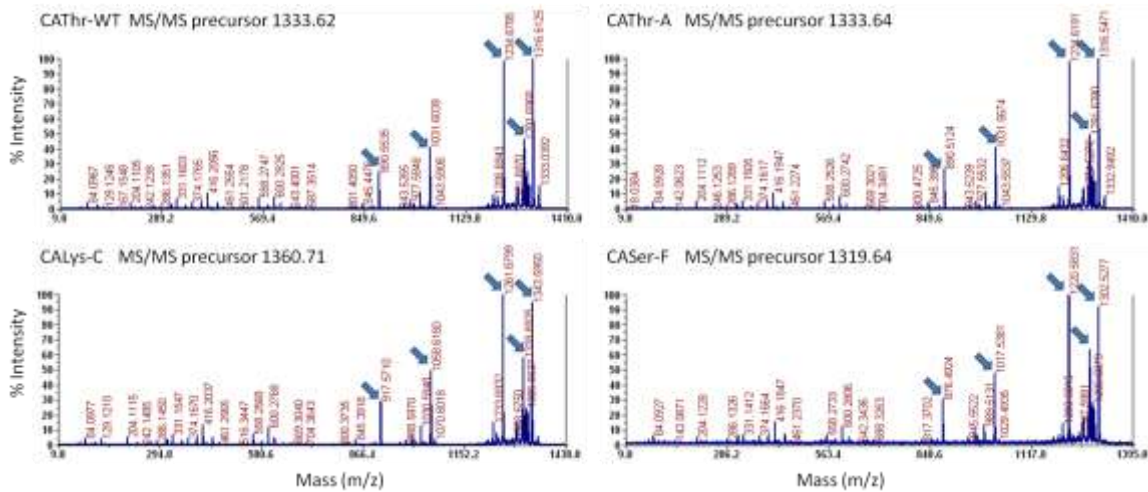


Figure 4-4. Fragmentation spectra of the main pyoverdine species from functional C-A domain substitution strains. Arrows indicate abundant ions used to confirm modifications to the C-terminus of pyoverdine.

Table 4-1. Ions detected by CID of major pyoverdine species produced from highly functional C-A domain substitution strains.

	Thr-WT	Thr-A	Lys-C	Ser-F
Ion	m/z			
[RDA + H - Suca] <sup>+</sup>	204.11	204.11	204.11	204.12
A <sub>1</sub> - Suca	317.19	317.18	317.18	317.16
A <sub>1</sub> - NH <sub>3</sub>	399.22	399.21	399.21	399.21
A <sub>1</sub>	416.21	416.19	416.20	416.18
B <sub>1</sub>	444.19	444.18	444.19	444.17
B <sub>2</sub>	600.29	600.27	600.28	600.28
B <sub>3</sub>	687.35	687.33	687.32	687.27
B <sub>4</sub>	845.45	845.40	845.39	845.36
Y <sub>7</sub> <sup>''</sup>	890.55	890.51	917.57	876.49
[RDAb + H] <sup>+</sup>	1031.60	1031.57	1058.62	1017.54
[M + H - Suca] <sup>+</sup>	1234.68	1234.62	1261.68	1220.58
[M + H - CH <sub>2</sub> N <sub>2</sub> ] <sup>+</sup>	1291.70	1291.64	1318.69	1277.61
[M + H - NH <sub>3</sub> ] <sup>+</sup>	1316.61	1316.55	1343.59	1302.53
MS/MS precursor	1333.62	1333.64	1360.71	1319.64

Shaded cells indicate ions expected to contain the residue added by the second module of PvdD. Suca, succinamide; RDA, chromophoric fragment from the RDA process; RDAb, peptide chain fragment released by the RDA process; ND, not detected.

#### 4.4.2 **Pyoverdine produced by C-A domain substitutions with limited activity**

Mass spectrometry was next used to detect pyoverdine in the supernatant of the C-A domain substitution strains that showed low levels of fluorescence when grown in liquid media (Figure 4-5). Unlike A domain substitutions, C-A domain substitutions are not limited by C domain acceptor site substrate specificity. Thus we predicted these might all be synthesising low levels of pyoverdine that had the terminal residue modified. However, neither wild type nor pyoverdine with the terminal residue modified were detected in the supernatant. Instead, a pyoverdine product of only 991.5 m/z was detected in all these strains except for strain CAAsp-G. The absence of pyoverdine from strain CAAsp-G fits with the fluorescence data showing no increase by CAAsp-G relative to the pyoverdine negative control (Figure 4-2, red bars). The 991.5 m/z product was also detected as a minor peak during the mass spectrometry of CAThr-A, CALys-C and CASer-F (Figure 4-3; although unlabelled in the spectra for CASer-F). The 991.5 m/z ion is equivalent to pyoverdine lacking the three last C-terminal residues ( $C_{41}H_{63}N_{14}O_{15}$ ; calculated 991.4592 m/z). This product has been previously detected as a breakdown product of pyoverdine, generated when the final Thr residue was not incorporated onto pyoverdine (Ackerley and Lamont, 2004). CID of the 991.5 m/z product identified ions consistent with the C-terminal ions missing the three terminal residues (Figure 4-6, Table 4-2). Although the less abundant ions were not detected with all strains, a near complete spectrum was obtained from strain CASer-B.

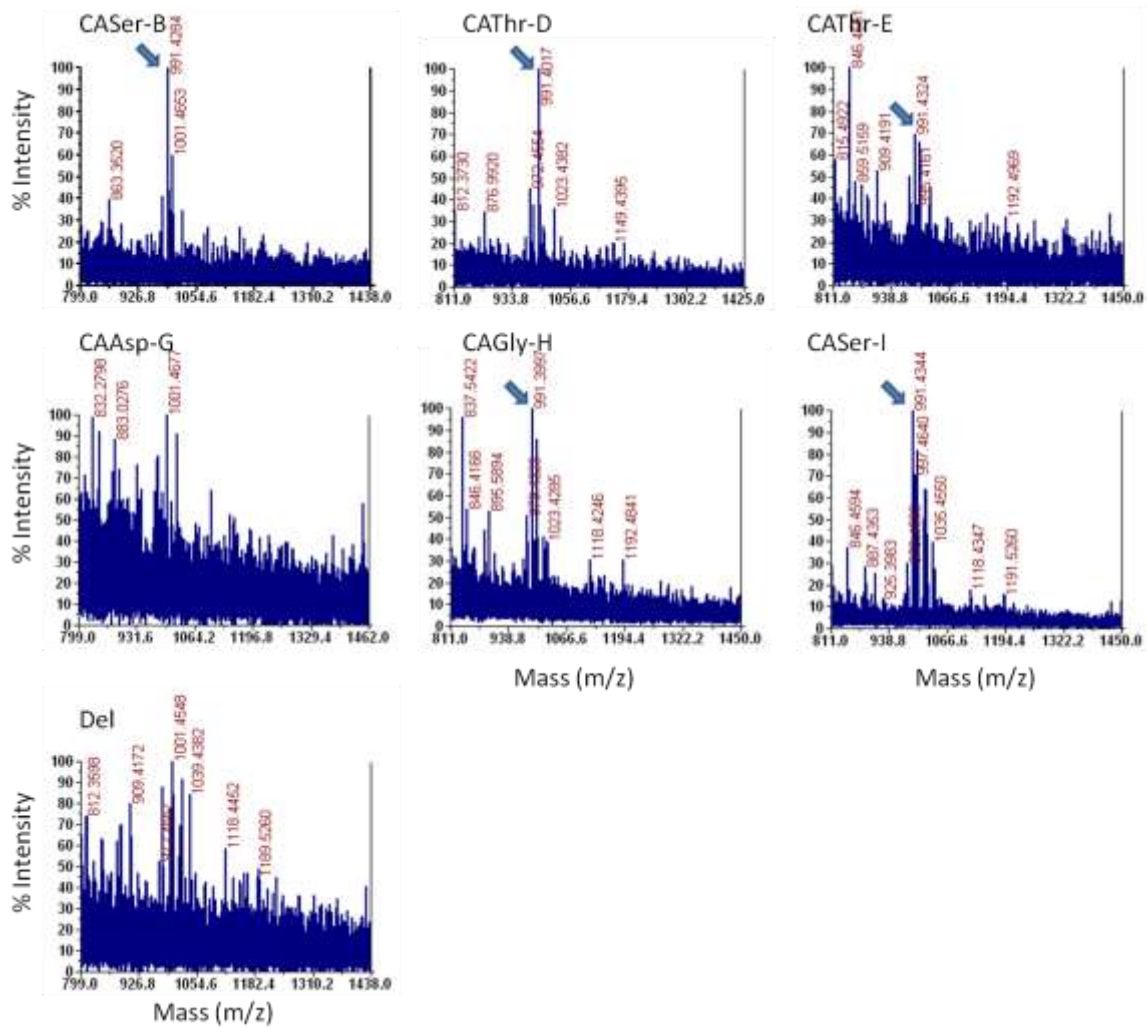


Figure 4-5. MALDI-TOF spectra obtained using the supernatant from functional C-A domain substitution variants with low levels or no fluorescence. Arrows highlight the truncated pyoverdine species.

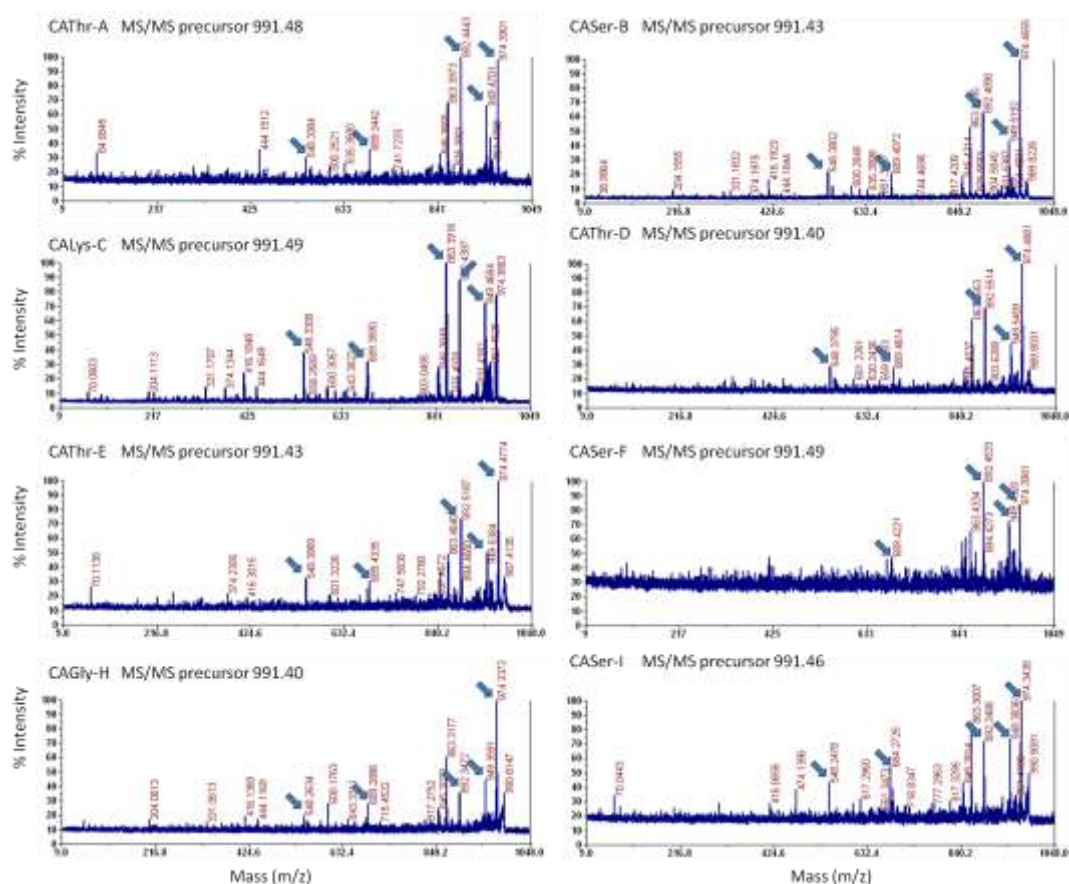


Figure 4-6. CID spectra of truncated pyoverdines obtained from C-A domain substitution strains. Arrows indicate ions corresponding to the five most abundant ions for wild type pyoverdine.

Table 4-2. CID of truncated pyoverdine produced by C-A domain substitution strains.

	Thr-A	Ser-B	Lys-C	Thr-D	Thr-E	Ser-F	Gly-H	Ser-I
Ion	m/z							
[RDA + H - Suca] <sup>+</sup>	ND	204.16	204.11	ND	ND	ND	204.06	ND
A <sub>1</sub> - Suca	ND	317.20	ND	ND	ND	ND	ND	ND
A <sub>1</sub> - NH <sub>3</sub>	ND	399.30	ND	ND	ND	ND	ND	ND
A <sub>1</sub>	ND	416.19	416.18	416.23	416.30	416.16	416.14	416.07
B <sub>1</sub>	444.15	444.18	444.16	ND	ND	ND	444.12	ND
B <sub>2</sub>	600.25	600.28	600.31	600.41	ND	600.35	600.18	600.24
B <sub>3</sub>	ND	ND	687.34	ND	ND	687.36	ND	687.29
B <sub>4</sub>	845.43	845.41	845.38	845.45	ND	ND	845.31	845.28
Y <sub>7</sub> ''	548.34	548.38	548.33	548.38	548.40	548.31	548.26	548.25
[RDAb + H] <sup>+</sup>	689.34	689.41	689.39	689.46	689.43	689.40	689.29	689.27
[M + H - Suca] <sup>+</sup>	892.44	892.49	892.44	892.55	892.52	892.43	892.35	892.35
[M + H - CH <sub>2</sub> N <sub>2</sub> ] <sup>+</sup>	949.47	949.52	949.47	949.54	949.54	949.47	949.36	949.38
[M + H - NH <sub>3</sub> ] <sup>+</sup>	974.39	974.47	974.39	974.49	974.48	974.40	974.34	974.34
MS/MS precursor	991.48	991.43	991.49	991.40	991.43	991.49	991.40	991.46

Shaded cells indicate ions expected to contain the residue added by the second module of PvdD. Suca, succinamide; RDA, chromophoric fragment from the RDA process; RDAb, peptide chain fragment released by the RDA process. \*Cells for ions detected by CID are empty due to corresponding precursor ion remaining undetected.

## 4.5 Discussion

In this chapter nine C-A domain substitution strains were created and tested for pyoverdine production. Three of these strains were able to incorporate the residue specified by the new A domain into pyoverdine. Of these, two incorporated novel amino acid residues into pyoverdine, confirming that modified pyoverdine peptides can be generated in high yield via C-A domain substitution. Moreover, the production of modified pyoverdines by strains CALys-C and CASer-F shows the new A domains in the C-A domain construct can load their predicted substrates onto the second T domain of PvdD and the TE domain can release the modified peptides without causing a substantial loss in activity. These results support the view that these A domains had the potential to be functional as A domain substitutions but were limited by C domain acceptor site specificity.

A direct implication of the hypothesis that A domain substitutions are mainly limited by substrate specificity is that the loss of function by C-A domain substitutions is caused by the introduction of the C domain. This was most likely the case for strains CAThr-D and CAThr-E since their respective A domain substitution strains were shown to be functional in Chapter 3. Of these, the loss of activity observed for CAThr-D could be explained in terms of the incompatible donor site stereo specificity of the new C domain (since it normally receives D-Thr). In contrast it is unknown why the C-A domain substitution for CAThr-E was non-functional. Moreover, similar to the A domain substitution results, it was expected that C-A domain substitution strains exhibiting low levels of fluorescence would be synthesising full-length pyoverdine modified at the final residue. The production of a truncated product by these C-A domain substitutions supports that the cause for loss in function is not associated with the A domain.

There are potentially a few reasons for how C-A domain substitutions could produce a truncated product. The absence of this product and decreased fluorescence exhibited by the *pvdD* deletion strain suggest that the truncated product is released from the modified PvdD enzymes. The reason for no release from PvdJ is unknown, but could be due to lower release of stalled peptides from T domains that act *in trans* or that a downstream C domain was involved in release. A potential way the truncated pyoverdine could form, after release from PvdD, is if changing the final residue of pyoverdine creates an unstable pyoverdine that degrades to a more stable product. This seems unlikely because

pyoverdines with a terminal Thr, Lys or Ser residue were able to be detected, suggesting a relatively high tolerance for alternative terminal residues and that strains CAThr-E, CASer-B and CASer-I could have synthesised a pyoverdine sufficiently stable for detection. Instead, we believe the detection of the truncated product and absence of full-length modified pyoverdines suggests that addition of a new C domain is causing peptide assembly to stall on the first T domain of PvdD, i.e. immediately upstream to the introduced C domain. As a result, this would mean condensation is not occurring at the newly introduced C domain and the pyoverdine missing the terminal Thr is released. As seen previously, the pyoverdine missing the terminal residue would partially degrade to form a pyoverdine missing the terminal three residues (Ackerley and Lamont, 2004). Based on this hypothesis, in trying to understanding why C-A domain substitutions in PvdD lose activity, we should look for factors that may cause peptide synthesis to stall. Three potential causes for peptide synthesis stalling are discussed below.

Previous work on domain substitution in PvdD suggested donor site specificity of the C domain could be a cause for the loss in activity of C-A domain substitutions (Ackerley and Lamont, 2004). Some specificity towards the terminal residue of the incoming peptide has been identified towards the C-terminus residue (Stein et al., 2005). In this regard the CAThr-E construct is very similar to PvdD module 2, in that the new C domain in its native context receives L-Thr residues at both the donor and acceptor sites. That this strain produced the truncated product suggests it was non-functional due to a reason other than C domain donor site substrate specificity. Moreover, both the functional CAThr-A and CALys-C strains contain C domains that normally receive hfOrn as the C-terminal residue in their donor sites, yet are clearly able to cope with L-Thr being in that position. In fact, despite the C domain from strain CASer-F naturally receiving Thr at its donor site, the two strains CAThr-A and CALys-C produced the highest levels of pyoverdine. The loss of function by strain CAThr-E, and strains CAThr-A and CALys-C being the most functional substitutions indicates factors other than C domain donor site specificity are limiting the activity of C-A domain substitution strains.

A second possible explanation for why peptide synthesis might have stalled is that C domains in a new context could be unable to accommodate some incoming peptides due to steric constraints. This hypothesis is related to substrate specificity, however is based on an entire peptide (not just the C-terminal residue) being unable to fit within the substrate channel of the new C domain. Structural studies of C domains show the

catalytic centre covered by a lid region within a channel between two subdomains (Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008), and it may be that some new incoming peptide chains are physically blocked from reaching the catalytic centre. Moreover, previous *in vitro* studies on C domain donor site specificity focused on single residues (Belshaw et al., 1999; Ehmann et al., 2000) or at most a tetrapeptide (Stein et al., 2005), rather than the larger peptide chains that substituted C domains further down an NRPS assembly line would normally receive. An experiment, which suggests peptide size can be important, tested the stereo specificity of the fifth C domain from tyrocidine synthesis (Clugston et al., 2003). For this C domain, there was no stereo specificity when an aminoacyl-CoA was loaded onto the upstream module. In contrast, when the native tetrapeptide containing either a terminal D- or L-amino acid was loaded onto the upstream module, only the D-isomer containing pentapeptide was formed; a result likely caused by racemisation of the terminal tetrapeptide residue by the upstream E domain and stereo specificity by the C domain. In addition, when the E domain of the upstream module was mutated to become non-functional, only the tetramer with a terminal D-amino acid was able to produce the peptapeptide. This suggests that, as well as stereo specificity, the size of an incoming peptide can be an important factor. Further tests of C domain donor site specificity using a range of peptides rather than single amino acids may be needed to interrogate the tolerance of C domains for alternative peptide chains.

A third possible cause for the loss of activity observed with most C-A domain substitutions is that introduced C-A domains may not communicate correctly with the neighbouring PvdD T domains. During a catalytic cycle T domains function to transfer substrates between multiple domains. Crystal structures of T domains suggest large conformational changes are needed for the T domain to transfer its substrate between catalytic domains (Section 1.2.2.1) and domain substitution experiments suggest T domains can be specific to the type of downstream domain (Section 1.3). These interactions are not completely understood, but suggest the T domain from the first module of PvdD may not be able to interact properly with the new C domains causing peptide synthesis to stall.

The work in this chapter confirms that modified pyoverdines can be created by C-A domain substitution. However unlike A domain substitutions, which appear limited by C domain acceptor site specificity, it is difficult to identify a specific cause for why the

majority of C-A domain substitutions lose activity. The potential causes raised in this discussion were that C domains may be inflexible in the peptide chains they can receive or may disrupt interactions with the upstream T domain. Of these hypotheses, the second could easily be bypassed because substituting T-C-A domains as a functional catalytic unit would keep the T-C domain interactions intact. One example of a previously functional T-C-A domain substitution has been published (Doekel and Marahiel, 2000) in which two module NRPS pathways were created and tested for activity. In total, one T-C-A and two T-C-A-T-TE domain substitutions were made and found to be functional in *in vitro* assays. However, without a larger sample size and direct comparison to C-A domain substitutions, it is unclear whether T-C-A domain substitution may offer a better alternative to C-A domain substitution. The following chapter compares the success of T-C-A versus C-A domain substitution, and examines rules governing interactions between T domains and downstream TE and C domains.



## Chapter 5: Investigation of the influence of T domains on the activity of C-A domain substitutions

### 5.1 Introduction

In the preceding chapters, staging plasmids were constructed in which A and C-A domains could be introduced into the second module of *pvdD*. For A domain substitution strains, the major cause for loss of function appeared to be due to the acceptor site specificity of the C domain from PvdD. In contrast, the reasons for loss of function for C-A domain substitutions were less clear.

Based on the production of a truncated product by the “non-functional” C-A domain substitution strains, it was hypothesised that the introduction of a heterologous C domain can cause peptide synthesis to stall, due to disruption of the ability of the T domain to pass its tethered substrate to the downstream C domain. This chapter tested the ability of T domains to interact with new catalytic partners. A particular focus was whether the T domain may be inhibited in its ability to interact with new downstream C domains introduced by C-A domain substitution.

#### 5.1.1 Additional nomenclature

Two new nomenclatures, described below and summarised in Figure 5-1, are used in this chapter to identify the type of domain immediately downstream to a T domain or to refer to specific interactions between a T domain and a specific catalytic partner.

The first nomenclature recognises the potential specialisation of T domains according to the type of domain located immediately downstream. When relevant, the type of downstream domain will be indicated in subscript. For example, T domains with C or E domains located immediately downstream will be referred to as T<sub>C</sub> or T<sub>E</sub> domains, respectively. Furthermore, to differentiate T<sub>C</sub> domains that interact *in cis* with a downstream C domain from T domains that interact *in trans* with a downstream C domain, the latter will be referred to as T<sub>Ct</sub> domains (i.e. unless specifically marked T<sub>Ct</sub>, a T<sub>C</sub> domain can be assumed to be linked to the downstream C domain *in cis*). The reason for this distinction is that T<sub>Ct</sub> domains may have additional recognition elements, in addition to COM domains, to allow them to associate with the correct downstream C domain (Chiocchini et al., 2006; Koglin and Walsh, 2009).

The second nomenclature is to distinguish between T domain interactions with different catalytic partners. In this context the term “interactions” refers to the ability of the T domain to receive its substrate from the upstream A domain or to then pass that substrate to another catalytic partner domain. If a domain substitution is functional, that indicates the new T domain interactions created by the domain substitution are not substantially inhibited. Five different interactions that are considered in this chapter are as follows:

- I. A/T domain interactions; the ability of the T domain to receive a substrate activated by the upstream A domain.
- II. C-/T domain interactions; the ability of a T domain to pass its substrate to the upstream C domain for condensation to occur (the “-” in C-/T indicates that these domains are not immediately adjacent to one another in the protein primary sequence).
- III. T/C domain interactions; the ability of a T domain to pass its tethered substrate to a downstream C domain.
- IV. T/TE domain interactions; the ability of a T domain to pass its tethered substrate to a downstream TE domain.
- V. T/E domain interactions; the ability of a T domain to pass its substrate to a downstream E domain for epimerisation.

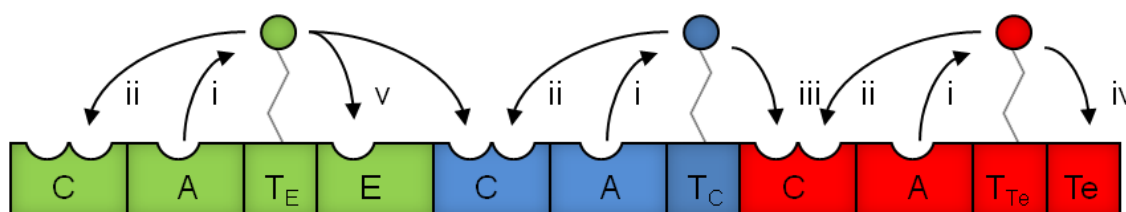


Figure 5-1. Nomenclature for T domain types and interactions shown for a hypothetical three module NRPS. Three T domain types are shown as T<sub>E</sub>, T<sub>C</sub> and T<sub>TE</sub> domains depending on the domain immediately downstream. T<sub>Ct</sub> domains are similar to T<sub>C</sub> domains except interact with the downstream domain *in trans*. The interactions between T domains and catalytic domains are labelled as per the description in the above text.

### 5.1.2 Evidence of T domain interactions being inhibited by domain substitution

As discussed in Section 1.2.2.1, the previously solved structures of T domains in association with surrounding domains indicate that T domains undergo dramatic conformational changes to be within a functionally relevant distance of all their surrounding catalytic partners. NMR (Frueh et al., 2008) and X-ray crystal (Liu et al.,

2011) structures of the T-TE domains from EntF have shown that the T-TE domains have a tight interface, and it has been suggested that this may explain why disrupting the T/TE junction destroys activity (Strieker et al., 2010). In addition, alignment (Owen, 2010) and domain substitution studies (Section 1.3) suggest T<sub>E</sub> and T<sub>TE</sub> domains are particularly specialised.

The interactions between T domains and their catalytic partners can be highly sensitive to single residue changes. The sensitivity of T domain interactions was previously tested using the T<sub>TE</sub> and TE domains of EntF, a single module NRPS involved in enterobactin synthesis (Zhou et al., 2006). By randomly mutating single residues of the EntF T<sub>TE</sub> domain to Ala, a process referred to as Ala scanning mutagenesis, it was found that key residues at positions +1, +21 and +24 relative to the conserved Ser residue that is the 4'-pp attachment site were intolerant of change for the *in vivo* production of enterobactin. *In vitro* assays determined that changing the +1 position to Ala impaired attachment of the 4'-pp group by the PPTase, and mutations at the +21 and +24 positions inhibited T/TE domain interactions (Zhou et al., 2006). Similar results were obtained using Ala scanning mutagenesis on the aryl-carrier protein (ArCP) domain of the NRPS EntB (Lai et al., 2006). Although not a T domain, the EntB ArCP is similar to a T<sub>Ct</sub> domain as it passes the attached substrate to the *in trans* C domain of EntF. Important residues of the ArCP were identified at the +4 and +23 positions relative to the conserved Ser, and the +19 position to a lesser extent, such that mutation of these residues inhibited the ArCP from passing the substrate to the downstream C domain. The +19 to +24 residues from both these studies are located within helix III of the carrier protein domains, indicating that helix III is particularly important for interaction with downstream domains (Lai et al., 2006; Zhou et al., 2006).

Later work substituted the ArCP of EntB for two alternative ArCPs (Zhou et al., 2007). Enterobactin functions as a siderophore, and the two variants with substituted ArCPs showed severely impaired growth in the presence of an iron chelator. The two ArCPs were mutated by error-prone PCR and improved variants in enterobactin production selected for by growth on media containing this iron chelator. After three rounds of mutagenesis and selection, rates of growth for both variant ArCP strains went from being severely inhibited on this medium to approximately half that of the wild type strain. Strains containing variants of one of the ArCPs were analysed for *in vitro* enterobactin synthesis and found to have a ~500 fold increase in the rate of enterobactin synthesis

compared to the corresponding non-evolved ArCP strain. Convergent mutations improving activity for both ArCPs were identified at the -8, +1, +23 and +24 positions relative to the conserved Ser residue. Analysing the -8, +1 and +24 mutants using *in vitro* assays revealed that the -8 position improved the interaction equivalent to an A/T domain interaction, the +1 position with the PPTase and the +24 position improved interactions equivalent to A/T and T/C domain interactions. Overall it was concluded that three different locations of the ArCPs interact with the PPTase, A domain and C domain (Zhou et al., 2007).

The sensitivity of T domain interactions to single residue mutations raises the question: how portable are T domains in relation to new domains? In addition to the work described above, the portability of T domains was previously tested by making domain substitutions within the yersiniabactin synthesising NRPS pathway (Suo, 2005). There are three T domains involved in yersiniabactin synthesis: the first is upstream to a cyclisation (Cy) domain, the second upstream to a ketoacyl reductase domain (a polyketide synthetase domain) and the third upstream to a TE domain. Substituting the first T domain with either the second or third T domains resulted in diminished yersiniabactin synthesis, at a rate of 30 % or 11 % of the wild type enzyme, respectively. Substituting the second T domain with the first or third T domains abolished the cyclisation activity of the upstream Cy domain, causing synthesis of a non-cyclised product at a reduced rate (Suo, 2005). This study showed that T domain substitutions can reduce activity. However, all these T domains differed in terms of the domain found immediately downstream in their native context, suggesting the substituted T domains may have been incompatible types.

#### 5.1.2.1 T domain substitutions and evolution in BpsA

Previous work in the Ackerley lab, by Dr Owen, tested T domain interactions using blue pigment synthetase-encoding gene A (*bpsA*), that he developed for use as an NRPS domain substitution model. BpsA is a single module NRPS expressed by *Streptomyces lavendulae* ATCC11924, which functions to dimerise two L-Gln residues into the blue pigment indigoidine (Figure 5-2) (Takahashi et al., 2007). As can be seen in Figure 5-2A, the A domain is unusual in that an oxidation (Ox) domain is located between motifs A8 and A9. This Ox domain functions to dehydrogenate the side-chain of L-Gln during indigoidine synthesis (Figure 5-2B). Homologous genes have been identified in other

indigoidine synthesising bacteria, including *Erwinia chrysanthemi*, *Vogesella indigofera* and *Phaeobacter* sp. Strain Y4I (Cude et al., 2012; Reverchon et al., 2002).

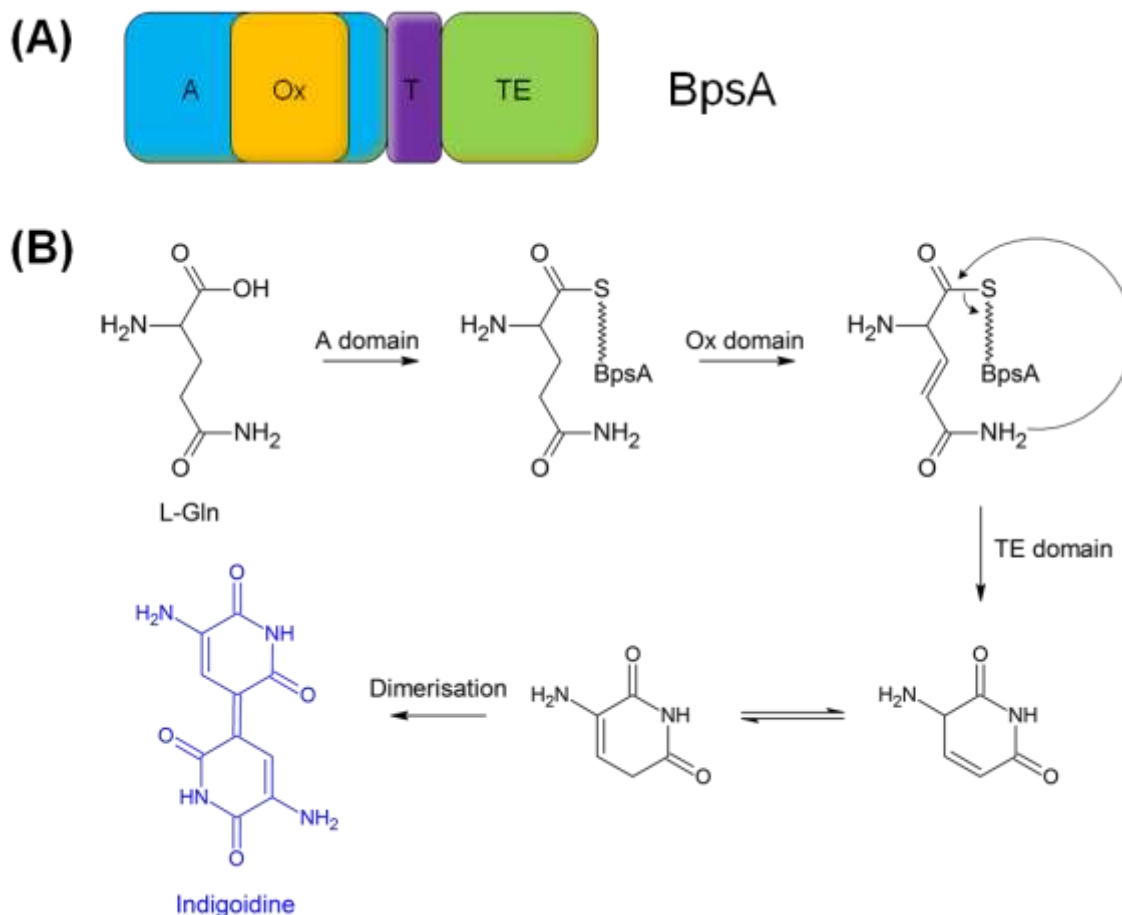


Figure 5-2. BpsA and indigoidine synthesis. (A) Domain architecture of the BpsA enzyme (Takahashi et al., 2007). The oxidation (Ox) domain highlighted in yellow is located between the A8 and A9 motifs of the A domain. (B) Proposed mechanism of indigoidine synthesis. The A domain activates L-Gln and attaches it to the T domain of BpsA. Two carbon atoms of the L-Gln side-chain are dehydrogenated by the Ox domain to form a double bond. The L-Gln residue is then cyclised during release by the TE domain. The released intermediates spontaneously dimerise in the presence of oxygen to form indigoidine. Basic mechanism as per Walsh and Wenciewicz (2012) with additional details derived from Reverchon et al. (2002) and Takahashi et al. (2007).

In Dr Owen's work (2010), the T<sub>TE</sub> domain from BpsA was substituted for six alternative T domains and one ArCP. Three of these were T<sub>TE</sub> domains and four (including the ArCP) were T<sub>C</sub> domains. The BpsA variants substituted with alternative T<sub>TE</sub> domains retained low levels of activity *in vivo*, whereas the ArCP and T<sub>C</sub> domains were completely inhibited in activity (Table 5-1). Among the T domains substituted into BpsA

were the two T domains from *pvdD*. It was shown that the activity of the BpsA variant containing the T<sub>C</sub> domain from the first module of *pvdD* could be significantly restored by directed evolution. From the improved T<sub>C</sub> domain variants, the +4 and +24 residues relative to the conserved Ser were found to be particularly important for restoration of activity. For example, one of the most successful mutants, named 3kf0(I/I), contained T→I and F→I mutations at the +4 and +24 positions, respectively, and exhibited 56 % *in vivo* activity relative to wild type BpsA. In naming 3kf0(I/I) and other BpsA variants, the letters in brackets indicate the new residues introduced at the +4 and +24 positions, respectively. Based on homology modeling, Dr Owen suggested that the +4 and +24 residues were located at the T/TE interface and may be important for T/TE domain interaction. When he analysed T domain sequences from a range of pseudomonads, he found T domain sequences were conserved differently at the +4 and +24 locations depending on whether the immediately downstream domain was a C, E or TE domain (Table 5-2). These results suggested the +4 and +24 positions may generally be vital for T/TE domain interactions.

Table 5-1. *In vivo* synthesis of indigoidine by BpsA variants.

Recombinant BpsA proteins made	
Substitution	<i>In-vivo</i> activity*
<i>E. coli</i> EntF PCP	5.1 +/- 0.3 %
<i>P. syringae</i> 1926 PCP	5.3 +/-0.3 %
<i>P. aeruginosa</i> PvdD PCP2	4.9 +/-0.4 %
<i>P. aeruginosa</i> PvdD PCP1	0 %
<i>P. syringae</i> 1926 PCP2	0 %
<i>B. subtilis</i> DhbF PCP1	0 %
<i>E. coli</i> EntB ArCP	0 %

Shaded cells indicate T<sub>TE</sub> domains, whereas unshaded cells are T<sub>C</sub> domains. *P. aeruginosa* PvdD PCP1 and PCP2 correspond to the first and second T domains from *pvdD*, respectively. \*Activity measured as the relative amount of indigoidine present in the supernatant after growth for 16 to 20 hours. Indigoidine levels were estimated based on levels of absorbance at 590 nm. Table provided by Dr Owen.

Table 5-2. Residues found at the +4 and +24 positions of *Pseudomonas* pyoverdine NRPS T domains relative to the 4'-pp attachment site.

Domain type	+4	+24
T <sub>C</sub> or T <sub>Ct</sub>	Variable (T,V,M,S,A)	F (1xL)
T <sub>E</sub>	L, I	F
T <sub>TE</sub>	V, L, I, M	F, L, I, M*

Data taken from *Pseudomonas* T domain alignments presented in Owen (2010). Sample size for each type of T domain ranged from 16-20. \*Differs from Table A1.1 in Owen (2010), because the entry in the table for this position was incorrect based on the sequence alignments presented.

## 5.2 Chapter overview and aims

All the previous studies discussed above have focused either on T domain interactions with domains other than C domains or the *in trans* interaction between ArCPs and a C domain. They suggest that T domains can be specialised in their interactions with downstream domains, and that substantial losses in activity can be caused by small changes that inhibit T domain interactions with catalytic partners. Typically C-A domain substitution results in a new T/C domain interaction, i.e. the upstream T<sub>C</sub> domain having to interact with a new downstream C domain. The primary aim of this chapter was to determine whether C-A domain substitutions frequently lose activity as a consequence of the new T/C domain interaction being impaired.

This chapter begins with a contribution to work that was initiated by Dr Owen (2010), a previous doctoral student under Associate Professor David Ackerley, examining T/TE domain interactions. Dr Owen's preliminary studies of T/TE domain interactions were further expanded here to determine the generality of his previous results. Based on those results, it was then hypothesised that C-A domain substitutions may inhibit T/C interaction and, as a consequence, that T-C-A domain substitutions may prove a more successful strategy for generating functional recombinant NRPS templates. To address this I generated three T-C-A domain substitution variants and tested the activity of these in comparison to the corresponding C-A domain substitution variants. The small sample size of three was a consequence of the PCR amplification and cloning being technically difficult for large T-C-A domain substitution constructs. Consequently, an alternative strategy of T domain substitutions in the first module of *pvdD* was adopted to investigate the effect of disrupting T/C domain interactions. Substitution of T<sub>C</sub> domains into the first module of PvdD creates new T/C domain interactions (with the C domain of PvdD module 2) that are equivalent to the new T/C domain interactions caused by C-A domain

substitutions. The only real difference is that substitution of T<sub>C</sub> domains also affects the upstream (module 1) A/T domain interaction, whereas C-A domain substitutions affect the downstream (module 2) A/T domain interaction; however, the results of Chapters 3 and 4 suggest that disruption of A/T domain interactions is not a major factor in impairing recombinant PvdD function.

As an adjunct to the T<sub>C</sub> domain substitution study, T<sub>CT</sub>, T<sub>E</sub> and T<sub>TE</sub> domains were also substituted into the first module of *pvdD* to test whether other types of T domain can function upstream to a C domain.

### 5.3 Substitution of evolved *pvdD* T domains back into *pvdD*

Work by Dr Owen (Owen, 2010) found that T<sub>C</sub> domains did not function upstream to the T<sub>E</sub> domain of BpsA, and that indigoidine synthesis could be restored to BpsA T<sub>C</sub> domain substitution variants by a few point mutations. To expand this work, using a different model system, it was decided to substitute the T<sub>C</sub> domain from the first module of *pvdD* for the T<sub>TE</sub> domain from the second module. In addition, variants of the *pvdD* T<sub>C</sub> domain that had been evolved for improved function in *bpsA*, carrying mutations at the +4 and +24 positions, were substituted at the *pvdD* T<sub>TE</sub> position. These substitutions are referred to as T2 domain substitutions to indicate their location in the second module of *pvdD*. The evolved T domains used for T2 domain substitutions were sourced from the BpsA variants 3kf0(I/I), 3PF2(T/L) and 3kf13(A/L). The BpsA variants 3PF2(T/L) and 3kf13(A/L) exhibit levels of *in vivo* indigoidine synthesis between 15 - 25 % that of the most active variant 3kf0(I/I) (Dr Owen, personal communication) and were selected to provide T domains from moderately active variants.

The T2 domain substitution plasmid, named pST2 (Figure 5-3), was constructed as per Section 2.8.2.1. It featured the same NotI restriction site as the A and C-A domain substitution plasmids, and a SpeI restriction site added between the T and T<sub>E</sub> domain of the second module. The T<sub>C</sub> domain from *pvdD* and its three improved variants were each amplified by PCR and ligated into the pST2 plasmid using a NotI/NotI fusion at the 5' end and NheI/SpeI fusion at the 3' end. This NheI/SpeI fusion changed the coding sequence from ER to AS, however, since this change was within the flexible linker region it was thought to be unlikely to influence activity. As a restriction site control, the T<sub>TE</sub> domain from *pvdD* was PCR amplified and ligated into the pST2 plasmid. When

transformed into the *pvdD* deletion strain, this control showed no reduction in pyoverdine production compared with wild type *P. aeruginosa* PAO1 (not shown).

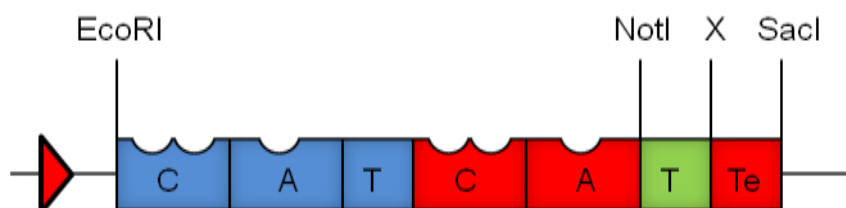


Figure 5-3. The domain organisation within the substitution plasmid pST2. Domains coloured blue and red were derived from the first and second module of *pvdD*, respectively. Domains in green were absent in the substitution plasmid and indicate the location where alternative domains would be inserted. X refers to a SpeI restriction site that was destroyed when ligating T domains into the plasmids via a compatible NheI sticky end.

### 5.3.1 Growth on solid media of T2 domain substitution strains

The four T2 domain substitution constructs and the restriction site control construct were transformed into *P. aeruginosa* PAO1  $\Delta pvdD$ , and the resulting strains tested for function by growth on solidified KB media (Figure 5-4). The strain containing the substituted  $T_C$  domain showed low levels of increased fluorescence compared to the *pvdD* deletion strain. This strain also exhibited low levels of growth on EDDHA supplemented media. In comparison, the three T2 domain substitutions with  $T_C$  domains that had been evolved for improved function in BpsA, were found to have increased levels of growth and fluorescence relative to the  $T_C$  domain substitution. The substitution strain containing the 3kf0(I/I) T domain, appeared to be the most active as measured by fluorescence and its ability to grow in the presence of EDDHA.

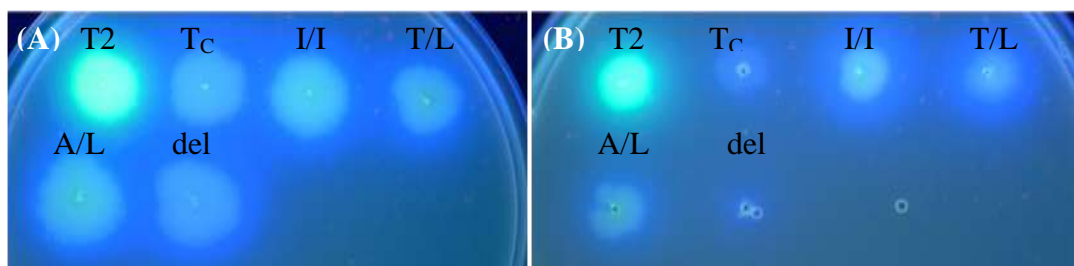


Figure 5-4. Fluorescence of T2 domain substitution strains on KB agar media in the (A) absence or (B) presence of EDDHA. Plates were inoculated with the T2 domain restriction site control (T2), T2 domain substitution strains, and the *pvdD* deletion mutant (del) then incubated for 24 h at 37 °C.  $T_C$  refers to the *pvdD* T1→T2 substitution, whereas residues at the +4 and +24 positions are indicated for evolved T domains. Photographs were taken under UV light.

### 5.3.2 Pyoverdine production in liquid media by T2 domain substitution strains

The T2 domain substitution strains were grown in liquid media to quantitatively assess pyoverdine production relative to the T2 restriction site control (Figure 5-5). The *pvdD* variant containing the substituted T<sub>C</sub> domain from *pvdD* exhibited approximately 6 % of wild type levels of pyoverdine, as measured by absorbance. Although function was substantially impaired, this result contrasted with T<sub>C</sub> domain substitutions in BpsA, which had in each case resulted in no indigoidine synthesis at all. The substitution variants with the T<sub>C</sub> domain evolved for improved activity in BpsA all synthesised relatively similar levels of pyoverdine, between 18 - 24 % of wild type levels based on absorbance. This indicated that increased fluorescence by the 3kf0(I/I) T domain substitution strain when grown on solid media may have been due to the subjective nature of the test. Nonetheless, these results confirm that T<sub>C</sub> domains evolved to function upstream to the BpsA TE domain can also have improved function upstream to the *pvdD* TE domain.

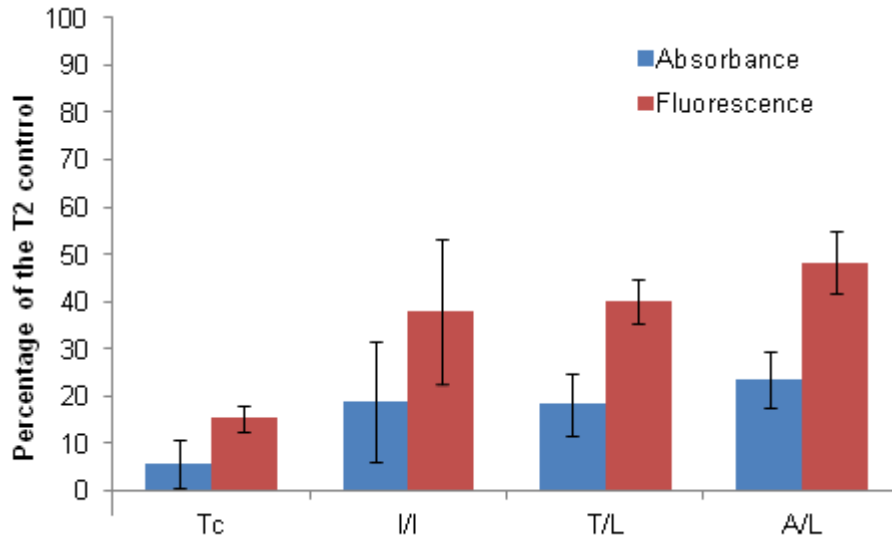


Figure 5-5. Pyoverdine production of T2 domain substitutions detected when grown in liquid media. T<sub>C</sub> refers to the *pvdD* T1→T2 substitution, whereas residues at the +4 and +24 positions are indicated for evolved T domains. Values are expressed as a percentage relative to the absorbance (400 nm) or fluorescence (ex. 400 nm/ em. 440 nm) measured for the T2 domain restriction site control strain, having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation. T<sub>C</sub>, *pvdD* T1→T2 substitution; I/I, 3kf0(I/I); T/L, 3PF2(T/L); A/L, 3kf13(A/L)

### 5.3.3 Mass spectrometry of T2 domain substitution strains

The T2 domain substitutions were next analysed by mass spectrometry to confirm the synthesis of wild type pyoverdine (Figure 5-6). All four T2 domain substitution strains were confirmed to be synthesising wild type pyoverdine by the presence of a peak corresponding to 1333.6 m/z. It was decided to not present CID data as no new pyoverdines were synthesised and the previous chapters showed that the 1333.6 m/z product was only present when wild type pyoverdine is synthesised.

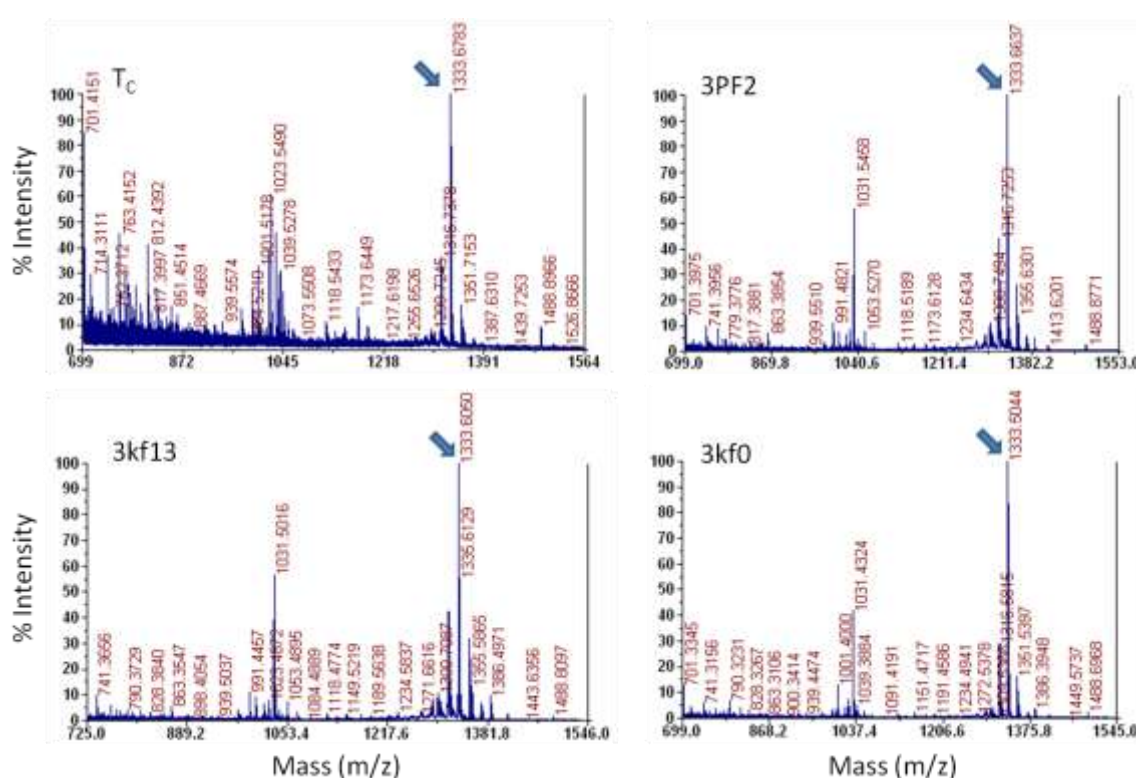


Figure 5-6. Mass analysis of T2 domain substitution strains. Arrows indicate the peak corresponding to wild type pyoverdine.

## 5.4 Bioinformatics of T domains

As discussed previously, bioinformatics of T domains by Dr Owen focused on T domains from pseudomonads and highlighted differences in the +4 and +24 residues relative to the Ser residue that serves as the 4'-pp attachment site dependent on the type of downstream domain. The reason for limiting sequences to a single genus was to highlight the differences in sequence that were specific for interaction with the downstream domain,

based on the assumption that the pyoverdine NRPS enzymes have evolved via repeated gene duplication and/or recombination events (Smith et al., 2005), leading to accelerated mutation of sequences required for optimal inter-domain interactions between recombined modules (Owen, 2010).

To determine whether the +4 and +24 positions were conserved dependent on the type of T domain for other bacteria, more extensive alignments were made using T domains from a diverse range of bacterial species (Figure 5-7, Appendix C). In addition, the residues implicated for T domain interactions in other studies, namely the -8, +1, +19, and +23 positions relative to the 4'-pp attachment site (Lai et al., 2006; Zhou et al., 2006, 2007), were also examined. In these alignments, the residues at the +1, +4, +19 and +23 positions were highly similar between T<sub>C</sub> and T<sub>TE</sub> domains, but each of these differed from T<sub>E</sub> domains. In contrast, the -8 position contained similar residues for all T domain types, and the +24 position was highly conserved as Phe for all T domains. These observations suggested that, despite the functional differences observed in both BpsA and PvdD assays described in this chapter, T<sub>C</sub> and T<sub>TE</sub> domains are highly similar to one another.

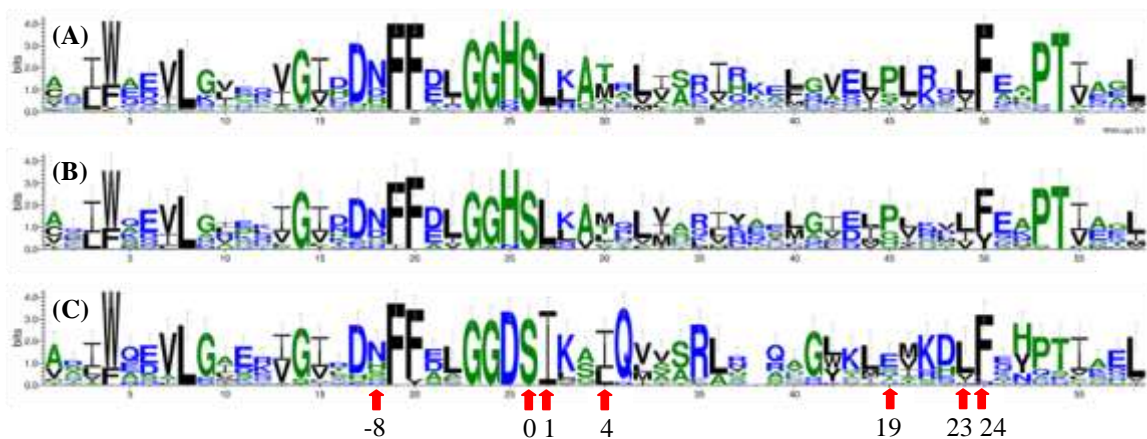


Figure 5-7. Sequence logos of T domains from bacterial strains other than pseudomonads based on clustalW alignments. All T domains were aligned together and separate graphs created for (A) T<sub>C</sub> domains, (B) T<sub>TE</sub> domains or (C) T<sub>E</sub> domains. Full alignments are shown in Appendix C. Each position of the sequence alignment in Appendix C is indicated by one column in the sequence logo. The height of the column indicates the sequence conservation measured in bits, and the height of individual letters within a column indicates their relative frequency at that position (Crooks et al., 2004). Residues implicated in other studies to be important for interaction with other domains are indicated by arrows. Image created using Weblogo (Crooks et al., 2004).

The similarity between T<sub>C</sub> and T<sub>TE</sub> domains suggests that these domains may be able to interconvert between types relatively easily. Despite biochemical evidence of key residues for T/TE domain interactions (Owen, 2010; Zhou et al., 2006), the sequence alignments showed that residues at these positions are conserved similarly between T<sub>C</sub> and T<sub>TE</sub> domains. This indicates that the residues important for T/TE domain interactions might be more specific for each T-TE domain combination, rather than optimisations for general T/TE domain interaction. This is highlighted by the original BpsA T domain substitution work of Dr Owen where, despite T<sub>TE</sub> domains retaining function in their new environment, indigoidine synthesis was decreased to approximately 5 % of wild type levels (Table 5-1). Given that T/TE domain interactions appear important for function and that T<sub>C</sub> and T<sub>TE</sub> domains are similar, this raised the questions of whether T/C domain interactions may also be prone to disruption and, consequently, whether C-A domain substitutions might disrupt T/C domain interactions.

Based on the T domain mutagenesis studies discussed in Sections 5.1, an inhibited T/C domain interaction could probably be restored by introducing only a few point mutations. However, if the important residues for correct T/C interaction differ between different T-C domain combinations, then this may require optimisation for each new T-C domain combination. An alternative strategy that might prove effective would be to keep evolutionarily optimised T/C domain interactions intact by substituting T-C-A domains together.

### 5.5 Attempts to avoid disruption of the T/C interface via T-C-A domain substitutions

This section compares the activity of T-C-A with C-A domain substitutions. T-C-A domain substitutions alter upstream T domain interactions of two T domains, e.g. the A/T and C-/T domain interactions for both the T domain introduced as part of the T-C-A domains and the T domain of the second module of *pvdD*. However, T-C-A domain substitutions keep the T/C domain interactions intact. As a consequence, if disrupting the T/C domain interactions cause C-A domain substitutions to lose activity, then T-C-A domain substitutions would likely be more successful. The 100 % success rate observed for five threonine-specifying A domain substitutions into PvdD (Ackerley and Lamont, 2004; Chapter 3) suggests that disruption of an established A/T domain

interaction is not a major factor limiting recombinant NRPS function. It is less clear whether the C-/T domain interaction may be disrupted, however, the T/C domain interaction was hypothesised to be more important due to the importance of T domain interactions with downstream domains as discussed in Section 5.1.

As mentioned earlier in Section 1.3.3, one T-C-A domain substitution has previously been created in the context of a dimodular enzyme and tested *in vitro* (Doekel and Marahiel, 2000). This previous work showed that T-C-A domain substitution can be successful, but differs from the experiments in this section as it did not seek to determine whether T-C-A domain substitution may offer a better alternative to C-A domain substitution and, as such, did not compare activity of the T-C-A domains substitution with an equivalent C-A domain substitution.

To make T-C-A domain substitutions, the vector pTCA (Figure 5-8) was constructed (Section 2.8.2.2) in a similar manner to the C-A domain substitution plasmid created in Section 3.2. The plasmid pTCA contained two NotI restriction sites for the non-directional insertion of T-C-A domains; however, this did not matter as the correct orientation of each inserted T-C-A domain was confirmed during PCR screening and sequencing. Synthesis of these constructs proved to be technically very challenging, due to difficulties encountered in amplifying and cloning large PCR products. Nonetheless, three T-C-A domain substitutions were ultimately successfully created, sequence-verified and transformed into the *P. aeruginosa* PAO1 *pvdD* deletion strain. These T-C-A domains were amplified using primers in Table 2-1 from modules Ser-B, Ser-F and hfOrn-J with the T domains for each T-C-A domain amplicon being from the upstream module (Figure 5-9). Following the same nomenclature used for C-A domain substitutions, the resulting strains containing these variant *pvdD* constructs will be referred to as TCASer-B, TCASer-F and TCAhfOrn-J, respectively. In addition, the C-A domain substitution strain CAhfOrn-J was created, so that it could be compared to TCAhfOrn-J.

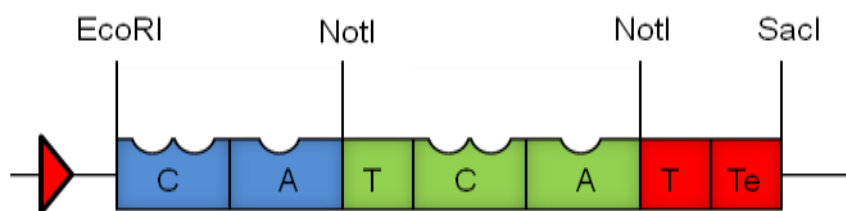


Figure 5-8. Domain organisation within the T-C-A domain substitution plasmid. Domains coloured blue and red were derived from the first and second module of *pvdD*, respectively. Domains in green were absent in the substitution plasmid and indicate where alternative domains would be added.

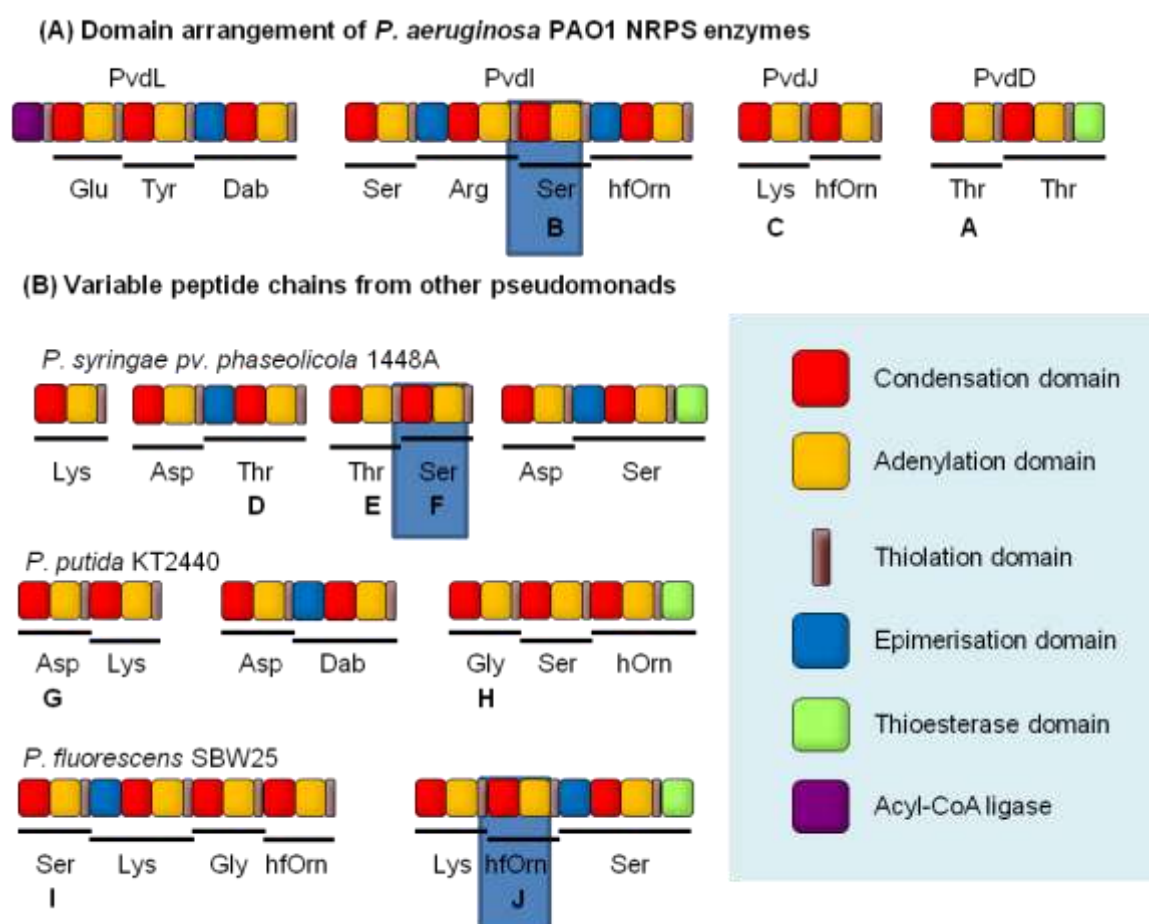


Figure 5-9. Location of domains used to source T-C-A domains for substitution. (A) Domain architecture of the four NRPS genes involved in *P. aeruginosa* PAO1 pyoverdine synthesis. (B) Domain architecture of genes involved in the synthesis of pyoverdine variable peptide chains from three other *Pseudomonas* strains. Figure modified from Figure 3-3.

### 5.5.1 Growth on solid media of T-C-A and C-A domain substitution strains

The CAThr-WT control, *pvdD* deletion strain, three T-C-A domain substitutions and the corresponding C-A domain substitution strains were spotted onto KB agar plates both with and without EDDHA (Figure 5-10). Out of the C-A domain substitution strains, CASer-F and CAhfOrn-J were fluorescent under UV and able to grow in the presence of EDDHA and CASer-B was not. The T-C-A domain substitution strains showed similar results to their corresponding C-A domain substitution strains in terms of fluorescence and growth in the presence of EDDHA. The fluorescence visible from the CAhfOrn-J and TCAhfOrn-J substitution strains suggested the successful creation of pyoverdine containing a terminal hfOrn residue. The substituted C domain from these two strains naturally receives a Lys residue at its donor site, supporting the previous conclusion (Section 4.5) that C domain side-chain specificity at the donor site is generally relaxed. The lack of any substantial difference between T-C-A and C-A domain substitutions suggested that retaining a native T/C interface might not provide a general means for improving the activity of NRPS domain recombination relative to the previously tested C-A domain substitution strategy.

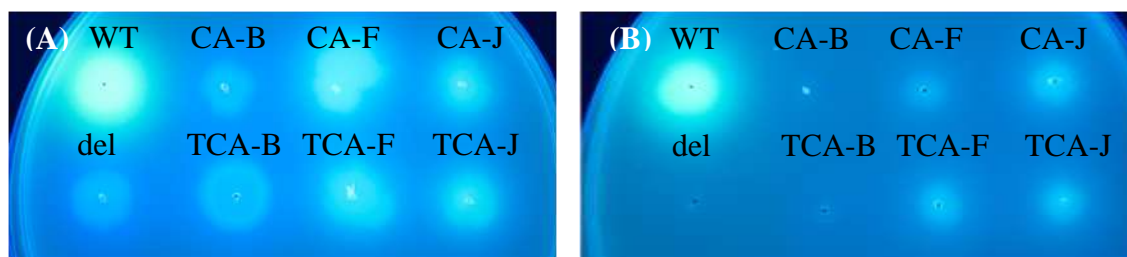


Figure 5-10. Growth of C-A versus T-C-A domain substitution strains on solid media, both (A) without and (B) with EDDHA. The substrate specified by each module was removed from nomenclature for clarity. Plates were inoculated with the restriction site control (WT), domain substitution strains, and the *pvdD* deletion mutant (del) then incubated for 24 h at 37 °C. Photographs were taken under UV light.

### 5.5.2 Growth in liquid media of T-C-A and C-A domain substitution strains

To more accurately quantify the levels of pyoverdine production, the absorbance (Figure 5-11A) and fluorescence (Figure 5-11B) of the T-C-A and C-A domain substitution strains were measured from the supernatant of 24 h cultures. For the domain substitutions with domains from modules Ser-B and hfOrn-J, there appeared to be little difference between T-C-A and C-A domain substitutions. On the other hand, slight increases in absorbance and fluorescence were detected for strain TCASer-F compared to CASer-F. Nonetheless, the differences were small and overall these results indicated that keeping the native T/C domain interactions intact had little effect for improving pyoverdine synthesis.

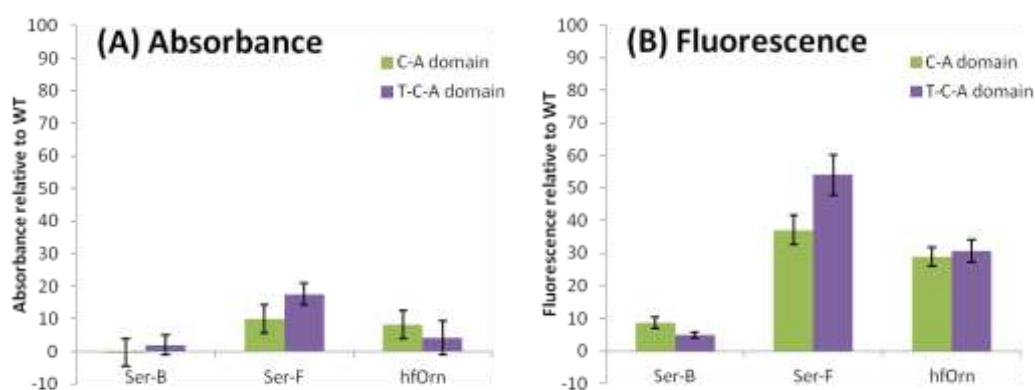


Figure 5-11. Pyoverdine production in liquid media from T-C-A versus C-A domain substitution strains. Values are expressed as a percentage relative to (A) the absorbance (400 nm) or (C) the fluorescence (ex. 400 nm/ em. 440 nm) measured for the CAThr-WT strain, having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation.

### 5.5.3 Mass spectrometry of T-C-A and C-A domain substitution strains

The T-C-A domain substitution strains and the corresponding C-A domain substitution strains were then analysed by mass spectrometry to confirm the production of the expected pyoverdines. Both TCASer-B and TCASer-F were found to be synthesising the same pyoverdine as their corresponding C-A domain substitutions, i.e. the truncated pyoverdine and full-length pyoverdine containing a terminal Ser residue, respectively. The strains TCAhfOrn-J and CAhfOrn-J both synthesised a product of approximately

1390.6 m/z, consistent with the terminal residue of pyoverdine being replaced with an hfOrn residue ( $C_{57}H_{88}N_{19}O_{22}$ ). As this pyoverdine was new and therefore not previously analysed, CID of this product was performed (Figure 5-12; Table 5-3). The five ions corresponding to the abundant ions analysed from other pyoverdines were consistent with the Thr residue being replaced by hfOrn (Figure 5-12) and full spectra were obtained confirming the remainder of pyoverdine (Table 5-3).

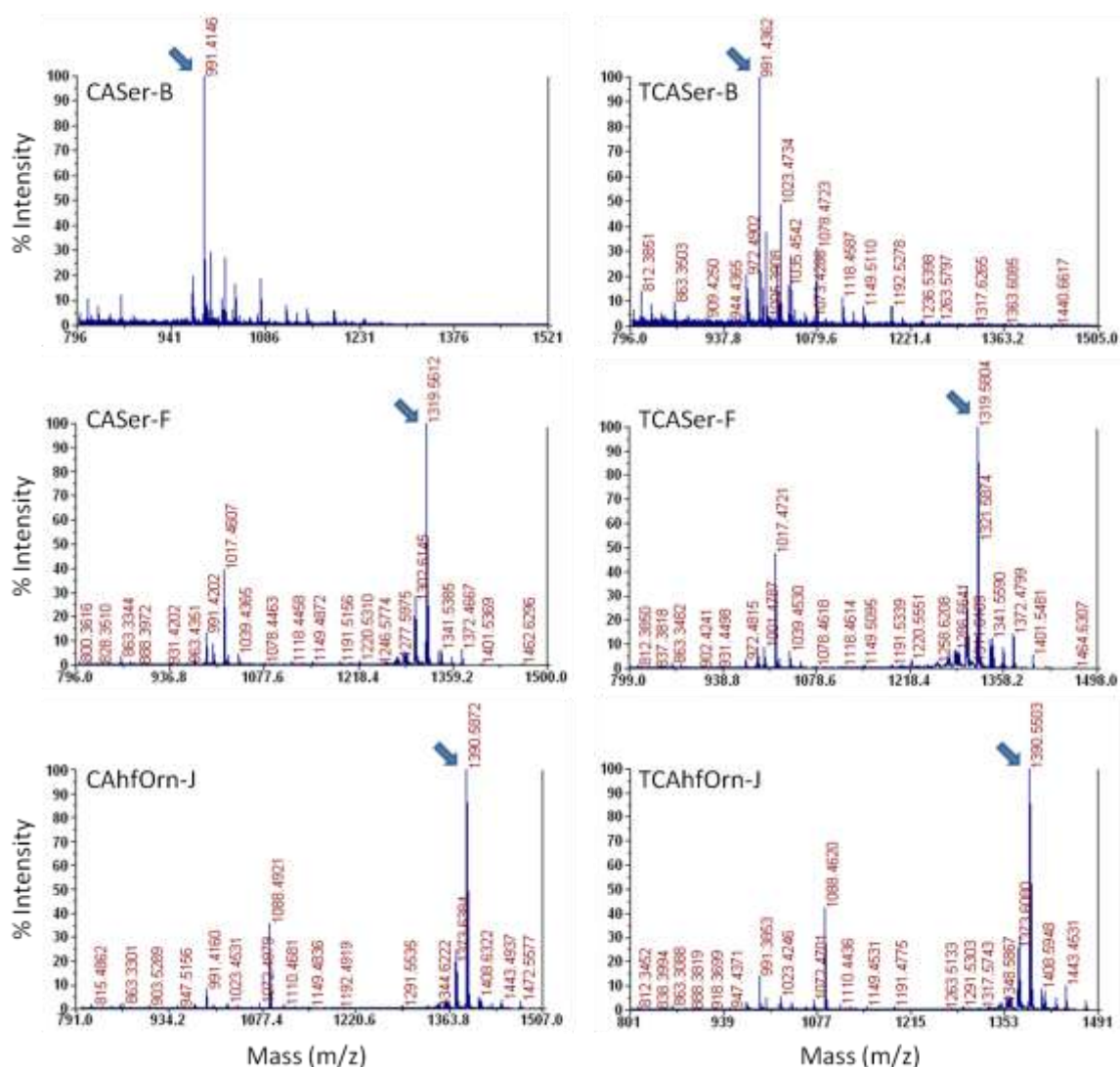


Figure 5-12. Mass spectra obtained from the supernatant of each of the T-C-A and C-A domain substitution strains tested in this section. Arrows indicate the most abundant peak corresponding to pyoverdine.

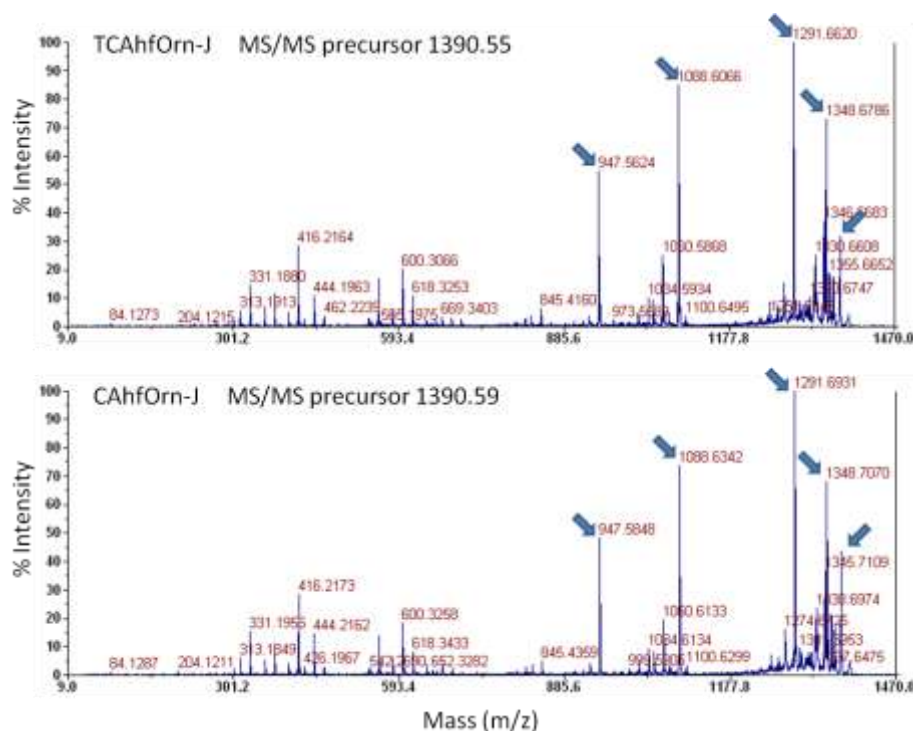


Figure 5-13. CID spectra of the pyoverdine synthesised by the strains TCAhfOrn-J and CAhfOrn-J. The five most abundant ions are indicated by arrows.

Table 5-3. Ions detected by CID of the pyoverdine synthesised by the strains TCAhfOrn-J and CAhfOrn-J.

	TCAhfOrn	CAhfOrn
Ion	m/z	
[RDA + H - Suca] <sup>+</sup>	204.12	204.12
A <sub>1</sub> - Suca	317.22	317.20
A <sub>1</sub> - NH <sub>3</sub>	399.20	399.22
A <sub>1</sub>	416.22	416.22
B <sub>1</sub>	444.20	444.22
B <sub>2</sub>	600.31	600.33
B <sub>3</sub>	687.34	687.31
B <sub>4</sub>	845.42	845.44
Y <sub>7</sub> ''	947.56	947.58
[RDAb + H] <sup>+</sup>	1088.61	1088.63
[M + H - Suca] <sup>+</sup>	1291.66	1291.69
[M + H - CH <sub>2</sub> N <sub>2</sub> ] <sup>+</sup>	1348.68	1348.71
[M + H - NH <sub>3</sub> ] <sup>+</sup>	1373.55	1373.60
MS/MS precursor	1390.55	1390.59

Shaded cells indicate ions expected to contain the residue added by the second module of PvdD. Suca, succinamide; RDA, chromophoric fragment from the RDA process; RDAb, peptide chain fragment released by the RDA process; ND, not detected.

## 5.6 Analysing the effect of disrupting the T-C junction via T domain substitutions

The T-C-A domain substitutions described in the previous section provided initial evidence that disruption of T/C interactions was not a primary factor in the loss of function observed for most C-A domain substitution constructs. However, both T-C-A and C-A domain substitutions were relatively difficult to construct because of low ligation efficiency of the relatively large insert into the vector. As an alternative means of investigating the effect of disrupting T-C domain interactions, it was decided to make T domain substitutions in the first module of *pvdD*. These T domain substitutions are referred to as T1 domain substitutions in this thesis. As substituting T<sub>C</sub> domains into this position would disrupt a previously optimised T/C domain interaction, a high success of substitutions of the T1 domain for alternative T<sub>C</sub> domains would be further evidence against the hypothesis that disrupting T/C domain interactions is a major factor in causing C-A domain substitutions to become non-functional. An added benefit of T1 domain substitutions was that, identical to the C-A domain substitution plasmid, the T1 domain substitution plasmid (pST1) had one fusion site added between an A and a T domain and a second between a T and a C domain (Figure 5-14A). Moreover, both T and C-A domain substitutions are highly similar in the T domain interactions changed, i.e. both T1 and C-A domain substitutions create new A/T domain, C-/T domain and T/C domain interactions (Figure 5-14B). Consequently, T1 domain substitutions mimic C-A domain substitutions in the fusion sites used and domain interactions altered. Thus, T1 domain substitutions not only test whether the T/C domain interaction is likely to be disrupted by C-A domain substitutions, but also whether any T domain interactions would be disrupted by using the fusion sites in this study.

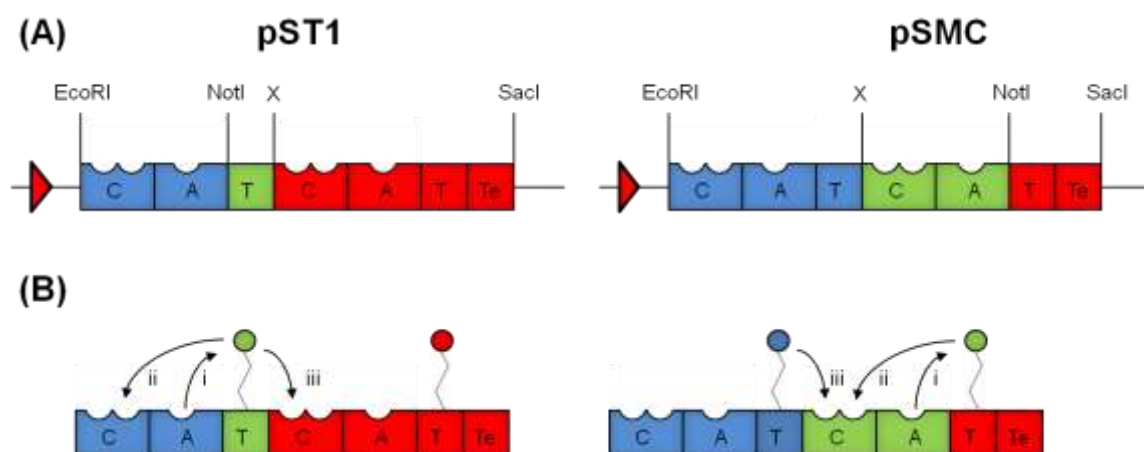


Figure 5-14. Comparison of T domain substitutions to C-A domain substitutions. (A) Location of restriction sites in the pST1 and pSMC plasmids. (B) T domain interactions altered by T domain substitutions compared with C-A domain substitutions. In both types of substitution, interactions between T domains and new domains are (i) an A domain passes a substrate to a new T domain, (ii) a T domain passes a substrate to a new upstream C domain and (iii) a T domain passes a substrate to a new downstream C domain.

### 5.6.1 Creation of T domain substitution variants

The primary aim of this section was to test the ability of substituted  $T_C$  domains to interact with their new C domain partners. As discussed above, it was expected that this would mimic C-A domain substitutions in terms of the location of the fusion sites and the T domain interactions that were thereby altered. Despite there being a number of studies that have previously focused on T domain interactions (Section 5.1), extensive literature searches identified no previous studies that had previously evaluated the ability of substituted T domains to interact with C domains. This may be because of greater difficulty in testing *in cis* interactions compared to *in trans* interactions (Zhou et al., 2006). The most similar studies to T domain substitution upstream to a C domain examined the *in trans* interactions between the ArCPs and the C domain of EntF (Lai et al., 2006; Zhou et al., 2007) or T domain substitutions that were conducted in the yersiniabactin synthetases upstream to a Cy domain (Suo, 2005). In addition, it has been suggested that keeping the type of T domain constant is important during recombination experiments (Baltz, 2012). With a secondary aim of providing greater understanding of these issues,  $T_{Ct}$ ,  $T_E$  and  $T_{TE}$  domains were also substituted into the first module of *pvdD*.

The T1 domain substitution plasmid, named pST1 and shown above in Figure 5-14A, was constructed as described in Section 2.8.2. Once created, eighteen T domains from

upstream of C, E or TE domains (Figure 5-15) as well as the T domain from the first module of *pvdD* were PCR amplified using primers from Table 2-1 and ligated into pST1 via NotI/NotI fusions at the 5' end and NheI/SpeI fusions at the 3' end. The eighteen T domain variants are referred to by the labels in Figure 5-15, in which T domains are labelled according to their type, e.g. C1-C6 refers to  $T_C$  domains, Ct1-Ct4 to  $T_{Ct}$  domains, etc.

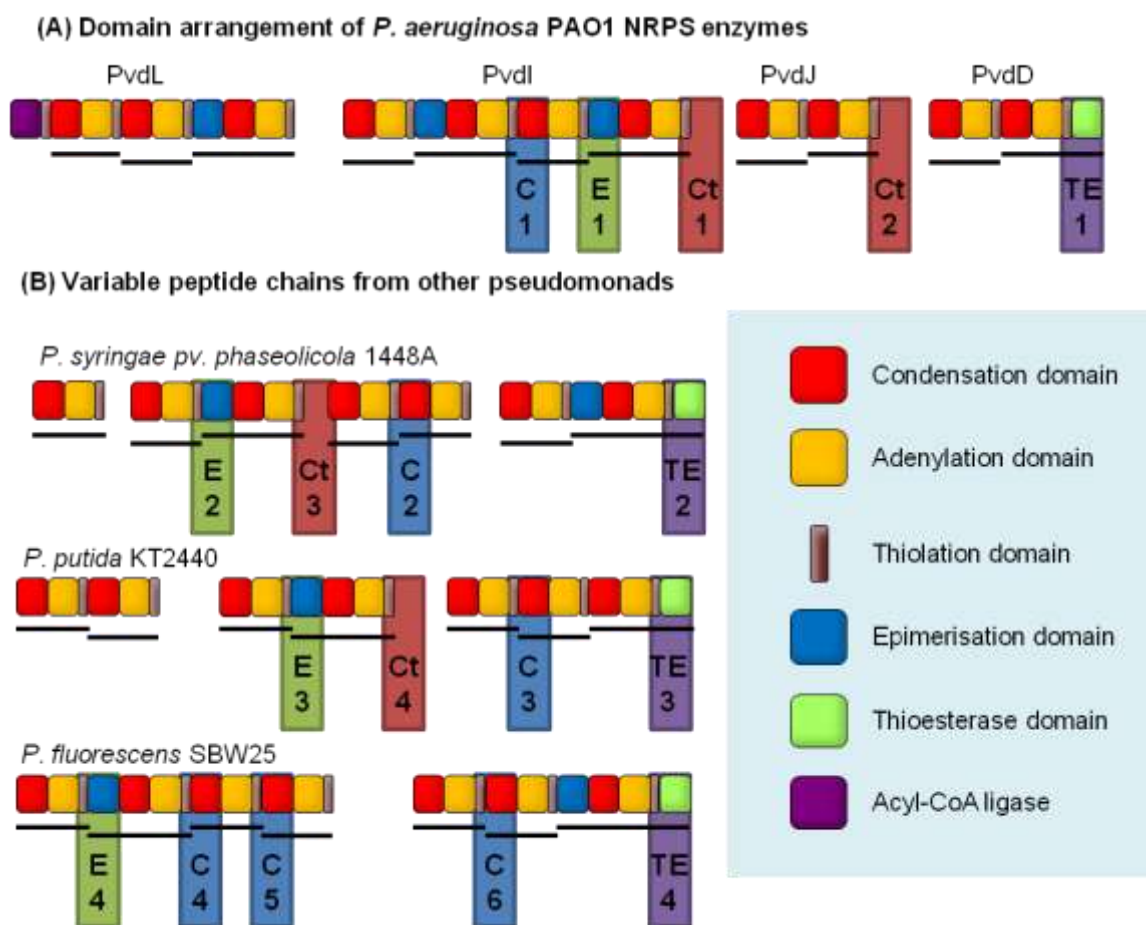


Figure 5-15. Location of T domains in *Pseudomonas* NRPS pathways used for T1 domain substitution. (A) Domain architecture of the four NRPS genes involved in *P. aeruginosa* PAO1 pyoverdine synthesis. (B) Domain architecture of genes involved in the synthesis of pyoverdine variable peptide chains from three other *Pseudomonas* strains. Figure modified from Figure 3-3.

## 5.6.2 Pyoverdine production by T domain substitution strains

### 5.6.2.1 Growth of T domain substitution strains on solid media

The T1 domain substitution variants were transformed into the *P. aeruginosa* PAO1 *pvdD* deletion strain and tested for pyoverdine production on solid media (Figure 5-16). It was found that all six T<sub>C</sub> domain variants were fluorescent at levels indistinguishable from the restriction site control and were all highly viable in the presence of EDDHA (Figure 5-16A). For the other types of substituted T domain (Figure 5-16, panels B - D), the strains generally showed similar fluorescence and growth in the presence of EDDHA as the restriction site control. The exceptions to this were the strains Ct2, Ct4 and TE3 that showed low levels fluorescence and low viability in the presence of EDDHA, and strains Ct3, E4 and TE1 that showed no increase in fluorescence and did not grow at all in the presence of EDDHA.

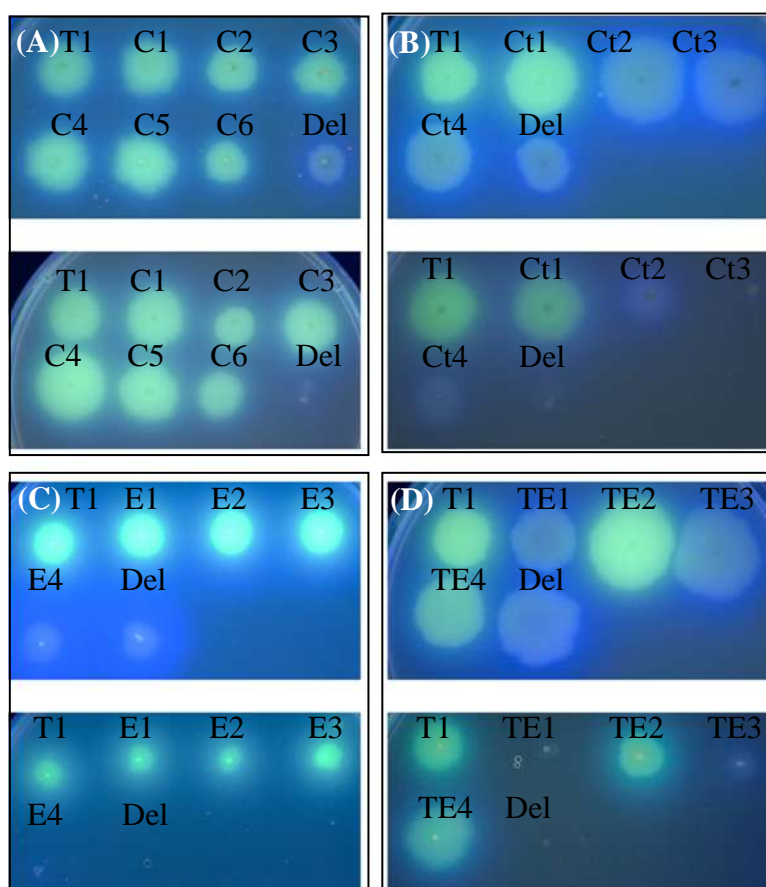


Figure 5-16. Growth of T1 domain substitution strains on solid media. Substitutions are grouped as (A) T<sub>C</sub> domain substitutions, (B) T<sub>Ct</sub> domain substitutions, (C) T<sub>E</sub> domain substitutions and (D) T<sub>TE</sub> domain substitutions. For each type of substitution, the top figure is for growth on KB media and the bottom figure is for growth on KB media supplemented with EDDHA. T1 refers to the restriction site control containing the native T<sub>C</sub> domain ligated back into pST1. Plates were incubated for 24 h at 37 °C, and photographs were taken under UV light.

#### 5.6.2.2 Growth in liquid media of T1 domain substitution strains

When grown in liquid media, the levels of absorbance of all the T1 domain substitution strains that were highly viable on EDDHA (Figure 5-16) were found to be similar to the T1 restriction site control, with the exception of slightly reduced absorbance observed for strain TE4 (Figure 5-17A). Strains Ct2, Ct4 and TE3, which all had low levels of viability in the presence of EDDHA, nonetheless all gave increased absorbance compared to the *pvdD* deletion control (at levels between 14 - 26 % of the restriction site control). The three strains not viable in the presence of EDDHA, namely strains Ct3, E4, and TE1, all showed levels of absorbance similar to the deletion control strain. For these strains, there appeared to be no increase in fluorescence (Figure 5-17B).

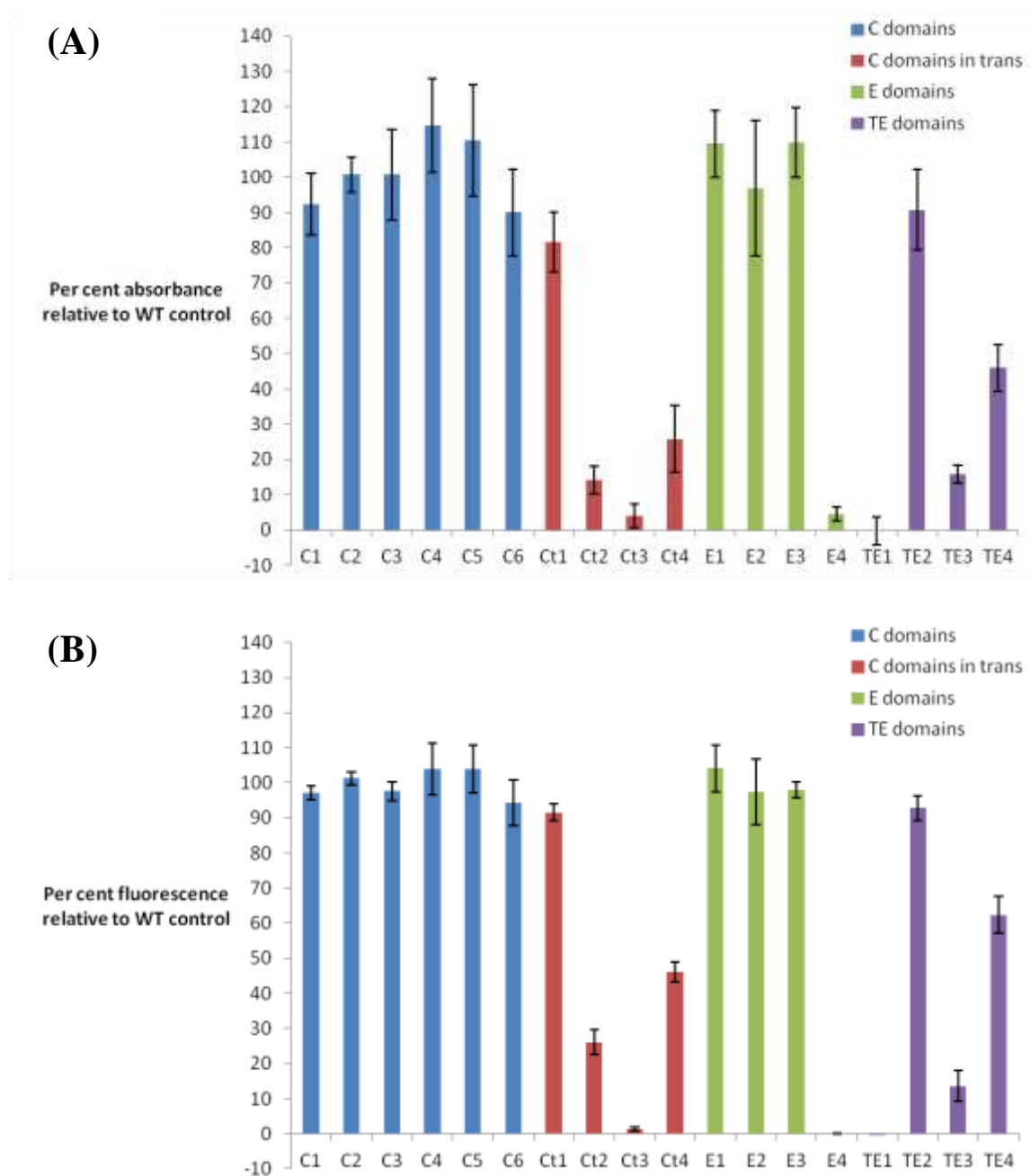


Figure 5-17. Pyoverdine production from T1 domain substitution strains grown in liquid media. Values are expressed as a percentage relative to the (A) absorbance (400 nm) or (B) fluorescence (ex. 400 nm/ em. 440 nm) measured for the Thr-WT strain, having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation.

In summary, these results indicate that T<sub>C</sub> domain substitution strains were all highly active in pyoverdine production. As T<sub>C</sub> domain substitutions disrupt the same T domain interactions as C-A domain substitutions, these results are inconsistent with the hypothesis that the non-functional C-A domain substitutions in Chapter 4 were inhibited by a disruption of the native T/C domain interface. For other T domain types, there appears to be some reduction in pyoverdine synthesis, however, only three appeared completely inhibited in pyoverdine synthesis. This indicates a relatively high portability for other types of T domain to function upstream of a C domain – in contrast with low T<sub>C</sub> domain portability upstream of the TE domain in the BpsA model (Section 5.1.2.1).

### 5.6.3 Mass spectrometry of T domain substitution strains

Next the T1 domain substitution strains with impaired function and one functional strain (namely strain C4) were analysed by mass spectrometry. Strain C4 was selected randomly to serve as a control as the functional T domain substitutions were most likely synthesising wild type pyoverdine due to wild type levels of activity and growth on EDDHA containing media. The strains Ct2, Ct4, TE3 and C4, that synthesised at least some pyoverdine as judged by absorbance, were all found to be synthesising wild type pyoverdine. Strain Ct3 was detected to synthesise the truncated 991.5 m/z product. In contrast, no pyoverdine production was detected from strains E4 and TE1. The absence of these products would suggest that peptide synthesis is stalling before reaching the new T domain or that the stalled peptide product is more stable on these T domains. Overall, these results indicate that only 3/18 T domain substitutions were completely inhibited in wild type pyoverdine production. The high success of T domain substitutions in wild type pyoverdine synthesis illustrated that T domains are highly portable in the context of *pvdD*.

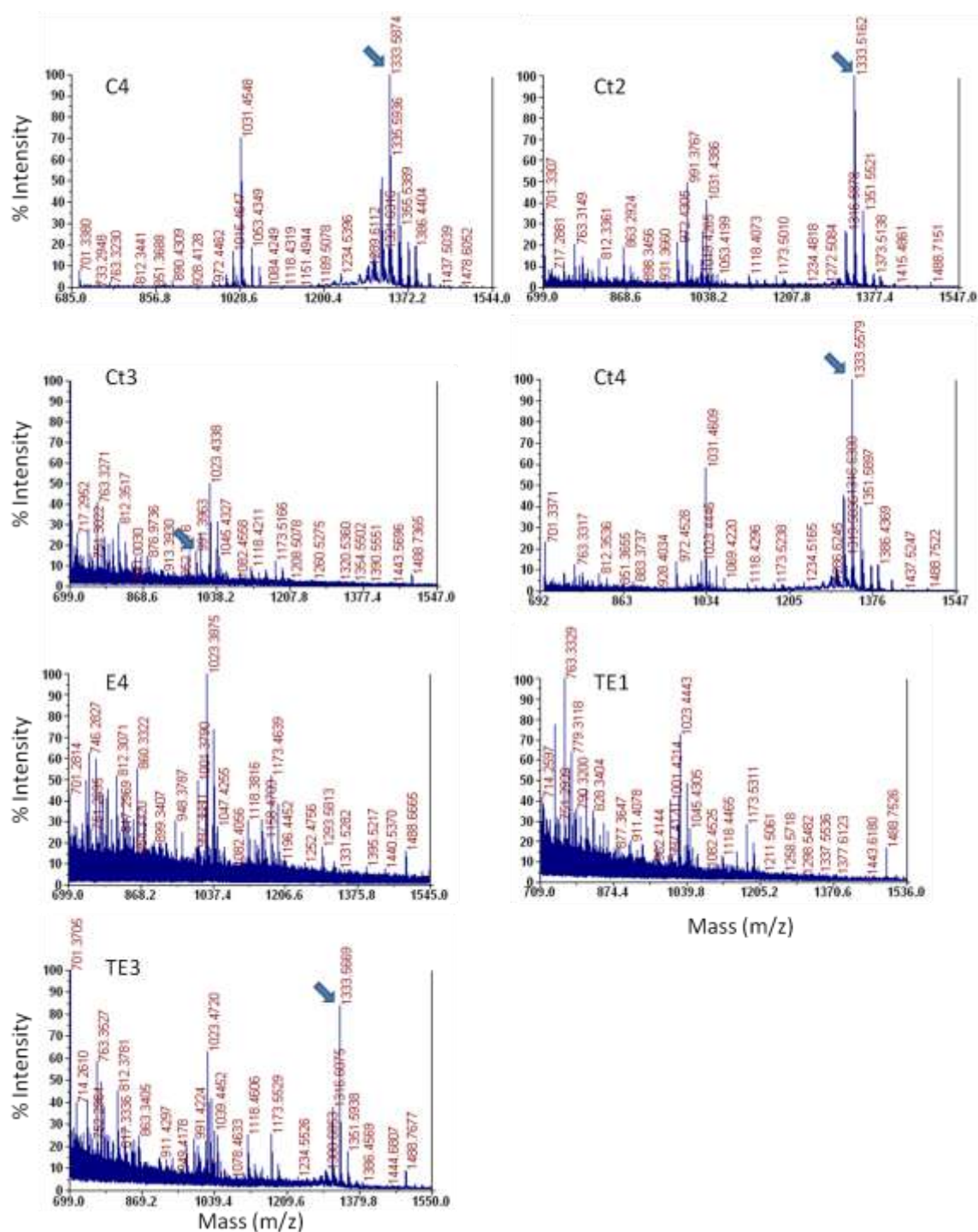


Figure 5-18. Mass spectra obtained from the supernatant of each of the T1 domain substitution strains. Arrows indicate peaks at 991.5 and 1333.6 m/z, corresponding to the truncated and wild type pyoverdine, respectively.

#### 5.6.4 Amino acid sequence alignments of T domains

The sequences of T domains were aligned in an attempt to highlight any key differences between functional T domains and non-functional T domain substitutions. Based on

mutagenesis studies, positions at -8, +1, +4, +19, +23 and +24 relative to the conserved 4'-pp Ser attachment site have been implicated in T domain interaction with the surrounding domains (Lai et al., 2006; Owen, 2010; Zhou et al., 2006, 2007). In general, residues that were present at these locations in strains that exhibited less than 80 % of wild type levels of pyoverdine synthesis could also be found at the equivalent location in the T domain from at least one highly active strain. However, some exceptions to this were seen at the -8, +19, 23 and +24 positions. At the -8 position, the strain Ct4 contained an Asp residue. Interestingly, this position was more variable across the functional strains. The +19 position was highly variable for all T domains. The T domains from Ct3 and TE3 contained Gly and Arg residues at this position, respectively, which were not observed in any of the functional T domain constructs. At the +23 position, the T domain from Ct2 contained a Met residue. All T domains from other strains contained a Leu residue at this position except for the functional strains C3 and TE2, which contained a Val and Phe residue, respectively. At the +24 position all strains except the T domains upstream to TE domains contained a Phe residue. The impaired T<sub>TE</sub> domain substitution strains contained either a Met or Ile residue at this position, whereas the functional strain contained a Leu residue. The differences between functional T domains and non-functional domains at these positions could be involved in the loss of function. Ultimately however, no clear patterns emerged from the sequence alignments.

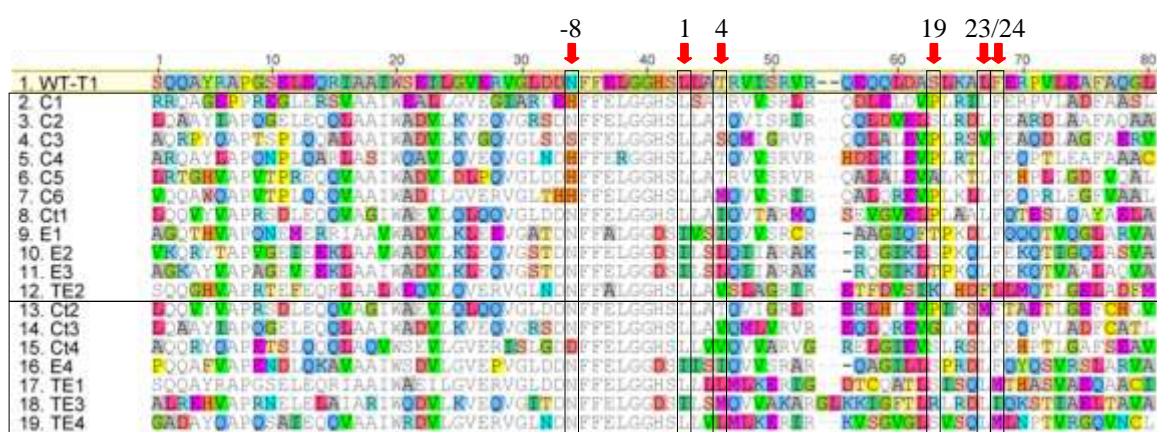


Figure 5-19. Alignment of T domains substituted into the first module of *pvdD*. T domains that allowed high levels of pyoverdine production are numbered 2-12 in left most column and those that cause impaired pyoverdine production are numbered 13-19. WT-T1 (1) refers to the native T<sub>C</sub> domain from *pvdD*. Arrows indicate residues previously implicated in interaction with other NRPS domains. Alignment generated in Geneious version 5.5.7 created by Biomatters. Available from <http://www.geneious.com/>.

## 5.7 Discussion

The first part of this chapter tested whether the T<sub>C</sub> domain from the first module of *pvdD* could function when substituted to act as the second T domain in *pvdD*; and whether its ability to function could be improved by directed evolution performed in the *bpsA* environment to improve communication with the BpsA TE domain. This work showed the domain substitution of the native T<sub>C</sub> domain into the T2 position severely impaired pyoverdine production, whereas evolved T<sub>C</sub> domains that had gained activity in the *bpsA* context had also gained activity in the *pvdD* T2 position. However, there was not a direct correlation between activity in BpsA and PvdD, with all evolved T<sub>C</sub> domains having similar levels of activity in PvdD. This suggested that although there may be some generality in terms of residues that are important for T/TE domain interactions, there may also be unique differences for each T-TE combination.

To examine how widespread the importance of the +4 and +24 positions might be, sequence alignments were generated of T domains from outside the pseudomonads. This found that the +4 and +24 positions exhibited no absolute differences between T<sub>C</sub> and T<sub>TE</sub> domains. Interestingly, the T<sub>C</sub> and T<sub>TE</sub> domains had several differences to T<sub>E</sub> domains. This supports some type of specificity for T<sub>E</sub> domains as observed when A and A-T domain substitutions were made upstream to the E domain of TycA (Linne et al., 2001). This difference could perhaps be due to T<sub>E</sub> domains interacting with multiple downstream domains, i.e. interacting with both a downstream E and C domain.

The lack of clear sequence differences between T<sub>C</sub> and T<sub>TE</sub> domains is interesting, as it has been shown in several systems that T/TE interactions can be inhibited by domain substitution (Doekel et al., 2008; Duerfahrt et al., 2003; Mootz et al., 2000; Owen, 2010). This raised the question of whether T/C domain interactions, which are altered by C-A domain substitution, are also susceptible to disruption. The importance of the T/C domain interaction was first examined by T-C-A domain substitutions, which sought to keep the native T/C domain interaction intact. Although the sample size was small, the three T-C-A domain substitutions all had highly similar levels of pyoverdine synthesis compared to the corresponding C-A domain substitutions. This indicated that disrupting T/C domain interactions was not causing C-A domain substitutions to lose activity. However, the examination of T-C-A domain substitutions was limited by the difficulty in amplifying and ligating larger DNA fragments. Moreover, a scarcity of C-A domain

substitutions with an upstream T<sub>C</sub> domain meant every additional T-C-A domain substitution required the corresponding C-A domain substitution to be constructed.

As an alternative means of testing whether C-A domain substitutions might inhibit T/C domain interactions, T<sub>C</sub> domain substitutions were made in the first module of *pvdD*. As discussed in Section 5.6, T<sub>C</sub> domain substitutions alter the native T/C domain interaction and a high success rate of T<sub>C</sub> domain substitution at this position would show that the T/C domain interaction is not inhibited by domain substitution. All six T<sub>C</sub> domain substitution strains were highly functional in pyoverdine synthesis at levels indistinguishable from the restriction site control strain. This strongly indicated that not only is the T/C interface tolerant of change, but also that T domain interactions with other domains would unlikely be disrupted by C-A domain substitution.

To test the portability of other different types of T domain, T<sub>Ct</sub>, T<sub>E</sub> and T<sub>TE</sub> domain substitutions were also made into the first module of *pvdD*. These domain substitutions had a high success rate with only one out of four of each type of T domain having wild type pyoverdine production entirely abolished. That three out of four T<sub>Ct</sub> domain substitutions showed reduced pyoverdine production is of interest, as they only differ from T<sub>C</sub> domains in that the interaction with the downstream C domain is *in trans*. A potential cause for this difference is that T<sub>Ct</sub> domains may require a more specific T/C domain interaction to ensure association with the correct downstream NRPS in the multienzyme assembly line (Chiocchini et al., 2006; Koglin and Walsh, 2009).

When the sequences of the non-functional substituted T domains were aligned against the functional T domain sequences, there appeared to be no clear cut pattern to the residues identified in other studies as being important for inter-domain communication. This fits with the hypothesis that residues vital for correct T domain interaction are likely to differ between T domains. It was decided to not attempt to identify the mutations that would restore activity to these T1 domain substitutions because the overall high success rate of T1 domain substitutions strongly suggests that T domain interactions, in the context of an elongation module, are highly tolerant of change. Despite this tolerance, the variability of T<sub>Ct</sub>, T<sub>E</sub> and T<sub>TE</sub> domain substitutions supports the rule that the type of domain downstream to a T domain should ideally be kept constant to maximise the likelihood of successful domain substitution.

To summarise, the collective results of all domain substitution experiments in this thesis indicate that a major limiting factor for A domain substitutions is the C domain acceptor site specificity. This suggests that the loss of activity of the A domain and/or ability of the A domain to communicate with the downstream T domain is unlikely to be the primary cause for loss in function of C-A domain substitutions. The implications of this chapter are that C-A domain substitutions in PvdD are unlikely to inhibit T domain interactions. This in turn suggests that the frequent loss in activity observed with C-A domain substitutions is due to a factor inherent to the C domain rather than any surrounding domains. One hypothesis discussed earlier (Section 4.5) that fits with these results is C-A domain substitutions may be limited by steric constraints at the donor site.



## Chapter 6: **Investigation of the acceptor site substrate specificity of a C domain**

### 6.1 **Introduction**

The ability to functionally substitute Thr specifying A domains (Chapter 3) and any T<sub>C</sub> domain (Chapter 5) into PvdD suggests that the large loss of function observed for C-A domain substitutions (Chapter 4) was not a consequence of introduction of the new A domain or of disrupting T domain interactions. This in turn suggests that the introduction of new C domains as part of C-A domain substitutions may have been the cause for the loss of function. It was hypothesised that this loss of function may have been caused by some C domains being unable to receive a non-cognate donor peptide due to steric constraints. This would also explain why the previous C domain donor site specificity study using single amino acids found no side-chain donor site specificity (Belshaw et al., 1999) and yet there was side-chain specificity identified for some dipeptides (Stein et al., 2005). Similarly, for donor site stereo specificity, Clugston et al. (2003) observed C domain stereo specificity towards a tetrapeptide, but not towards the single amino acid substrate. The studies by Stein et al. (2005) and Clugston et al. (2003) indicate that the size and composition of the incoming donor peptide can influence activity. Thus, both A and C-A domain substitutions may be limited by the C domain – the former by acceptor site selectivity, and the latter by donor site constraints. Overall, a greater understanding of C domain specificity is needed as it may aid in future efforts to perform functional NRPS domain substitutions.

Despite the potential importance of C domain specificity, little is known about the residues involved. Based on solved structures, the native substrates were modeled into the C domain active site of VibH (Keating et al., 2002) and the first C domain involved in CDA synthesis (Bloudoff et al., 2013). While it might be expected that this would provide reasonable estimates of the residues involved in determining specificity, confirming the exact nature of substrate docking would require the structure to be solved in complex with the substrate, and this has not yet been achieved (Bloudoff et al., 2013).

Another potential way to identify specificity conferring residues is via sequence analysis. However, from phylogenetic studies it was found that C domains align according to the type of donor substrate (e.g. fatty acid, D-amino acid or L-amino acid) rather than by the

side-chain of either the donor or acceptor substrates, and consequently no specificity conferring residues have been definitively identified on this basis (von Döhren et al., 1999; Rausch et al., 2007; Roongsawang et al., 2005).

This chapter aimed to identify, or at least narrow down, the location of residues involved in acceptor site specificity of the C domain using mutagenesis. Previous C domain mutagenesis studies have focused on the identification of catalytic residues rather than those involved in substrate specificity (Bergendahl et al., 2002; Keating et al., 2002; Roche and Walsh, 2003; Stachelhaus et al., 1998). In this study, the reason for choosing to examine specificity at the acceptor site instead of the donor site was because the tight amino acid side-chain specificity inferred at the C domain acceptor site (Sections 1.2.3.2) suggests the specificity conferring residues are more defined than the residues that might be generating steric constraints at the donor site. Relaxation of acceptor site specificity also builds on the A domain substitution work in Chapter 3, where it was concluded that altering or relaxing C domain acceptor site specificity might allow non-Thr specifying A domain substitutions to successfully synthesise new peptides (Section 3.5). A limitation of changing acceptor site specificity is that it might in turn generate problems with steric constraints at the donor site of the downstream C domain. However, as A domain substitutions alter only one residue of the donor peptide – in contrast with C-A domain substitutions, which potentially change the entire peptide received by the substituted C domain – the A domain substitutions seem less likely to induce steric limitations. Moreover, this study focused on the terminal module of PvdD, meaning that donor site specificity of a downstream C domain would not be a problem.

As a model system to examine the substrate specificity of C domains, it was chosen to focus on the first C domain of PvdJ, and to try and identify residues that could be altered to change the acceptor site specificity from L-Lys to L-Thr. In this chapter, the first C domain from *pvdJ* is referred to as JC<sub>Lys</sub> and the second C domain from *pvdD* as DC<sub>Thr</sub>. The J and D in the nomenclature JC<sub>Lys</sub> and DC<sub>Thr</sub> refer to the domains being sourced from *pvdJ* and *pvdD*, respectively. It was hypothesised that changing the key acceptor site residues of JC<sub>Lys</sub> to those from DC<sub>Thr</sub> would alter the acceptor site specificity of JC<sub>Lys</sub> to Thr. JC<sub>Lys</sub> was selected based on this C domain having previously been shown to function in PvdD as a C-A domain substitution (Section 4.3), proving that this C domain is able to receive the pyoverdine intermediate from the first module of PvdD. Also, JC<sub>Lys</sub> has a relatively high amino acid sequence identity to DC<sub>Thr</sub> (73.6 % identity). The high

sequence identity is useful for differentiating the residues likely to be involved specificity towards Lys over Thr.

## 6.2 Prediction of the residues most likely to be involved in substrate specificity

### 6.2.1 Sequence comparison of JC<sub>Lys</sub> and DC<sub>Thr</sub>

This section aimed to narrow down the residues likely to be involved in C domain acceptor substrate specificity. Despite the high sequence similarity between JC<sub>Lys</sub> and DC<sub>Thr</sub>, there were still too many differences across a large stretch of DNA for precisely targeted techniques to offer much hope of altering the substrate specificity of JC<sub>Lys</sub>. When the two C domains were aligned, it was observed that there were three regions of high sequence diversity interspersed by regions of high sequence identity (Figure 6-1). These regions of high sequence diversity are labelled 1 to 3 in Figure 6-1 and will be referred to as variable regions 1, 2 and 3. To guide the prediction for which of these regions were likely to be important for Thr specificity, a homology model of DC<sub>Thr</sub> was created to identify residues within 15 Å of the second His residue of the HHxxxDG motif proposed to be involved in catalysis (as described in Section 1.2.3.2). This distance was chosen based on a meta-analysis by Morley and Kazlauskas (2005) of multiple random mutagenesis studies that found changing residues within 15 Å of the catalytic residues was most likely to alter the substrate specificity of enzymes.

### 6.2.1 Structural comparison of variable regions of JC<sub>Lys</sub> and DC<sub>Thr</sub>

Homology modelling is based on the core tertiary structure of proteins being conserved to a higher level than amino acid sequence. This is shown by proteins that have sequence identity of over 50 % having few structural differences within the core of the protein (Lesk and Chothia, 1986). As such, the main structural differences between similar proteins are generally found in regions on the end of secondary structure, loops between secondary structure and peripheral secondary structure elements (Lesk and Chothia, 1986). It has been suggested, as a general guideline, that when the sequence identity is greater than 25 - 30 % it is likely a homology model of sufficient quality for identifying residues for site-directed mutagenesis can be created (Cavasotto and Phatak, 2009; Hillisch et al., 2004).

JC <sub>Lys</sub>	1	----FARLPQPQTRQEMDNPLPSYAQERQWFLWQLEPESAYHIPTALRL	
DC <sub>Thr</sub>	1	TTDAVSTIPLADRQQP---LALSFAQERQWFLWQLEPESAAYHIPALRL	
JC <sub>Lys</sub>	47	RGRLD IASLQRSFAALVERHESLRTRIARMGDEWVQVVSADVSLALEVEV	1
DC <sub>Thr</sub>	48	RGRLD VDALQRSFDSLVARHETLRTFRLEGGRSYQQVQPAVSVSIERE-	
JC <sub>Lys</sub>	97	Q RGLDEQRLLERVEAEIARPFDLQGP LLRVTLLEVDAD EHV LVMVQH HI	
DC <sub>Thr</sub>	97	QFG--EEGLIERIQAI VVQPFDLERG P LLRVNLLQLAEDDHV LVLVQH HI	
JC <sub>Lys</sub>	147	VSDGWSMQLMVEELVQLYAAYSQGLDVVLPALPI <b><u>QYADY</u></b> ALWQRSWMEAG	
DC <sub>Thr</sub>	145	VSDGWSMQVMVEELVQLYAAYSQGLDVVLPALPI <b><u>QYADY</u></b> ALWQRSWMEAG	
JC <sub>Lys</sub>	197	EKERQLAYWTGLLGGEQPV IELPLDHP RQPLRSYRGAQLDLELEPHLALA	2
DC <sub>Thr</sub>	195	EKERQLAYWTGLLGGEQPV LELPFDRPRPARQSHRGAQLGFELSRELVEA	
JC <sub>Lys</sub>	247	LKQLVQRKGVTFM LLLASFQALLHRYSGQADIRVGVP IANRNRVETERL	
DC <sub>Thr</sub>	245	VRALAQREGASSFM LLLASFQALLYRYSGQADIRVGVP IANRNRVETERL	
JC <sub>Lys</sub>	297	IGFFVNTQVLKADINGRMGFDELLAQARQRALEAQA H <b><u>QDL PFE</u></b> QLVEALQ	
DC <sub>Thr</sub>	295	IGFFVNTQVLKADLDGRMGFDELLAQARQRALEAQA H <b><u>QDL PFE</u></b> QLVEALQ	
JC <sub>Lys</sub>	347	PER SLGHNPLFQVMFNHQADSR SANQGVQLPGLSLERM EWRS SSVAFDLT	3
DC <sub>Thr</sub>	345	PER NASHNPLFQVLFN HQSEIRSVTPEVQLEDLRLEGLAWD GQT AQFDLT	
JC <sub>Lys</sub>	397	LDVHEAEDGIWASFGYATDLFEASTVERLARHWQNLLRGIVAEPGRPVAE	
DC <sub>Thr</sub>	395	LDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQRLGE	
JC <sub>Lys</sub>	447	LPLLLD	
DC <sub>Thr</sub>	445	LP-LLD	

Figure 6-1. Nucleotide sequence alignment of JC<sub>Lys</sub> and DC<sub>Thr</sub>. The three highly variable regions are highlighted in yellow and indicated by 1-3 in the right hand column. The C domains were split into three regions based on these variable sections with the recombination site for each region indicated in bold and underlined. Alignment performed according to Section 2.10.3.

#### 6.2.1.1 Creation and selection of a structural model of DC<sub>Thr</sub>

Using the template identification tool from the Swiss-Model server (<http://swissmodel.expasy.org/>), the crystal structure of the T-C domains spanning modules 5 and 6 of TycC involved in tyrocidine synthesis (pdb: 2JGP) (Samel et al., 2007) was judged to be the best template for homology modelling based on 38 % amino acid sequence identity with DC<sub>Thr</sub>. In addition to the 2JGP structure, it was decided to

create homology models using the C domain structures from VibH (pdb: 1L5A) (Keating et al., 2002) and SrfA-C (pdb: 2VSQ) (Tanovic et al., 2008) that had 17 % and 23 % sequence identity, respectively, as well as an E domain structure from TycA (pdb: 2XHG) (to be published; Samel, Heine, Marahiel, Essen) with 19 % sequence identity. Multiple structures were used to create homology models, because when sequence identity between the template and model drops below 40 %, using multiple template modelling along with selection of good models has been shown to generally yield better quality models (Larsson et al., 2008). In addition to using Modeller 9.11, homology models were created using the semi-automated Swiss-Model servers and several fully-automated servers (Section 2.10.4). The aim of this was to create a diverse set of homology models from which the best model could be selected.

The quality of the top models from each method (52 in total) was then estimated using QMEAN6 and QMEANclust. The statistical potential QMEAN6 was used as it provides a measure of the “degree of nativeness” of models, in other words, whether the model is characteristic of experimental structures. It is based on calculating statistical potentials from 6 variables within the model, which is then normalised to high quality experimental structures giving a Z-score (Benkert et al., 2008, 2011). QMEAN6 has been shown to perform well at predicting model quality when compared to either other statistical potentials (Benkert et al., 2009b; Rykunov et al., 2009) or other quality estimation methods (Benkert et al., 2008). In contrast, QMEANclust was developed to select the most accurate model from a set of models (Benkert et al., 2009b). It incorporates the information from QMEAN6 together with the structural density of local structures from many models. In this context, structural density refers to how similar a structure is to the other structures in a group. The reason for focusing on structural density is that local structures that are maintained in a similar conformation in a diverse set of models are more likely to be correct (Benkert et al., 2009b; Ginalska et al., 2003; Shortle et al., 1998). The evaluation of QMEANclust in CASP8 and CASP9, a series of large-scale bioinformatic experiments testing the accuracy of structural model prediction tools, found that QMEANclust was among the top model quality predictors in terms of predicting the global accuracy of models (Cozzetto et al., 2009; Kryshchuk et al., 2011) and stood out as the best at predicting accuracy at the residue level during CASP8 (Cozzetto et al., 2009).

From the 52 models analysed by QMEAN6 and QMEANclust (0), the model RaptorXmsa was selected to work with as it scored well in both measurements. QMEANclust found that most areas of RaptorXmsa were predicted to have high accuracy at the residue level and that the areas of predicted inaccuracy were located in loops, the ends of secondary structure and one peripheral helix (Figure 6-2A). The model RaptorXmsa aligned well to the C domain structure from TycC (Figure 6-2B) with a root-mean-square deviation for the backbone  $\alpha$  carbons of 0.381 Å. Overall, these results suggest that the core residues of the model are likely to be of high enough accuracy for predicting the residues in close proximity to the conserved catalytic His residue.

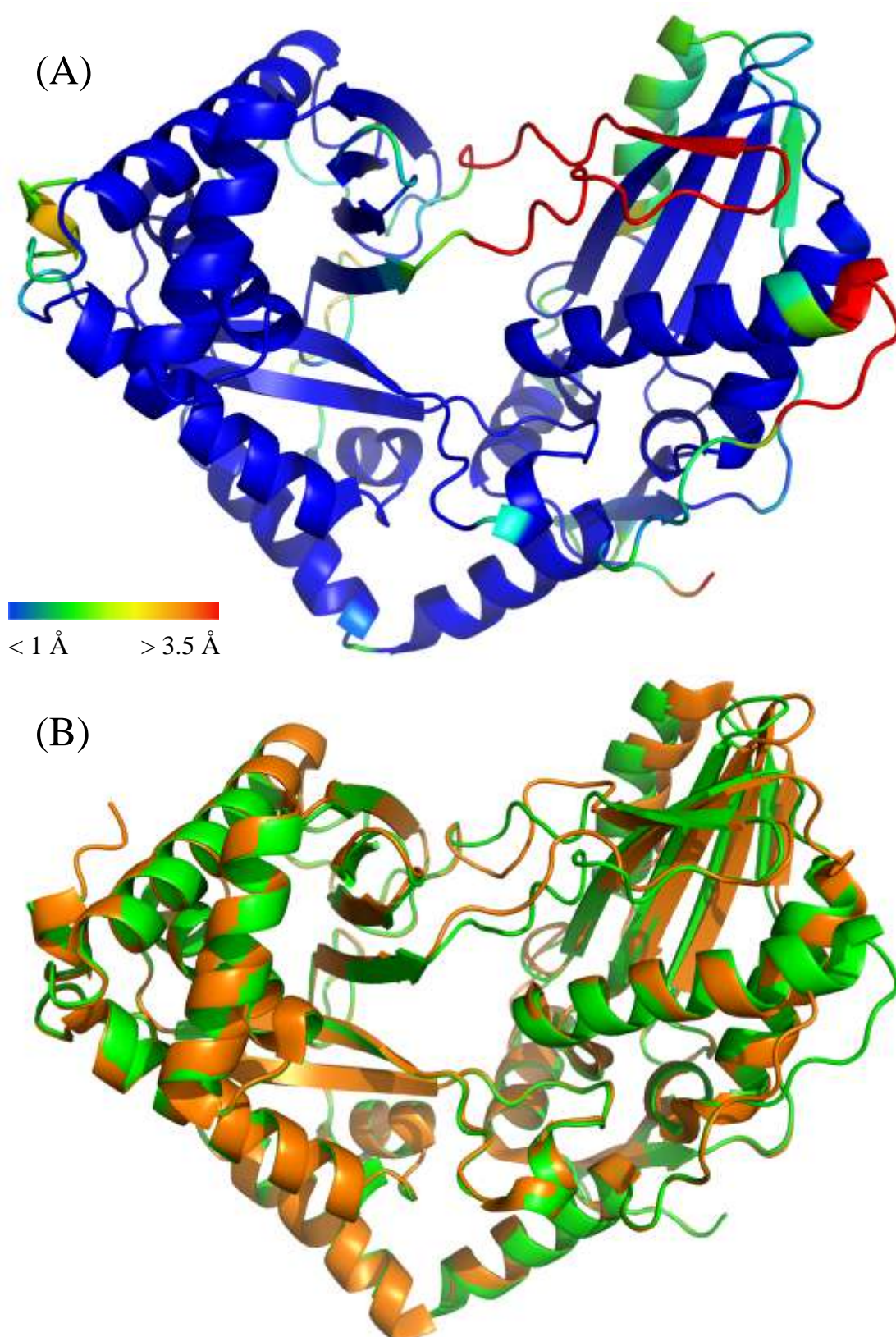


Figure 6-2. Analysis of the structural model RaptorXmsa. (A) The predicted error at the residue level based on QMEANclust. The backbone is coloured according to the level of predicted error measured in Å. (B) Alignment of RaptorXmsa (green) to the C domain from the structure 2JGP (orange).

### 6.2.1.2 Analysis of non-identical residues between JC<sub>Lys</sub> and DC<sub>Thr</sub>

The residues that were non-identical in the alignment of JC<sub>Lys</sub> and DC<sub>Thr</sub> were mapped onto the model of DC<sub>Thr</sub> and coloured according to the three variable regions of the C domain (Figure 6-3). These residues were deemed to be of particular interest as they were most likely to be involved in differentiation between Thr and Lys. The non-identical residues from the first variable region (Figure 6-3, red) were entirely located within one subdomain of the C domain and those from the second variable region (Figure 6-3, blue) were located within the other subdomain. The non-identical residues of the third variable region (Figure 6-3, green) were located in the same subdomain as the second variable region, with the addition of a large loop predicted to flex over the catalytic channel to the subdomain containing the first variable region.

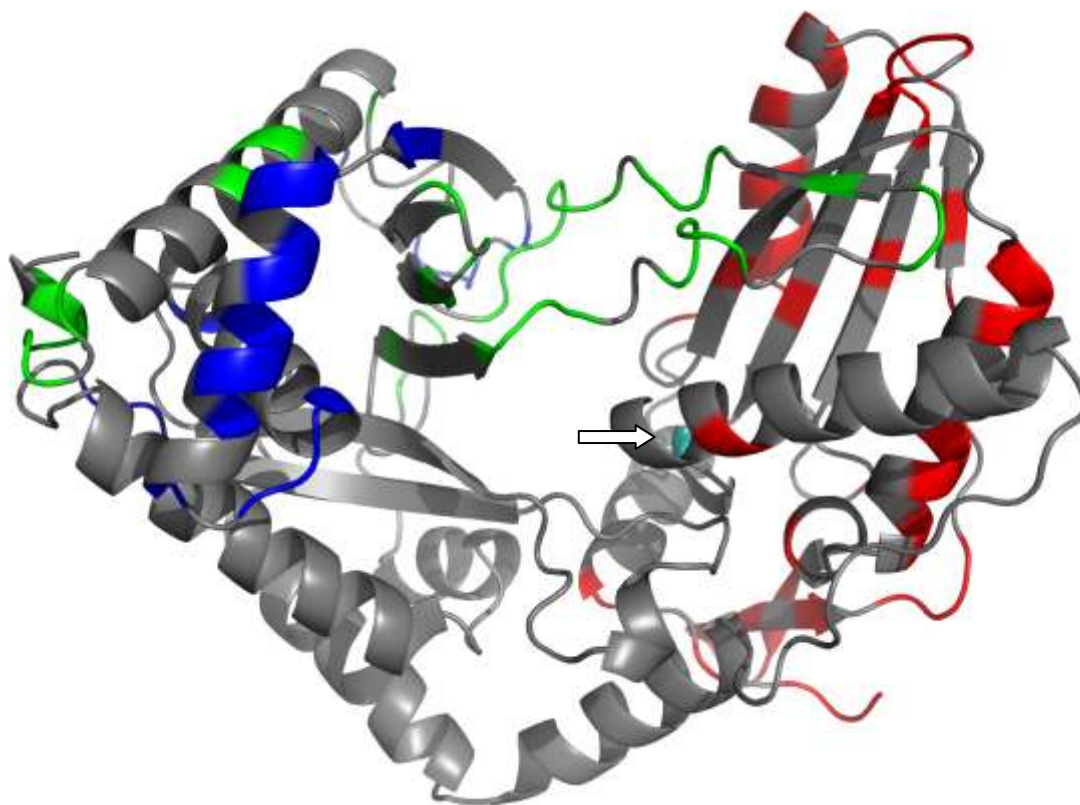


Figure 6-3. Residues differing between JC<sub>Lys</sub> and DC<sub>Thr</sub> mapped onto the model of DC<sub>Thr</sub>. Non-identical residues within variable regions 1, 2 and 3 are coloured red, blue and green, respectively. Non-identical residues and variable regions were based on sequence alignment shown in Figure 6-1. The catalytic His residue is indicated by an arrow and coloured in cyan.

### 6.2.1.3 Analysis of residues within 15 Å of the conserved catalytic His residue

Next the residues within 15 Å of the catalytic His side-chain were examined (Figure 6-4). This showed that all three variable regions had non-identical residues close to the catalytic His residue. The non-identical residues of the first variable region within 15 Å (Figure 6-4, red) were thought unlikely to be involved in substrate specificity because they did not have a direct line of sight to the catalytic residue. From the second variable region, there was only one non-identical residue within 15 Å of the catalytic His (Figure 6-4, blue). The third variable region had six non-identical residues within 15 Å (Figure 6-4, green), all of which had a direct line to the catalytic His. These observations suggested that the third variable region was the most likely region to be involved in substrate specificity. Four out of six of the non-identical residues from the third variable region were within the flexible loop covering the catalytic channel. This loop interacts through hydrogen bonding with the other subdomain in all C domain crystal structures (Bloudoff et al., 2013; Keating et al., 2002; Samel et al., 2007; Tanovic et al., 2008). Moreover, the surface area of this loop in contact with the other subdomain suggests that it may remain intact during catalysis, though it may need to open to let the peptide chain through (Bloudoff et al., 2013). One way the loop could be involved in determining substrate specificity is if the correct acceptor substrate is required for it to open.

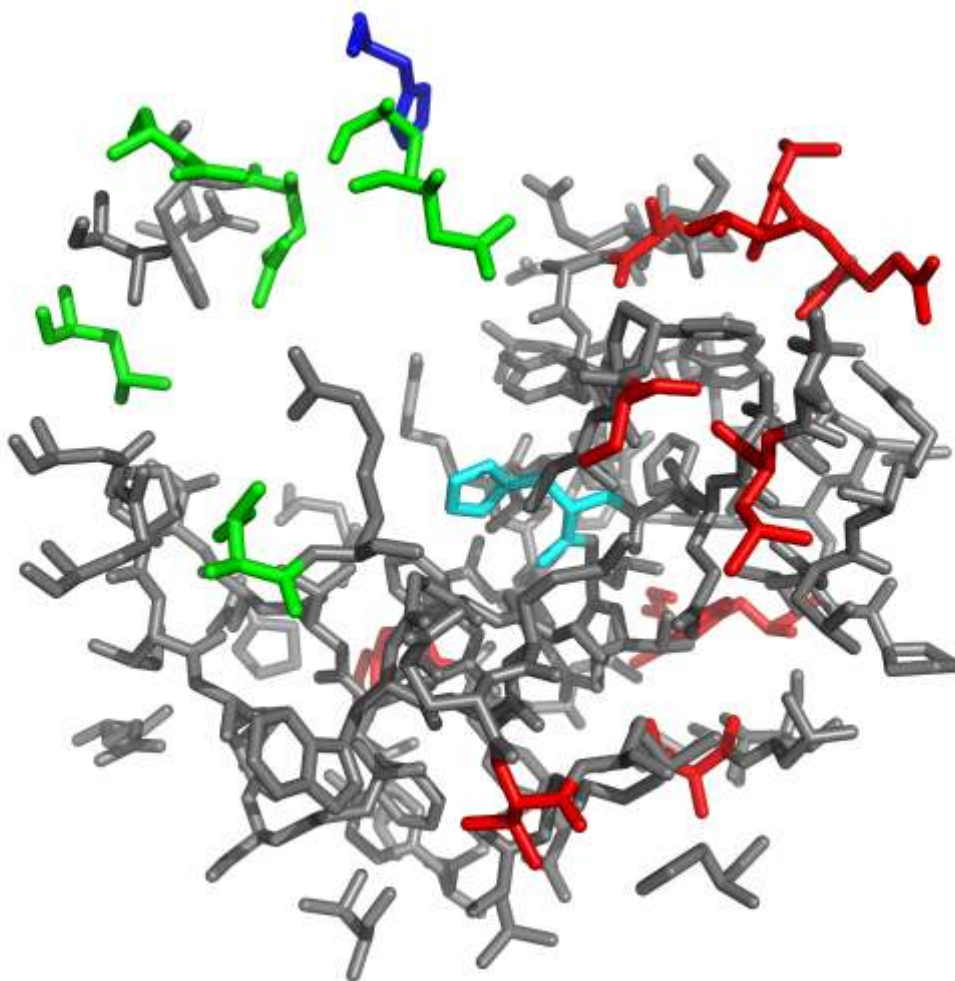


Figure 6-4. Residues non-identical to JC<sub>Lys</sub> and within 15 Å of the catalytic His residue of DC<sub>Thr</sub>. Residues within 15 Å of the catalytic His residue were identified in Swiss-PdbViewer (Guex and Peitsch, 1996). Non-identical residues within variable regions 1, 2 and 3 are coloured red, blue and green, respectively.

### 6.3 Probing the role of residues close to the catalytic His residue

From the modeling described in the section above, it was predicted that non-identical residues from the third variable region of JC<sub>Lys</sub> and DC<sub>Thr</sub> were most likely to be involved in determining acceptor site specificity and those from the first variable region were least likely to be involved. The strategy for research described in this section aimed to confirm the role of residues in the third variable region before focusing on the substrate specificity at the residue level. The reason for this approach was to mitigate any errors due to modelling or conformational changes that result from substrate binding and catalysis. The A and T domains of NRPS enzymes have been shown to be highly dynamic in structural studies (Sections 1.2.1.1 and 1.2.2.1) and analysis of C domains also suggests some

rotation between the subdomains that shifts the average distance between backbone  $\alpha$  carbons by 6 – 11 Å (Bloudoff et al., 2013). It was suggested this shift could be part of a catalytic mechanism (Bloudoff et al., 2013). However, the differences in C domain structures may be due to the structures all being created from different types of C domain, i.e. three of the structures are of different types of C domain based on their donor site specificity and VibH has unusual features as described in Section 1.2.3.1.

### 6.3.1 Testing the function of chimeric C domains

#### 6.3.1.1 Creation of a vector suitable for directed evolution

A secondary aim of the research described in this section was to create a substitution construct suitable for directed evolution. The plasmid pSW196 was useful in the work described in previous chapters, but its integration into *P. aeruginosa* PAO1 had relatively low efficiency. This did not pose a problem for domain substitutions because the constructs were constructed in *E. coli* before being transformed into *P. aeruginosa* PAO1, and because a large number of recombinant *P. aeruginosa* colonies were not required for further study. However, adopting a random or semi-random strategy to identify the residues involved in substrate specificity could involve screening large libraries of mutants. Since low transformation efficiency could be a problem for library screening, it was decided to use a non-integrating plasmid for the work described in this chapter.

Previous work using plasmid based complementation of *pvdD* found that a lower copy number plasmid containing *pvdD* resulted in increased pyoverdine levels compared to a high copy number plasmid (Ackerley and Lamont, 2004). This may have been due to the lower copy number plasmid over-expressing PvdD to a lower level, and hence establishing a better stoichiometric ratio between PvdD and the other pyoverdine NRPS enzymes (Ackerley and Lamont, 2004). To address this issue, during construction of pDEC (the C domain substitution plasmid used in this chapter) the  $P_{BAD}$  promoter from pSW196 was incorporated into the base plasmid pUCP22 (West et al., 1994). The pUCP22 plasmid contains a  $P_{TAC}$  promoter and was the high copy number plasmid used previously (Ackerley and Lamont, 2004). The  $P_{BAD}$  promoter was added since it has tighter regulation than the  $P_{TAC}$  promoter (Guzman et al., 1995). The domain architecture within pDEC is shown in Figure 6-5 and its construction is described in Section 2.8.3.1. This plasmid contained a version of *pvdD* lacking  $DC_{Thr}$  (i.e. without the second

C domain of *pvdD*) and contained restriction sites to insert new C domains in its place. The upstream restriction site was SpeI (labelled “X” in Figure 6-5) and allowed insertion of C domains via SpeI/XbaI fusion at the 5’ terminus. This restriction site was identical to the one used in pSMC for C-A domain substitutions (Figure 3-1B). The downstream site was SalI (labelled “S”) to allow insertion of C domains via XhoI/SalI fusion at the 3’ terminus. This restriction site was designed to be a silent, i.e. it did not alter the amino acid sequence of PvdD.

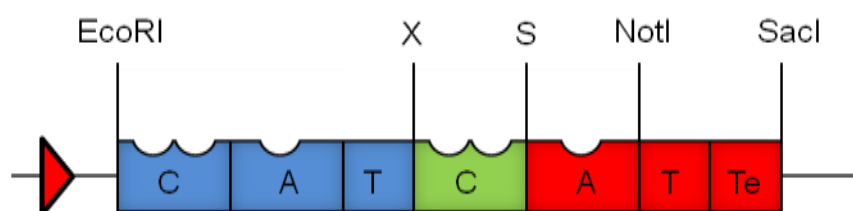


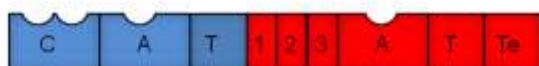
Figure 6-5. Domain architecture within the plasmid pDEC. Domains coloured blue and red were derived from the first and second modules of *pvdD*, respectively. Domains in green were absent in the base substitution plasmid and indicate where alternative domains would be added. X refers to a SpeI restriction site that was destroyed when ligating C domains into the plasmid via a compatible XbaI sticky end. S refers to a SalI restriction site that was destroyed when ligating C domains into the plasmid via a compatible XhoI sticky end.

### 6.3.1.2 Creation of chimeric C domains

To determine whether changing the third variable region of the C domain was sufficient to alter acceptor substrate specificity, three chimeric C domains were constructed (Figure 6-6) and ligated into the pDEC vector, as described in Section 2.8.3.2. The three chimeric C domains were based on JC<sub>Lys</sub> and had either one or two variable regions exchanged for the same region from DC<sub>Thr</sub>. The first of these, J12-D3, contained the first and second variable regions from JC<sub>Lys</sub> and the third variable region of DC<sub>Thr</sub>. In naming J12-D3 and other chimeric C domains, the origin of each variable region of the C domain is indicated, e.g. for J12-D3, “J12” indicates the first and second regions are identical to those from JC<sub>Lys</sub> and “D3” indicates the third variable region is identical to that from DC<sub>Thr</sub>. The C domain J12-D3 contains the first two variable regions from a C domain specific at the acceptor site for Lys and the third variable region from a C domain specific at the acceptor site for Thr. As such, the ability of the J12-D3 chimeric C domain to function in the context of *pvdD* would show this C domain can receive a Thr monomer at the acceptor site, and support the hypothesis that the residues conferring acceptor site

specificity are mainly located within the third variable region. However, the second and third variable regions were in the same subdomain, and it was also possible that changing the entire third variable region may disrupt residues interacting between the second and third variable regions. To test whether or not the second variable region might be involved in modulating the function of variable region three, a second chimeric C domain, named J1-D23, containing the first variable region from JC<sub>Lys</sub> and the second and third variable regions from DC<sub>Thr</sub> was constructed that kept the entire subdomain intact. For the strain containing J1-D23 to be capable of pyoverdine synthesis would be consistent with the third variable region being involved in acceptor site specificity. However, if J12-D3 was non-functional, it might also suggest that the acceptor site specificity is mediated by the second variable region. To control for this, the chimera J1-D2-J3 was created containing the first and third variable regions of JC<sub>Lys</sub> and the second variable region of DC<sub>Thr</sub>. In addition, a control with the complete JC<sub>Lys</sub> domain and a restriction site control containing the DC<sub>Thr</sub> domain were constructed. The chimeric C domain constructs were all sequence verified and transformed into the *P. aeruginosa* PAO1 *pvdD* deletion strain.

(A) Variable regions of the second C domain of PvdD



(B) Chimeric C domains

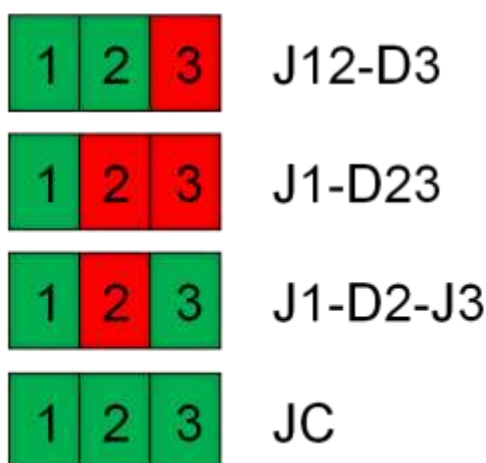


Figure 6-6. Domain architecture of *pvdD* and the chimeric C domains created to replace DC<sub>Thr</sub>. (A) The domain architecture of *pvdD* with the three variable regions of DC<sub>Thr</sub> indicated as 1-3. (B) The chimeric C domains created in this study. Green regions indicate origin from JC<sub>Lys</sub> and red regions indicate origin from DC<sub>Thr</sub>.

### **6.3.1.3 Analysis of the strains expressing chimeric C domain constructs on solid media and in liquid media**

Having created and transformed the chimeric C domain constructs, the levels of pyoverdine produced by the different strains were assessed both on solid media and in liquid media. The concentration of L-arabinose used in the assays was reduced to 0.1 % (v/v) because initial fluorescence tests in liquid media indicated that, at this concentration, the restriction site control was not impaired in pyoverdine production compared to the wild type strain and that the strain containing J12-D3 had maximum fluorescence at this concentration (Appendix E). When grown on solid media (Figure 6-7, panels A and B), the strain containing J1-D23 was the only chimeric C domain strain viable in the presence of EDDHA. This strain also had high levels of fluorescence, whereas the other strains were indistinguishable from the empty plasmid control. When grown in liquid media (Figure 6-7C), the strain containing J1-D23 produced high levels of pyoverdine as indicated by absorbance at levels of 67 % of the DC<sub>Thr</sub> control strain. This result confirms that the first variable region was unlikely to be involved in acceptor site specificity. The other chimeric C domain strains had no increased pyoverdine synthesis as measured by absorbance, but exhibited slight increases in fluorescence with the greatest increase for the strain containing J12-D3.

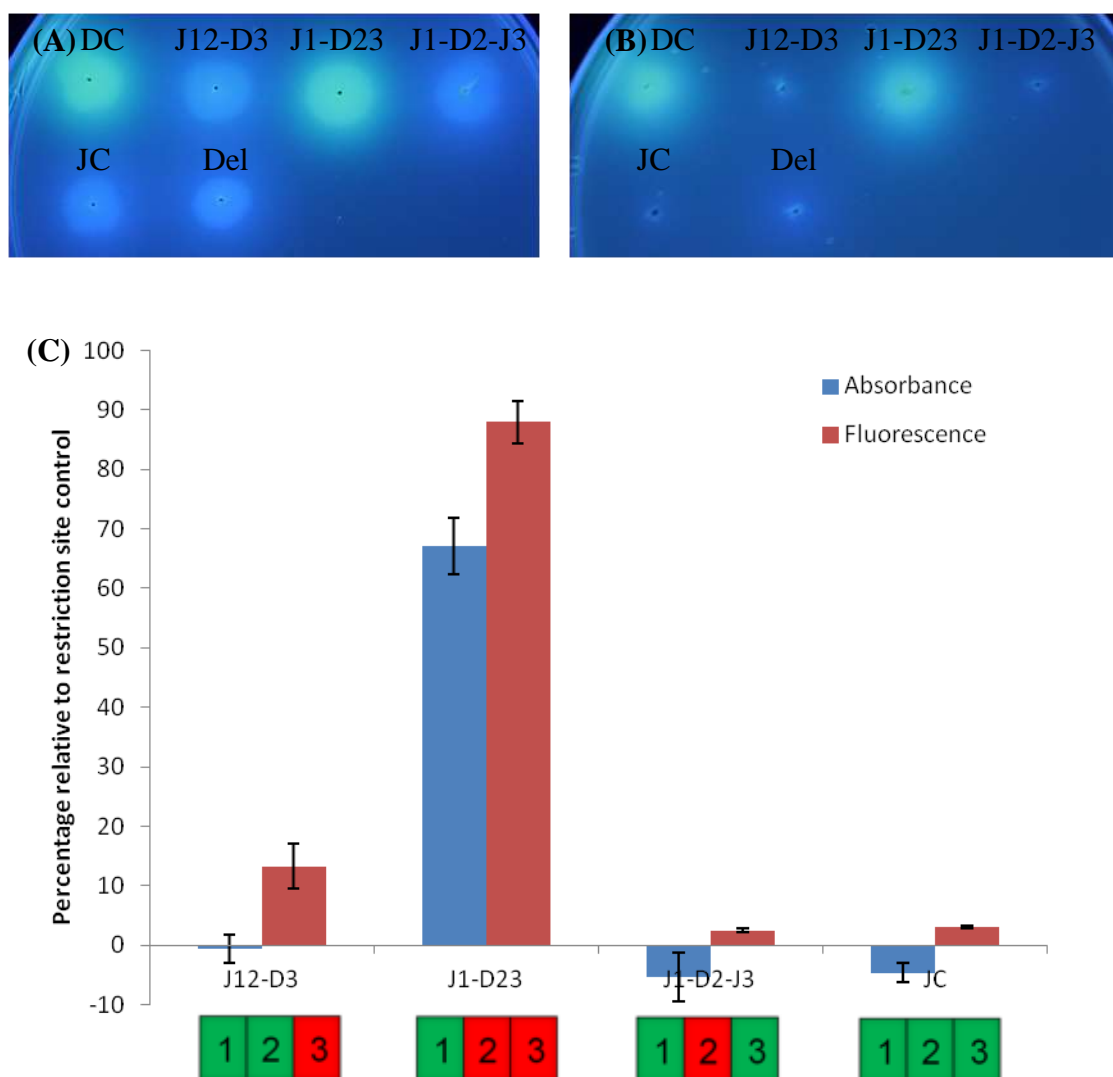


Figure 6-7. Pyoverdine production by chimeric C domain strains measured on solid media and in liquid media. (A) Growth of the chimeric C domain strains on solid KB media. The restriction site control is labelled DC, the strains with chimeric C domain labelled according to the C domain, the JC<sub>Lys</sub> substitution labelled JC and deletion strain labelled Del. Plates were incubated for 24 h, then photographs taken under UV light. (B) Growth on KB media supplemented with EDDHA. Conditions and labelling identical to (A). (C) Percentage of pyoverdine production from chimeric C domain strains grown in liquid media. Values are expressed as a percentage relative to the absorbance (400 nm) or fluorescence (ex. 400 nm/ em. 440 nm) of the restriction site control strain (DC), having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. A pictorial representation of each chimeric C domain was added below the *x*-axis, in which green regions indicate origin from *pvdJ* and red from *pvdD*. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation.

#### 6.3.1.4 MALDI-TOF detection of pyoverdine from chimeric C domain strains

The chimeric C domain strains were next analysed by mass spectrometry (Figure 6-8). The strain producing high levels of pyoverdine and able to grow on EDDHA containing media, namely J1-D23, was confirmed to be synthesising wild type pyoverdine. The chimeric C domain strain containing J12-D3 had an abundant peak corresponding to truncated pyoverdine of 991.5 m/z and a very small peak corresponding to wild type pyoverdine at 1333.6 m/z. This was unusual in that for all other strains capable of synthesising wild type pyoverdine the mass peak corresponding to wild type pyoverdine was always the most abundant. This unusual observation could perhaps indicate that acceptor site specificity had been partially or completely altered, but that the C domain was severely impaired in activity as a consequence of the recombination. In contrast, only truncated pyoverdine was detected from the J1-D2-J3 strain and the JC<sub>Lys</sub> control. This result was similar to that from non-functional C-A domain substitution strains (Section 4.4.2), however, was unexpected because this C domain was shown to be able to receive the donor peptide from the first module of PvdD in the context of a C-A domain substitution. An alternative explanation for peptide synthesis stalling at the first T domain of PvdD, for this case, is that the Thr specifying A domain from the second module of PvdD may not be able to activate Lys (even at low “leaky” levels) due to the charge and/or increased size of Lys relative to Thr. This is consistent with extensive testing of substrate promiscuity of the Phe specifying A domain from TycA, which found that the A domain had reduced activity against larger substrates and no activity with charged amino acids (Villiers and Hollfelder, 2009).

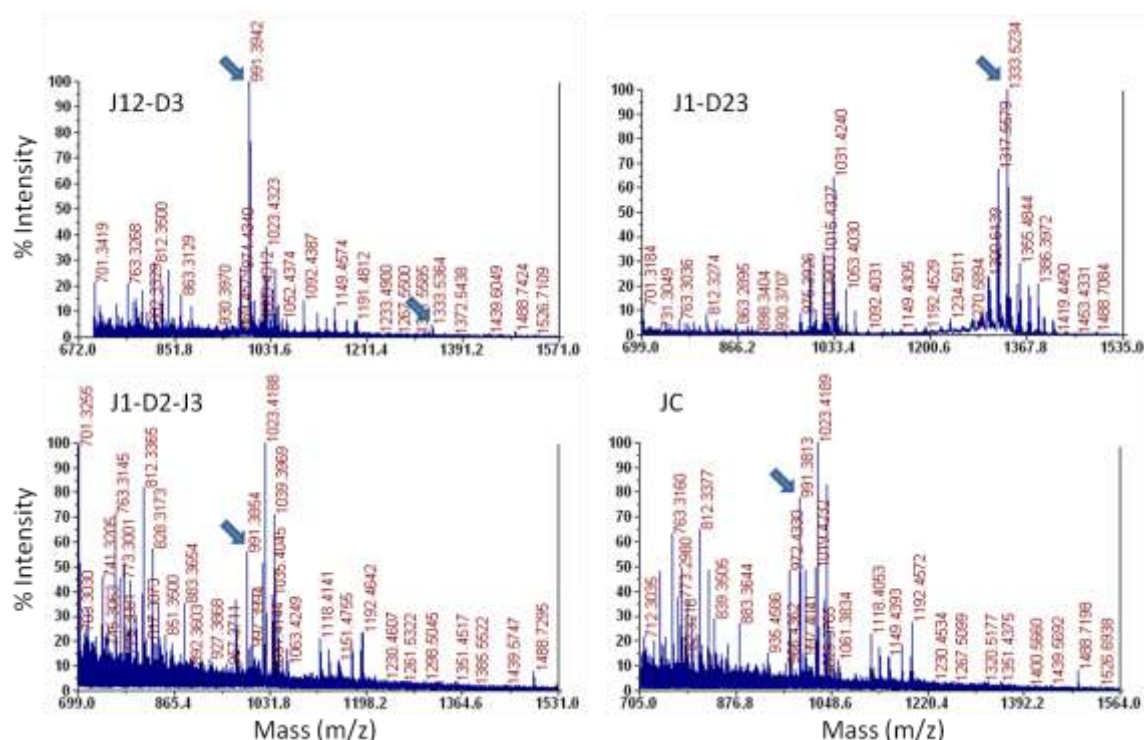


Figure 6-8. Mass spectrometry of the supernatant from strains expressing *pvdD* variants that contain chimeric C domains. Arrows indicate abundant peaks corresponding to either 991.5 or 1333.6  $m/z$  and an additional minor peak corresponding to 1333.6  $m/z$  for the strain containing the J12-D3 chimeric C domain.

### 6.3.2 Altering the loop covering the active site of the C domain

The above results suggested that the region responsible for acceptor site specificity was within the second and/or third variable regions of  $DC_{Thr}$ . From the structural modelling in Section 6.2.1, it was found that most non-identical residues between  $JC_{Lys}$  and  $DC_{Thr}$ , and close to the catalytic His residue, were located within the loop from variable region three that covered the active site. It was hypothesised that if these residues were involved in substrate specificity, then changing the loop of  $DC_{Thr}$  to the one from  $JC_{Lys}$  would inhibit activity. To test this, the loop from  $DC_{Thr}$  was replaced with the loop from  $JC_{Lys}$  as described in Section 2.8.3.3. Replacement of the C domain loop was confirmed by sequencing, and the resulting construct was then transformed into the *pvdD* deletion strain.

### 6.3.2.1 Testing pyoverdine production by the C domain loop substitution strain

The function of the C domain loop substitution strain was tested in comparison to the *pvdD* deletion strain and the DC<sub>Thr</sub> restriction site control strain. When grown on solid KB media (Figure 6-9, panels A & B), the loop substitution strain was fluorescent and grew in the presence of EDDHA at similar levels to the restriction site positive control. Testing the absorbance and fluorescence in liquid media found the levels of absorbance and fluorescence were identical to the positive restriction site control strain (Figure 6-9, panel C). The loop substitution strain was also confirmed to be synthesising wild type pyoverdine by mass spectrometry (Figure 6-9, panel D). These results indicated that the new loop did not impair the ability of the C domain to receive the Thr acceptor substrate, suggesting that it does not play a substantial role in substrate selectivity.

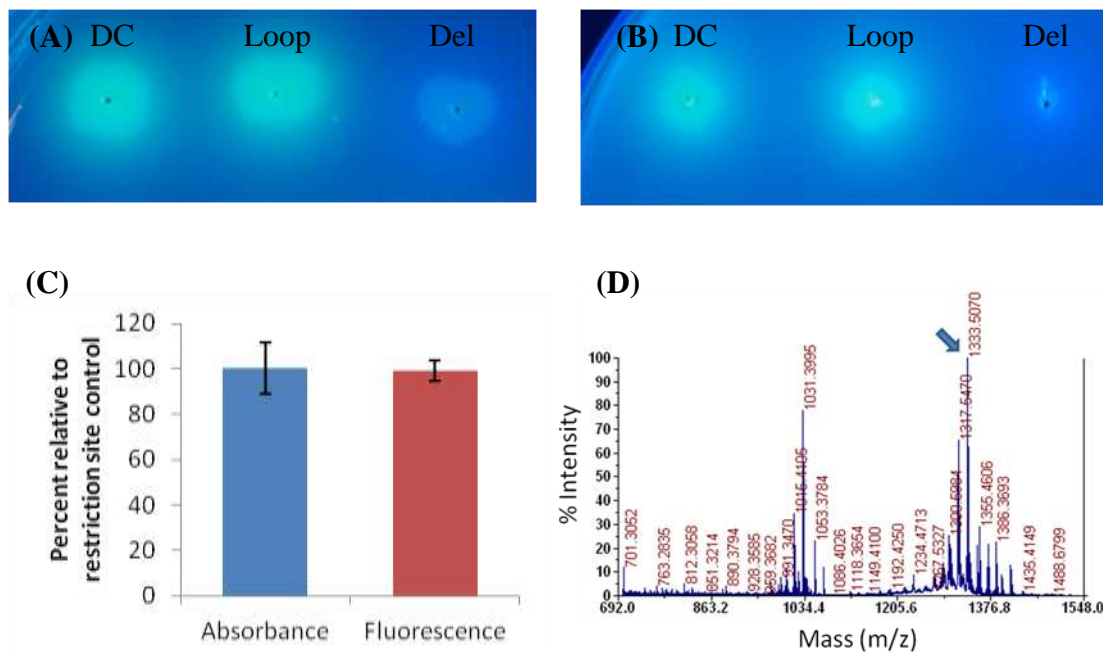


Figure 6-9. Analysis of pyoverdine production by the loop substitution strain. **(A)** Growth of the chimeric C domain strains on solid KB media. The restriction site control is labelled DC, the strains with chimeric C domain labelled according to the C domain, the JC<sub>Lys</sub> substitution labelled JC and deletion strain labelled Del. Plates were incubated for 24 h, then photographs taken under UV light. **(B)** Growth on KB media supplemented with EDDHA. Conditions and labelling identical to (A). **(C)** Percentage of pyoverdine production from the loop substitution strain grown in liquid media. Values are expressed as a percentage relative to the absorbance (400 nm) or fluorescence (ex. 400 nm/em. 440 nm) of the restriction site control strain (DC), having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation. **(D)** Mass spectrum obtained using the supernatant of the loop substitution strain. Arrow indicates peak consistent with 1333.6 m/z.

### 6.3.3 Recombination of the third variable region of JC<sub>Lys</sub> and DC<sub>Thr</sub>

The high levels of pyoverdine production from the strain containing the J1-D23 chimeric C domain suggested that residues within the second and third variable regions were sufficient for receiving a Thr monomer at the acceptor site. In contrast, the low levels of pyoverdine from strain J12-D3 indicated that either substituting the third variable alone disrupted protein structure, or the second variable region has a role in acceptor site specificity or in mediating activity of the third variable region. Moreover, the high level of pyoverdine production for the strain that had the third variable region loop from DC<sub>Thr</sub> replaced with the loop from JC<sub>Lys</sub> indicated that the predictions of residues involved in acceptor site specificity may be incorrect. This suggested the residues predicted to be close to the active site might not be the primary determinants of acceptor site specificity.

In an attempt to identify residues of the third variable region of DC<sub>Thr</sub> that were essential for activity in DC<sub>Thr</sub> with a Thr acceptor substrate, it was decided to shuffle the third variable region of JC<sub>Lys</sub> and DC<sub>Thr</sub>, and then ligate the recombined region downstream to the first and second variable regions of DC<sub>Thr</sub>. To ligate the recombined third variable region downstream to the first and second variable regions of DC<sub>Thr</sub>, the first two variable regions from DC<sub>Thr</sub> were PCR amplified and ligated into pDEC (the plasmid created for chimeric C domain substitutions in Section 6.3.1.1) to create the plasmid pTRN. Thus, pTRN contained *pvdD* lacking the third variable region of the C domain from the second module of *pvdD*. The plasmid pTRN contained SpeI/SacI restriction sites to ligate in the recombined third variable region. Both restriction sites were silent in terms of the PvdD amino acid sequence. The full details of construction for pTRN are described in Section 2.8.3.4.

As an initial test, the Staggered Elongation Process (StEP) (Section 2.6.3.4) was used to shuffle the third variable regions from JC<sub>Lys</sub> and DC<sub>Thr</sub> to determine whether this technique would provide diversity of C domains. StEP is a process whereby, during amplification of multiple templates by a DNA polymerase, a short elongation time at low temperature is used such that elongation is halted and, for homologous sequences, the truncated amplicon can bind either template in the next cycle (Zhao et al., 1998). This results in recombination of the domains in areas of homology. Here, StEP was used to recombine the third variable region from the C domains JC<sub>Lys</sub> and DC<sub>Thr</sub> using a forward primer that binds JC<sub>Lys</sub> and the reverse primer that binds DC<sub>Thr</sub>, and vice versa. The

recombined third variable regions were ligated into pTRN, and the ligated plasmids were transformed into the *pvdD* deletion strain. Active C domain variants were screened for in a 96-well plate assay using fluorescence according to methods in Section 2.9.4. In the fluorescence screening assays, using the JC<sub>Lys</sub> forward and DC<sub>Thr</sub> reverse primers generally resulted in strains that were highly efficient at synthesising pyoverdine (Figure 6-10A), whereas using the JC<sub>Lys</sub> forward and DC<sub>Thr</sub> reverse primers resulted in strains with generally lower levels of activity (Figure 6-10B).

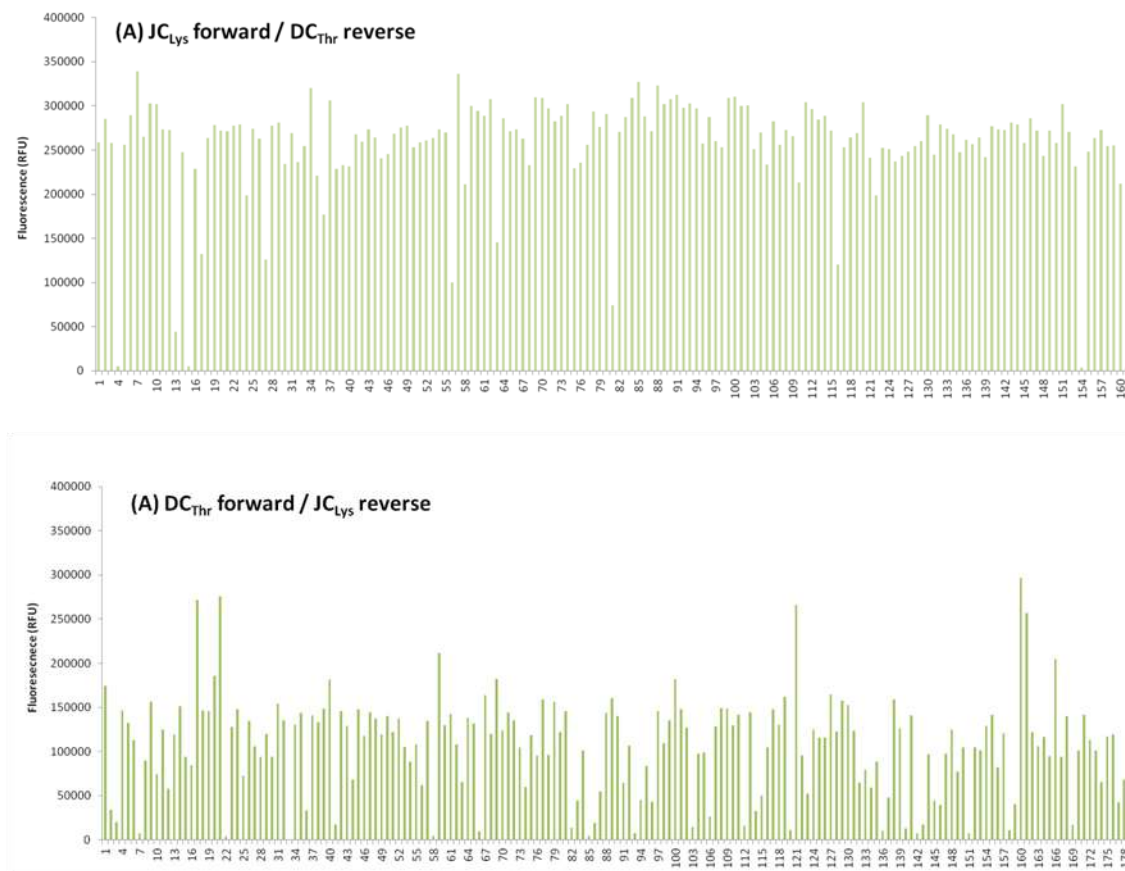


Figure 6-10. Fluorescence based screening of StEP library. (A) The fluorescence of StEP library hits generated with the Lys-C forward primer and Thr-WT reverse primer. (B) The fluorescence of StEP library hits generated with the Thr-WT forward primer and Lys-C reverse primer. Data are from one replicate.

To assess the levels of recombination, fifteen strains were selected to span a range of fluorescence and the plasmids from these strains were purified, sequenced and re-transformed into the *pvdD* deletion mutant. These strains were then reassessed for pyoverdine production by measuring fluorescence after growth in M9 medium (Figure 6-11). The strains containing C domains amplified using the JC<sub>Lys</sub> forward and DC<sub>Thr</sub>

reverse primers are labelled J1-J7, whereas the strains with C domains amplified with the opposite primers are labelled D1-D9. From the sequenced strains there were three levels of activity, i.e. highly active strains, moderately active strains, and one strain (J1) that had only trace levels of fluorescence. The C domain from the J1 strain contained an internal stop codon, explaining why only trace levels of fluorescence were observed for this strain.

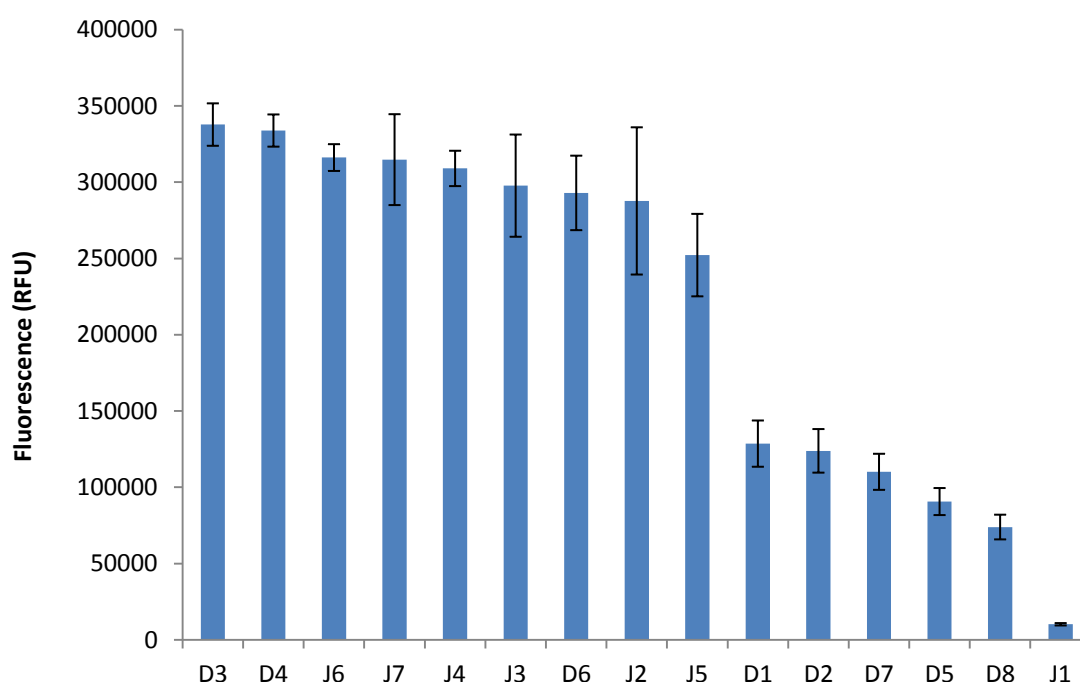


Figure 6-11. Fluorescence of strains that were sequenced from the StEP library. Strains labelled D1-D8 were amplified using the primers DC<sub>Thr</sub> forward and JC<sub>Lys</sub> reverse. Strains labelled J1-J7 were amplified using the primers JC<sub>Lys</sub> forward and DC<sub>Thr</sub> reverse. Data are the mean of 4 independent replicates, and error bars indicate 1 standard deviation.

The sequences for the shuffled third variable region from the sequenced strains from the StEP library, not including J1, were then aligned according to whether they were highly functional or moderately functional (Figure 6-12). The highly functional strains all had sequences highly similar to DC<sub>Thr</sub> (Figure 6-12, top) whereas moderately active strains were more related to JC<sub>Lys</sub> (Figure 6-12, bottom). In each case where the JC<sub>Lys</sub> forward and DC<sub>Thr</sub> reverse primers had been used, the JC<sub>Lys</sub> forward primer was found to bind the C domain DC<sub>Thr</sub>, resulting in amplification of the third variable region from DC<sub>Thr</sub>. An exception to this was strain J5 that had one cross-over close to the 5' terminus of the third

variable region. More frequent recombination caused by a single cross-over was observed when the DC<sub>Thr</sub> forward and JC<sub>Lys</sub> reverse primer set was used. For the highly active constructs generated using these primers (D3, D4 and D6), recombination occurred very close to the 3' end of the third variable region, suggesting the sequence upstream to this point needed to be DC<sub>Thr</sub> sequence for high levels of activity. In contrast, the moderately functional constructs were most similar to JC<sub>Lys</sub>. In fact D1 and D2 had the identical amino acid sequence to the third variable region of JC<sub>Lys</sub>. Interestingly, strain D5 (which exhibited similar fluorescence to the strains D1 and D2 containing the third variable region of JC<sub>Lys</sub>) had a third variable region composed of the first half from DC<sub>Thr</sub> and the second half from JC<sub>Lys</sub>. This indicated that the residues of DC<sub>Thr</sub> important for increased activity may be upstream to recombination point for strain D5 (Figure 6-12, highlighted in bold). Overall, the moderate levels of fluorescence observed for the strains containing the JC<sub>Lys</sub> third variable region suggested the third variable region might have less influence on acceptor site specificity than predicted based on the distance from the catalytic His residue in Section 6.2.1.3.

DC	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
JC	ALQPERSLGHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D3	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D4	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D6	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J2	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J3	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J4	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J5	ALQPERSLGHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J6	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
J7	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D1	ALQPERSLGHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D2	ALQPERSLGHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D5	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D7	ALQPERSLGHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
D8	ALQPERNASHNPLFQVLFNFHQSEIRSVTPEVQLEDLRLEGLAWDGGQTAQF
DC	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
JC	DLTLDVHEAEDGIWASFGYATDLFDASTVERLARHWONLLRGIVANPRQR
D3	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
D4	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
D6	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J2	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J3	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J4	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J5	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J6	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
J7	DLTLDIQEDENGIWASFDYATDLFDASTVERLAGHWRNLLRGIVANPRQR
D1	DLTLDVHEAEDGIWASFGYATDLFDASTVERLARHWONLLRGIVANPRQR
D2	DLTLDVHEAEDGIWASFGYATDLFDASTVERLARHWONLLRGIVANPRQR
D5	DLTLDIQEDEN <b>GIWASFG</b> YATDLFDASTVERLARHWONLLRGIVANPRQR
D7	DLTLDVHEAEDGIWASFGYATDLFDASTVERLARHWONLLRGIVANPRQR
D8	DLTLDVHEAEDGIWASFGYATDLFDASTVERLARHWONLLRGIVANPRQR
DC	LGELPLLDAPEEROTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
JC	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI
D3	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARALRPOAVALI
D4	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARLRPOAVALI
D6	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARLRPOAVALI
J2	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
J3	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
J4	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
J5	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
J6	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
J7	LGELPLLDAPEERQTLSEWNPAORECAVQGTLOORFEEQARORPOAVALI
D1	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI
D2	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI
D5	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI
D7	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI
D8	VAELPLLDEERDCISRAWAENADEGGIPPLVQLQIQEQARLRPOAVALI

Figure 6-12. Sequence alignments of the third variable region from strains that were highly or moderately active in pyoverdine synthesis identified by fluorescence screening of the StEP library. DC and JC are the sequences from DC<sub>Thr</sub> and JC<sub>Lys</sub>, respectively. Sequence alignments are continued over three lines, and for each line the sequences in the top rectangle are from highly active strains and the sequences in the bottom rectangle are from moderately active strains. Amino acid residues that are different from the sequence of DC<sub>Thr</sub> are highlighted. The area of recombination for strain D5 is highlighted in bold. Alignments generated in Geneious version 5.5.7 created by Biomatters. Available from <http://www.geneious.com/>.

### 6.3.3.1 Evaluation of the J12-D3 chimeric C domain strain

That the two StEP mutants containing the third variable region from JC<sub>Lys</sub> (D1 and D2) were functional in the context of *pvdD* suggests that acceptor site specificity is not

primarily defined by residues from this region. The strain D1 was analysed further in this section and is referred to as D12-J3. It should be highlighted that it contains the opposite variable regions to the previously tested J12-D3 strain (Section 6.3.1.3), which although unable to grow in the presence of EDDHA, was confirmed to have increased fluorescence in liquid media and to synthesise wild type pyoverdine by mass spectrometry, albeit the peak was unusual in its low abundance. The D12-J3 strain was tested alongside J12-D3 for direct comparison of pyoverdine production. When grown on solid media (Figure 6-13, panels A and B), D12-J3 had low levels of growth in the presence of EDDHA and low levels of fluorescence. In contrast (and consistent with previous results in Section 6.3.1.3), J12-D3 was not viable in the presence of EDDHA and did not have increased fluorescence compared to the *pvdD* deletion strain. In liquid media, D12-J3 showed increased fluorescence and absorbance relative to J12-D3 (Figure 6-13, panel C) and was confirmed to be synthesising wild type pyoverdine by mass spectrometry (Figure 6-13, panel D). Overall, these results show that D12-J3 was more functional than J12-D3. This suggests that the second variable region may have a larger role in acceptor site specificity than the third variable region and that the previous hypothesis of the importance of the third variable region (Section 6.2.1.3) was incorrect. Although not helpful in pinpointing the substrate specificity residues, that both J12-D3 and D12-J3 were able to synthesise at least some wild type pyoverdine suggests that the C domains may have slightly relaxed acceptor site substrate specificity rather than altered specificity. Further work would be needed to confirm whether these results are indeed a useful step towards development of a relaxed acceptor site specificity C domain.

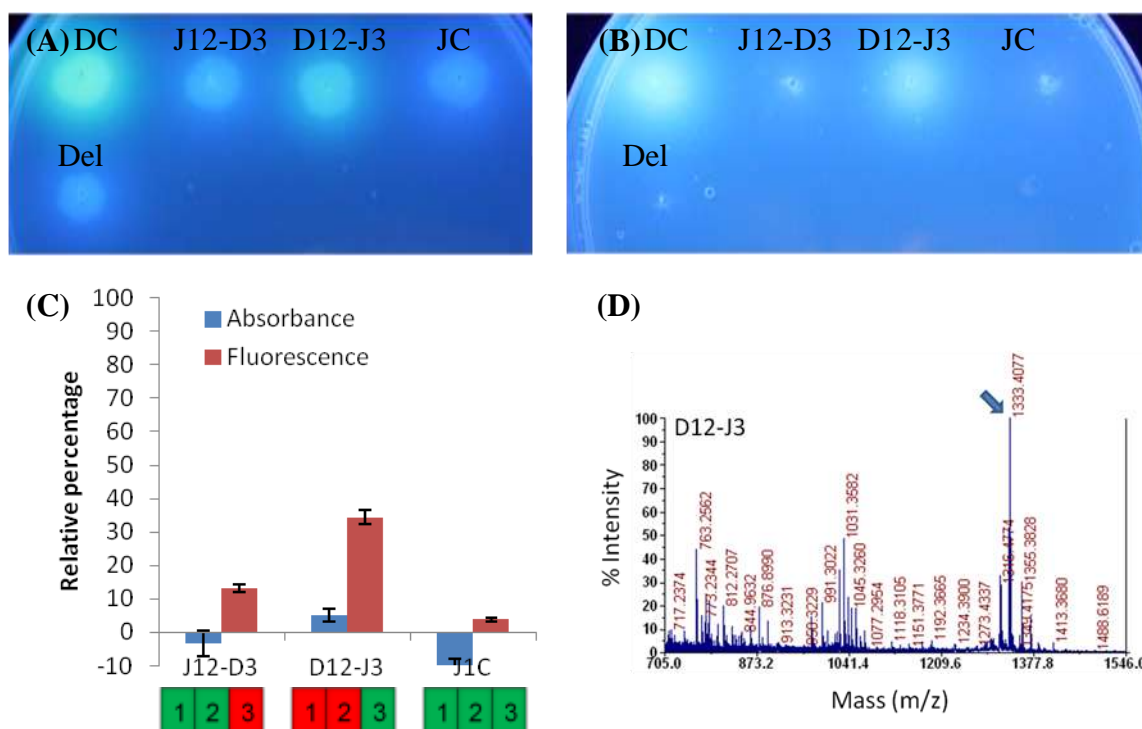


Figure 6-13. Analysis of pyoverdine synthesised by D12-J3 compared with J12-D3. (A) Growth of J12-D3 and D12-J3 strains on solid KB media. The restriction site control is labelled DC, the JCLys substitution strain is labelled JC and *pvdD* deletion strain is labelled Del. Plates were incubated for 24 h, then photographs taken under UV light. (B) Growth on KB media supplemented with EDDHA. Conditions and labelling identical to (A). (C) Percentage of pyoverdine production from J12-D3, D12-J3 and JCLys substitution strains grown in liquid media. Values are expressed as a percentage relative to the absorbance (400 nm) or fluorescence (ex. 400 nm/ em. 440 nm) of the restriction site control strain, having first been zeroed against the absorbance/fluorescence of the *pvdD* deletion mutant. A pictorial representation of each chimeric C domain was added below the x-axis, in which green regions indicate origin from *pvdJ* and red from *pvdD*. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation. (D) Mass spectrum obtained using the supernatant of the J12-D3 substitution strain. Arrow indicates peak consistent with 1333.6 m/z.

## 6.4 Discussion

This chapter aimed to narrow down the residues involved in defining the acceptor site specificity of JC<sub>Lys</sub>. A key hypothesis underpinning much of the research in this chapter was that residues close to the catalytic His would be most likely to be involved in substrate specificity. Based on homology modelling of DC<sub>Thr</sub>, the non-identical residues between JC<sub>Lys</sub> and DC<sub>Thr</sub> that are close to and have access to the catalytic His were identified mainly within the third variable region. Therefore, it was hypothesised that chimeric C domains that contained the third variable region of DC<sub>Thr</sub> would have the correct acceptor site specificity to function in the context of *pvdD*. The observations that (i) the J1-D23 strain was able to synthesise high levels of pyoverdine; and (ii) whereas the

J1-D2-J3 strain did not make any detectable levels of pyoverdine the J12-D3 strain did, albeit at only trace levels; together indicated that the third variable region might be more important than the first or second variable region in determining substrate selectivity. However, in conflict with this conclusion, the D12-J3 strain (i.e. with the opposite C domain variable regions to J12-D3), displayed increased activity relative to the J12-D3 strain, indicating that the second variable region might in fact be more important than the third in terms of accommodating a Thr residue at the acceptor site.

On the face of it, that both J12-D3 and D12-J3 were found to synthesise at least some wild type pyoverdine is a contradictory result. However, this may indicate a slight relaxation of acceptor site specificity for one or both of J12-D3 and D12-J3. Possible explanations for why substrate specificity may have been relaxed include that acceptor site specificity may be dependent on both the second and third variable regions such that acceptor site specificity was partially changed in each of J12-D3 and D12-J3, and/or that disrupting the subdomain caused structural changes that allowed greater access for alternative substrates to the acceptor site. Further testing would be required to definitively identify which regions are involved in specificity and whether the specificity of chimeric C domains was indeed relaxed. One way to test this would be by substituting the chimeric C domains into the second module of *pvdD* in partnership with the Lys specifying A domain of *pvdJ*. Since the Lys specifying A domain substitution strain in Chapter 3 was shown to synthesise low levels of wild type pyoverdine, using this Lys specifying A domain may allow synthesis of pyoverdine containing either a terminal Lys or Thr residue and could therefore be advantageous in indicating whether any of the chimeric C domains had relaxed specificity and whether the second variable region is most important for Lys condensation.

Although not yet definitive, the results in this chapter indicate that the second variable region of the C domain is the most important region for acceptor site substrate specificity. If this is true it follows that the predictions based on homology modelling were likely incorrect, because only one residue of the second variable region was within 15 Å of the catalytic His residue. Possible explanations for this incorrect prediction include that the 15 Å cut off that was imposed was too small, the model had large errors, or that there are large structural rearrangements upon substrate binding that substantially alter the location of key residues. The structural model appeared to be of reasonable accuracy with predicted errors mostly below 1 Å, and the distance of residues within the 15 Å of the

catalytic His residue was selected to be conservative. As discussed earlier, it has been suggested that large conformational changes may occur during substrate binding and catalysis by C domains (Bloudoff et al., 2013). As such, determining the precise residues for substrate specificity may support that large conformational changes occur during catalysis.

The shuffling of the third variable regions of JC<sub>Lys</sub> and DC<sub>Thr</sub> was an initial test to determine whether StEP would efficiently recombine the third variable regions. The primers used were not designed specifically for StEP PCR, and it was subsequently found that the forward primers could bind both sequences. This reduced level of recombination means new primers would be needed for StEP to result in higher levels of recombination. Nonetheless, the StEP recombination experiment showed that some activity was retained when the third variable region of *pvdD* was replaced with the equivalent sequence from JC<sub>Lys</sub>. In addition, for strains with C domains created using the DC<sub>Thr</sub> forward and JC<sub>Lys</sub> reverse primers, high levels of activity were associated with recombination very close to the 3' terminus of the third variable region. Since this recombination event was rare, few recombined C domains that were generated using these primers were subsequently found to have wild type levels of pyoverdine production (Figure 6-10B). When the cross-over location using these primers was in the middle of the gene (e.g. for strain D5), moderate levels of pyoverdine production were observed that were comparable to strains which contained the entire third variable region from JC<sub>Lys</sub>. Since the D5 strain contains the 5' half of the third variable region identical to DC<sub>Thr</sub>, it suggests that the 3' half of the third variable region of JC<sub>Lys</sub> may cause reduced activity seen for the moderately active strains.

To identify where the non-identical residues between JC<sub>Lys</sub> and DC<sub>Thr</sub> and within the 3' half of the third variable region were located in the context of a C domain, the residues were mapped onto 2VSQ (the structural model of SrfA-C) based on sequence alignment (Figure 6-14). The 2VSQ structural model is shown here, rather than RaptorXmsa, because 2VSQ contained an entire module and the C domain substitutions in this chapter contained part of the A domain up until the A1 motif. In Figure 6-14, the non-identical residues from the 3' half of the third variable region that were part of the C domain are indicated in orange, whereas all the residues from the A domain that were substituted with the C domain are shown in green. From the residues mapped onto 2VSQ, only one non-identical residue was close to the catalytic His residue, and all the remainder were at the C/A domain interface or part of the A domain. The location of these residues suggests

that the D5 strain may have lost activity by disrupting the C/A domain interface, and supports that the C/A interface is indeed important as previously suggested by Tanovic et al. (2008). Moreover, the reduced activity by disrupting the C/A domain interface may also explain the variation in levels of pyoverdine synthesis by Thr specifying A domain substitutions in Chapter 3, indicating that disrupting the C/A interface can reduce activity but often does not completely abolish activity. Additional screening of the StEP library may allow these residues to be further narrowed down.

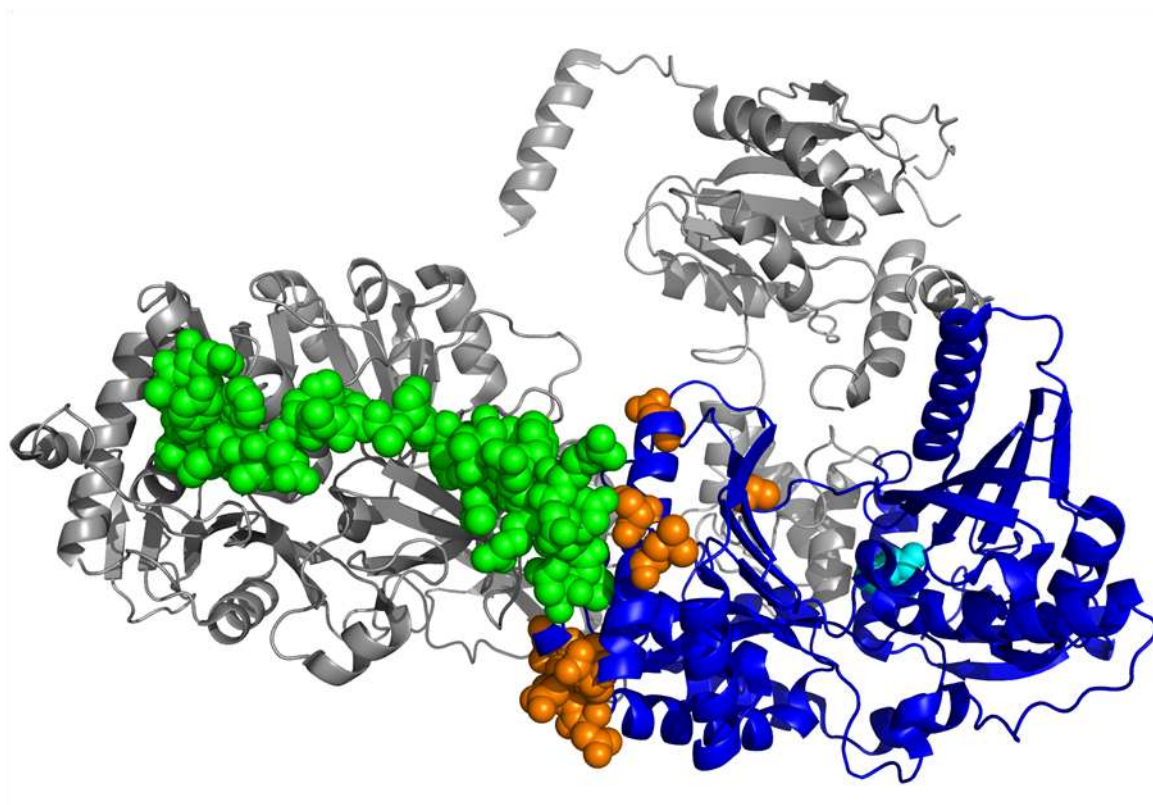


Figure 6-14. Location of residues potentially important for high levels of activity mapped onto 2VSQ, the structure of SrfA-C, based on sequence alignment. The C domain backbone is coloured in blue. Residues substituted as part of the C domain substitution are shown in green if they were part of the A domain and in orange if they were part of the C domain and non-identical between JC<sub>Lys</sub> and DC<sub>Thr</sub>. The catalytic His residue is shown in cyan.

The work in this chapter set out to narrow down the residues involved in substrate specificity of the C domain and to create a staging plasmid that might ultimately prove more suitable for directed evolution. For the first of these aims, it was attempted to narrow down residues based on predictions of residues close to the catalytic site. These predictions suggested most non-identical residues between JC<sub>Lys</sub> and DC<sub>Thr</sub> that were close to the catalytic His were located within the third variable region. Although not

entirely clear, the results suggest the initial hypothesis was incorrect and that the second variable region is most important for acceptor site specificity. More work is needed to understand acceptor site specificity and to test whether chimeric C domains created in this chapter are a step towards a C domain with relaxed acceptor site specificity. For the second aim of this chapter, a new plasmid was created and found to allow high levels of pyoverdine production. Moreover, in initial tests this plasmid was suitable for screening a small number of shuffled C domains. Overall, this chapter establishes a foundation for further work on C domain specificity and provides a plasmid suitable for directed evolution experiments.



## Chapter 7: **Key findings and future directions**

### 7.1 **Research motivation**

Non-ribosomal peptides comprise a diverse set of peptides synthesised by NRPS enzymes. These peptides are of interest as many of them exhibit antibiotic, anticancer, anti-inflammatory, immunosuppressant or antiviral activities. In addition, the assembly line-like mode of synthesis and semi-autonomous nature of domains and modules indicates that they could potentially be manipulated as a means of creating new bioactive products. This would be advantageous because most new medically relevant non-ribosomal peptides are in fact synthetic or semi-synthetic derivatives of natural product scaffolds (Butler, 2005; Newman and Cragg, 2012; Walsh, 2003). The ability to functionally modify NRPS enzymes by domain substitution may provide a direct biosynthetic route to high yield generation of derivatives of current natural product scaffolds and potentially the *de novo* construction of entirely new NRPS pathways. However, historically domain substitution experiments have almost invariably caused a large loss in function, limiting product yield. Although a number of reasons have been suggested, the precise causes for the loss of function in domain substitution experiments remain unknown. The research in this thesis aimed to develop pyoverdine as a model for domain substitution and to identify constraints to domain substitutions.

### 7.2 **Key findings**

The research described in the third chapter of this thesis aimed to determine whether the function of A domain substitutions is mainly limited by the acceptor site substrate specificity of the partner C domain or by disruption of the C/A domain interface. It was found that Thr specifying A domain substitution strains all yielded relatively high levels of pyoverdine, whereas non-Thr specifying A domain substitution strains resulted in only trace levels of wild type pyoverdine. The high success of Thr (but not non-Thr) specifying A domain substitutions was consistent with previous A domain substitutions in module one of *pvdD* by Ackerley and Lamont (2004). Collectively, these results indicated that A domain substitutions in *pvdD* are mainly limited by the substrate specificity of the C domain rather than by disrupting the C/A domain interface. An implication of these

results is that if C domain acceptor site specificity can be overcome, then A domain substitution may be a viable approach for creating new products.

As an alternative, C-A domains were substituted together to bypass acceptor site constraints. The results of this research confirmed that some C-A domain substitutions can produce modified full-length pyoverdines. However, in contrast to A domain substitutions, the success of C-A domain substitutions did not follow any obvious predictable rule. Moreover, a key difference from A domain substitutions was that C-A domain substitution strains exhibiting trace levels of fluorescence were found to be synthesising a truncated pyoverdine. This suggested that the peptide was being released prior to condensation by the introduced C domain. C domain donor site specificity – at least in terms of the incoming C-terminal residue – did not appear to be the primary limitation because the C-A domain substitutions that gave the highest yield of modified pyoverdine products contained C domains that normally receive L-hfOrn at the donor site, rather than the L-Thr they were receiving post-substitution into PvdD. Further evidence for relaxed donor site specificity was provided by the synthesis of pyoverdine from the strain CAhfOrn-J in Chapter 5, which also functioned in *pvdD* and contained a C domain that normally receives L-Lys at the donor site.

Since donor site specificity was unlikely to be the main cause for C-A domain substitutions to be impaired in pyoverdine production, two alternative hypotheses were developed to explain why C-A domain substitutions are frequently non-functional. These were that T/C domain interactions may be impaired as a consequence of C-A domain substitutions or that the substituted C domains may not be capable of receiving the incoming peptide chain due to steric constraints at the donor site.

Chapter 5 addressed the first hypothesis, by studying T domain interactions to investigate whether the function of C-A domain substitutions may be impaired due to disrupting T domain interactions with the surrounding domains. This work introduced a second NRPS model system, the single-module indigoidine synthetase BpsA, which was particularly well suited for T domain substitution. Previous directed evolution experiments by Dr Owen suggested that there are key residues (in particular at the +4 and +24 positions relative to the conserved Ser site of 4'-pp attachment) that can be functionally distinct in the otherwise closely related T domains found upstream of either C or TE domains. The *pvdD* system was used to show that residues at the +4 and +24

positions mutated to improve the ability of the first T domain of PvdD to interact with the TE domain of BpsA also improved its ability to interact with the TE domain of PvdD. However, despite the ability of specific mutations to dramatically improve the interaction of T<sub>C</sub> domains with TE domains, sequence and phylogenetic analyses found that T<sub>C</sub> and T<sub>TE</sub> domains were highly related to one another. Returning to the over-arching theme of identifying possible constraints to NRPS recombination studies, it was proposed that T<sub>C</sub> domains may similarly have optimised interactions with the downstream C domain and that C domains newly introduced by C-A domain substitution may therefore be unable to interact with the upstream T domain. If this issue were a major cause for decreased or inhibited function in C-A domain substitution studies, then it could potentially be bypassed by substituting T-C-A domains together to keep the T/C domain interactions intact. However, when three T-C-A domain substitutions were created and tested in comparison to the corresponding C-A domain substitutions, no major differences in pyoverdine production were observed between the T-C-A and C-A domain substitution strains. Since results from A domain substitutions had already indicated that A/T domain interactions are usually not substantially impaired in domain substitution experiments (at least, at the recombination sites used in this work), these results suggested that keeping the native T/C domain interactions intact had little effect.

To disprove the hypothesis that impaired T/C domain interactions were the major cause for loss of function for C-A domain substitutions, it was decided to substitute T domains into the first module of *pvdD*. This shift from T-C-A domain substitutions was because the T domain substitution constructs were much easier to build due to the small size of the insert. In addition, substitution of T<sub>C</sub> domains into this position were conceptually similar to C-A domain substitutions in that they established a new A/T domain interaction, C-/T domain interaction and T/C domain interaction, and used equivalent fusion sites. This meant that they tested whether any T domain interactions might be disrupted by C-A domain substitutions, rather than solely focusing on the T/C interactions. In all, six T<sub>C</sub> domain substitutions were constructed, and it was found that none of these were impaired in activity relative to the restriction site control. In contrast, twelve other types of T domain (four of each of T<sub>Ct</sub>, T<sub>E</sub> and T<sub>TE</sub> domains) were also substituted and the resulting substitution strains found to have variable levels of pyoverdine synthesis. Nonetheless, the overall success rate was still high, with only three T domain substitution strains being completely inhibited in pyoverdine production. This indicates that there is

some specialisation of T domains and, consequently, that it would ideally be preferable (albeit not necessarily essential) to keep the type of domain downstream to a T domain constant in order to increase the chance of success for domain substitutions. Overall, these results suggest that T domains are relatively tolerant in interacting with new domains. The most important implication of this is that disrupted T/C domain interactions were unlikely to be the major cause of impaired NRPS function during the earlier C-A domain substitution experiments.

The second hypothesis mentioned above for why C-A domains become non-functional when substituted was that steric constraints at donor site of the new C domain may block the peptide chain from accessing the catalytic site of the C domain. This would in turn suggest that C domains have potential to be limiting for domain substitutions because of constraints at both the donor and acceptor sites. The final research chapter aimed to create a plasmid suitable for directed evolution of C domains and to narrow down the region of the C domain responsible for acceptor site substrate specificity. It was decided to focus on acceptor site specificity instead of donor site specificity because it was thought that the former would be more defined. The residues within close proximity to the catalytic His residue were predicted based on homology modelling and hypothesised to be the residues most likely to be involved in determining acceptor site substrate specificity. Several chimeric C domains were created to test this. From these results, it appears that the residues predicted to be close to the catalytic His have only small effects on acceptor site substrate specificity and the second variable stretch of amino acids appeared to have the most involvement in substrate specificity. However, results were not definitive and further work will be needed to identify the key acceptor site residues. Although not successful in identifying substrate specificity conferring residues, the plasmid constructed was found to be highly efficient in terms of expressing *pvdD* in a manner that did not impair pyoverdine synthesis (in contrast with previous studies where higher copy number of *pvdD* was associated with decreased pyoverdine production (Ackerley and Lamont, 2004)). In addition, initial tests using StEP to recombine the third variable regions of two C domains found that this plasmid was suitable for directed evolution experiments. Thus, the work described in this chapter provides a solid foundation for future work aiming to use directed evolution to interrogate, and perhaps overcome, C domain specificity constraints.

In light of the frequent loss of activity that has been observed for NRPS domain substitution studies (Baltz, 2012; Giessen and Marahiel, 2012), perhaps one of the more surprising results of this thesis was in fact how tolerant *pvdD* was for domain substitutions. This is shown by the success of Thr specifying A domain substitutions, the three functional C-A domain substitutions that led to modified full-length pyoverdine products, the similarity in activity recorded for equivalent T-C-A and C-A domain substitutions, and the tolerance of many modifications of the C domain reported in the final chapter. Factors that may have contributed to non-functional substitutions were predicted to be acceptor site specificity of the C domain, steric constraints at the donor site of the C domain, some specialised T domains not functioning with a downstream C domain or by making large changes within a subdomain of the C domain in Chapter 6. The final two factors can frequently be avoided when attempting to make new products using domain substitution, e.g. by ensuring the type of domain downstream to a T domain is not changed and making substitutions that keep domains intact. Consequently, a more global conclusion from this thesis is that domain substitutions that do not interfere with C domain constraints are likely to yield a functional *pvdD* enzyme. An implication of this is that if the limitations of the C domain could be overcome, then it may be that NRPS enzymes become relatively tolerant to domain substitution as a means of creating new products.

### 7.3 Limitations of this research

#### 7.3.1 Limitations in the use of pyoverdine NRPS enzymes

This project limited the substituted domains to those involved in pyoverdine synthesis from four *Pseudomonas* strains. The diversification of pyoverdine side-chains is proposed to be a mechanism to stop siderophore cheating, i.e. to inhibit other strains utilising the synthesised siderophore (Schalk and Guillon, 2013; Visca et al., 2007). A suggested evolutionary mechanism for the diversification of pyoverdines is that the pyoverdine uptake receptor is first modified, which in turn creates pressure for the rapid evolution of the peptide side-chain to enable interactions with the new receptor (Smith et al., 2005). It has been hypothesised that many different pyoverdine genes in nature have derived from recombination and horizontal gene transfer of NRPSs between pseudomonads (Smith et al., 2005), and pyoverdine pathways may therefore have become generally more

amenable to domain recombination experiments. As such, it is unclear how transferable these results may be to other systems or with distantly related domains. However, an important reason for focusing on related domains in this thesis was to maximise the likelihood of functional recombination events, and thereby highlight factors that might be inhibiting function when recombination resulted in impaired activity. Although additional factors may limit the ability to substitute more distantly related domains, understanding the factors inhibiting the function in one system provides a starting point for determining why other pathways might be non-functional.

A second limitation of this study was that the modified derivatives of pyoverdine that were successfully generated have limited uses. Other products exist, such as antibiotics, where making derivatives could prove to be highly valuable (e.g. the daptomycin and related antibiotic derivatives generated by Baltz and coworkers; (Doekel et al., 2008; Nguyen et al., 2010)). However, pyoverdine has an advantage in that the wild type pyoverdine from *P. aeruginosa* PAO1 and the derivatives created in this study are easily detectable in chromogenic or fluorogenic assays, and, as a siderophore, allow growth in the presence of the iron chelator EDDHA. This makes pyoverdine a valuable model for testing domain substitutions and, as suggested above, the lessons learned in this system could aid in modifying other pathways. In addition, an advantage added by this research is that *pvdD* has now been characterised in-depth by domain substitution, i.e. in terms of both the number and variety of domain substitutions.

### 7.3.2 Limitations in testing pyoverdine production

In this study, the domain substitution strains were only tested by pyoverdine synthesis using *in vivo* assays. A consequence is that the level of stability and expression for individual recombinant proteins is unknown, and it may be that the activity of substituted domains was impaired in a manner that could have been more easily quantified in *in vitro* assays. However, the purification of NRPS enzymes is technically very challenging and can be variable, causing insolubility or inactivity. In particular, functional PvdD has only been successfully purified once before (Ackerley et al., 2003), and a large number of attempts to repeat this successful purification have all met with failure (DF Ackerley, *personal communication*). Moreover, *in vitro* assays commonly focus on reactions of individual domains and are not always predictive of *in vivo* activity. The focus of this research was on identifying robust ways in which domain substitutions can be performed

in the context of an NRPS pathway with the ultimate goal of creating new products *in vivo*. As such, the *in vivo* level of pyoverdine synthesis is substantially the most important metric for testing the success of domain substitution variants.

Another limitation was that MALDI mass spectrometry does not provide a quantitative measurement of the different types of pyoverdine synthesised. Many factors can affect the peak intensity including different ionisation efficiencies of similar peptides, small changes in sample preparation and saturation of the ion detector (Albrethsen, 2011; Szájli et al., 2008). As a consequence, when truncated and full-length pyoverdines were detected in the supernatant, their relative yields remain unknown. However, the most important aspect of this study was detecting when domain substitution strains were able to create the full-length pyoverdine, and MALDI mass spectrometry was valuable in determining this.

In the final chapter of this thesis, a total of 338 colonies were screened using fluorescence in a 96-well plate based assay. The screening method could be greatly scaled up, however, screening in 96-well plates is often limited to small libraries with a maximum limit of screening  $10^5$  variants a day under ideal conditions, whereas selection or fluorescence-based screening can screen much larger libraries of up to  $10^9$  variants (Dietrich et al., 2010; Leemhuis et al., 2009). As such, the screening method that was used here is highly limiting for the directed evolution techniques that could potentially be used on *pvdD*. Nonetheless, the modified pyoverdines in this study generated high levels of fluorescence and enabled growth in the presence of the iron chelator EDDHA. These characteristics provide the potential to create high-throughput screens focused on the siderophore or fluorescent properties of pyoverdine.

## 7.4 Future directions

### 7.4.1 Improve the activity of A domain substitutions

The work presented in Chapter 3 suggested that the functionality of A domain substitutions is primarily limited by acceptor site specificity of the C domain. However, there was still variation in the level of activity observed for the different Thr specifying A domain substitution strains. It would be useful to find ways to improve the activity of A domain substitution strains because it would add value if the acceptor site specificity of

the C domain can be altered or relaxed. One of the main hypotheses for why A domain substitutions may be unsuccessful is that they disrupt an important C/A domain interface (Tanovic et al., 2008). Although this appeared not to have been a major factor in limiting pyoverdine synthesis in the A domain substitution experiments presented here, it may have been a contributing factor to the Thr specifying A domain substitutions having variable levels of activity. The C/A domain interface hypothesis is in fact supported by initial results in Chapter 6, where it appeared that residues at this interface were essential for high levels of activity. In particular, residues at the C/A domain interface and important for high levels of activity were found to be located in the 3' half of the third variable region of the C domain. This indicates that using a new fusion site for A domain substitutions that is located partially into the C domain might avoid disrupting the C/A domain junction. Further screening of the two StEP libraries that recombined the third variable regions from JC<sub>Lys</sub> and DC<sub>Thr</sub> could be useful in confirming this and for indicating the best location for a fusion site. Another option would be to test the least functional Thr specifying A domain substitution with different fusion sites located further into the C domain. For a fusion site found to improve activity, the equivalent fusion site could be tested in other Thr specifying A domains to determine whether the new fusion site is generally able to improve activity.

#### **7.4.2 Perform T1 domain substitutions of T<sub>C</sub> domains from non-pyoverdine NRPS enzymes**

In Chapter 5, T domain substitutions into the first module of PvdD found that 6/6 T<sub>C</sub> domain substitution strains were highly functional for pyoverdine production and 9/12 domain substitution strains created using other types of T domain were able to synthesise pyoverdine at varying levels. These high success rates for pyoverdine production suggest that T domains are highly portable, and hence T domain function is unlikely to be substantially impaired by C-A domain substitution. This is important, as the ability of T domains to interact with new domains is essential for domain substitution experiments to be successful. However, because the T domains in this study were sourced exclusively from *Pseudomonas* pyoverdine NRPS enzymes, it is unknown how generally applicable these T domain substitution results may be. Thus, it would be of interest to repeat these T domain substitutions using a range of T domains from other NRPS enzymes to determine how portable other T domains are in interacting with new domains. The most

important T domains to substitute would be T<sub>C</sub> domains because substitution of T<sub>C</sub> domains mimics the most similar scenario to C-A domain substitutions in terms of the interactions altered.

#### 7.4.3 **Identify residues involved in acceptor site specificity of the C domain**

The research in this thesis highlights the need to understand the substrate specificity of the C domain. The work in Chapter 6 attempted to narrow down the acceptor site substrate specificity of the C domain from the second module of *pvdD* because this may enable non-Thr A domain substitutions to be successful. To narrow down the acceptor site specificity it was attempted to identify which region of the C domain is involved. Based on homology modelling and prediction of the residues likely to be in close proximity to the catalytic His residue, it was hypothesised that residues within the third variable region of the C domain would most likely be involved in substrate specificity. However, the residues of the third variable region appeared to have only a small influence on substrate specificity, with greater influence instead from the second variable region. Moreover, it was found that strains containing diametrically opposed combinations of variable regions of the C domain could both synthesise at least trace levels of wild type pyoverdine. Further testing of each of the chimeric C domains will be needed to identify whether substrate specificity was altered or relaxed, and which region(s) play the most important role in determining substrate specificity. As such, it may be useful to substitute the Thr specifying A domain of the second PvdD module with the Lys specifying A domain of PvdJ for each of the chimeric C domain constructs. Results from Chapter 3 suggest the PvdJ A domain is able to activate Thr with low efficiency in addition to Lys at high efficiency. Studying the pyoverdine production of strains containing these constructs may enable clearer results for identifying when specificity is altered versus relaxed. Following this, the substrate specificity could be further defined using the constructed pUCP22 based plasmid and directed evolution techniques. The choice of techniques for directed evolution would be highly dependent on the throughput of the screening method available and whether the acceptor site specificity of the C domain was narrowed down by testing the chimeric C domains with an alternative A domain.

#### 7.4.4 Characterise the donor site constraints of the C domain

An important corollary to identifying acceptor site constraints would be to define donor site constraints. An advantage for relaxing constraints that may exist at the donor site over focusing on acceptor site specificity is that changing acceptor site specificity may subsequently lead to problems with donor site constraints at the downstream C domain. However, altering donor site constraints has the limitation that – if based primarily on steric constraints – it may not be defined by consistent rules. Nonetheless, identifying the location of residues that block access to the catalytic site of the donor site of a C domain might prove useful in identifying potential problematic residues in other C domains or assist in identifying better recombination sites within the C domains. From the work in this thesis, the C domain that may be particularly useful for donor site specificity studies would be the C domain from the strain CAThr-E. The strain CAThr-E was found to synthesise the truncated pyoverdine product and was hypothesised to be limited by donor site steric constraints. As this C domain is not limited by acceptor site specificity (i.e. “expects” to receive Thr in the acceptor site) it would be an ideal choice for studying donor site specificity in the PvdD model.

#### 7.4.5 Development of large scale screens

Future experiments could also focus on developing a high throughput screen or selection for pyoverdine synthesising strains. As mentioned in the limitations section above, the initial screen used in this study would limit throughput, and hence the range of directed evolution techniques that could be used on *pvdD*. However, the siderophore and fluorescent properties of pyoverdine give the potential to create powerful selections or screens for pyoverdine synthesis. The siderophore property is useful as it could be used to establish a conditional viability selection for directed evolution experiments. This would most likely require selection on solid media because secreted pyoverdine in liquid media could also be taken up by strains not synthesising pyoverdine themselves. The fluorescent property of pyoverdine is useful as it may enable the use of fluorescence activated cell sorting (FACS) to screen for pyoverdine producing strains. Similar to selection, a potential problem for FACS based screening is that pyoverdine is secreted and could be taken up by less functional strains. However, it has been shown that *P. aeruginosa* PAO1 strains with mutations of the iron uptake receptor FpvA and efflux pump PvdRT-OpmQ are inhibited in pyoverdine uptake and have increased intracellular pyoverdine (Hannauer

et al., 2010). Both mutations would likely be needed in concert because there is still some secretion of pyoverdine when PvdRT-OpmQ alone is mutated (meaning that it is available for other strains to take up), yet deletion of FpvA alone does not lead to the same accumulation of pyoverdine (Hannauer et al., 2010). FACS would likely need to be combined with secondary screening because the fluorescence on a cell-to-cell basis in FACS based screening can differ greatly due to variation (Dietrich et al., 2010), which may cause overlap in the populations of the pyoverdine producing and pyoverdine deficient strains. For secondary screening a fluorescence based screen in 96-well plates would be valuable.

## 7.5 Concluding remarks

The potential to create new non-ribosomal peptides has created interest in NRPS domain substitution. However, it has time and time again proven difficult to alter non-ribosomal peptides using this method. This thesis has greatly extended previous work developing the enzyme PvdD as a model system to study the limitations to domain substitution. The enzyme PvdD was chosen for study because the pyoverdine product is fluorescent, required for growth in an iron limited environment, and there is a large amount of diversity in the naturally evolved peptide chains between *Pseudomonas* strains. The work in this thesis adds to this in that PvdD has now been extensively characterised by domain substitution and was shown to be highly amenable to certain forms of domain substitution.

The main aim of this thesis was to identify barriers to effective domain recombination for the creation of novel non-ribosomal peptides. It was found that A domain substitutions are mainly impaired by C domain acceptor site specificity and hypothesised that C-A domain substitutions are inhibited by steric constraints at the C domain donor site. This provides more definition for why domain substitutions in *pvdD* lose activity and establishes a firm foundation for further work to identify how these constraints can be addressed, defined and overcome. In particular, this research may guide future development of relaxed specificity C domains, which could prove to be invaluable tools for targeted modification of biotechnologically relevant non-ribosomal peptide products.



## Appendix A. Accession numbers of NRPS genes used as templates for domain substitution, and the substrate specificity of key A domains.

The accession numbers of genes used for domain substitution in this study are shown in Table A-1. Gene and protein sequences were obtained from the Pseudomonas Genome Database (Winsor et al., 2011), and were sourced from the genome sequences of *P. aeruginosa* PAO1 (Stover et al., 2000), *Pseudomonas syringae* pv. phaseolicola 1448A (Joardar et al., 2005), *Pseudomonas putida* KT2440 (Nelson et al., 2002) and *Pseudomonas fluorescens* SBW25 (Department of Energy Joint Genome Institute).

Substrate specificity of modules was based on previous annotation by Ackerley and Lamont (2004) (*P. aeruginosa* PAO1); Owen & Ackerley (2011) (*P. syringae* 1448a); Moon et al. (2008) (*P. fluorescens* SBW25); and Ravel & Cornelis (2003) (*P. putida* KT2440). Binding pocket residues for substituted A domains are shown in Table A-2 and were predicted using the PKS/NRPS analysis tools available at <http://nrps.igs.umaryland.edu/nrps/> (Bachmann and Ravel, 2009).

Table A-1. Accession numbers of genes used for domain substitution

Accession Number	Gene Name	Organism*
NP_251092	<i>pvdI</i>	<i>P. aeruginosa</i> PAO1
NP_251090	<i>pvdJ</i>	<i>P. aeruginosa</i> PAO1
NP_251089	<i>pvdD</i>	<i>P. aeruginosa</i> PAO1
YP_274152.1	<i>pspph1924</i>	<i>Pseudomonas syringae</i> pv. phaseolicola 1448A
YP_274153.1	<i>pspph1925</i>	<i>Pseudomonas syringae</i> pv. phaseolicola 1448A
YP_274154.1	<i>pspph1926</i>	<i>Pseudomonas syringae</i> pv. phaseolicola 1448A
NP_746337.1	<i>pp4220</i>	<i>Pseudomonas putida</i> KT2440
NP_746336.1	<i>pp4219</i>	<i>Pseudomonas putida</i> KT2440
YP_002872133.1	<i>pvdI</i>	<i>Pseudomonas fluorescens</i> SBW25
YP_002872133.1	<i>pvdJ</i>	<i>Pseudomonas fluorescens</i> SBW25

Table A-2. Residues lining the A domain binding pockets and substrate specificity predictions

Module	1	2	3	4	5	6	7	8	Prediction
Thr-WT	D	F	W	N	I	G	M	V	Thr
Thr-A	D	F	W	N	I	G	M	V	Thr
Thr-D	D	F	W	N	I	G	M	V	Thr
Thr-E	D	F	W	N	I	G	M	V	Thr
Ser-B	D	V	W	H	V	S	L	I	Ser
Ser-F	D	V	W	H	V	S	L	I	Ser
Ser-I	D	V	W	H	V	S	L	I	Ser
Lys-C	D	G	E	D	I	I	T	V	Lys
Asp-G	D	L	T	K	V	G	H	V	Asp
Gly-H	D	I	L	X <sup>1</sup>	I	G	L	I	Gly
Orn-J	D	G	E	V	C	G	G	V	Orn

<sup>1</sup> The Gly-H A domain had an insertion at this point so that this residue did not align to the subject A domain.

## Appendix B. Optimisation of liquid media based assay for measuring pyoverdine levels

To determine whether pyoverdine could be quantified through absorbance or fluorescence, strains were first grown in liquid media for 24 h and absorbance and fluorescence measured (Figure B-1). The background for fluorescence measurements was relatively low and differences in measurements between the culture and supernatant were small. In contrast, there was a high background in absorbance. As a result, measurements of the cell culture were increased and more variable than the supernatant. This increase was also observed for the *pvdD* deletion strain, indicating that it was due to the presence of cells rather than increased pyoverdine within cells.

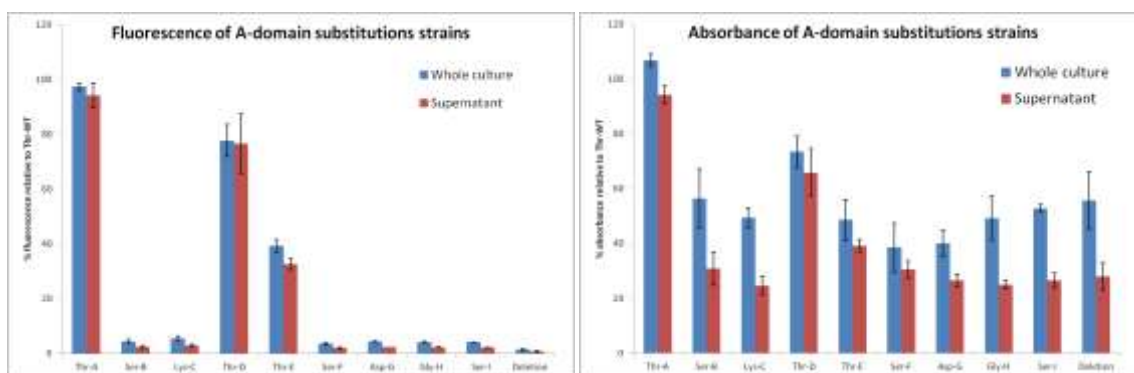


Figure B-1. Fluorescence (left) and absorbance (right) of the culture and supernatant of A domain substitution strains relative to the Thr-WT control strain. Values zeroed to a media control. Data are the mean of 3 independent replicates, and error bars indicate 1 standard deviation.

To assess the quantification ability of fluorescence versus absorbance, a dilution series was made using the supernatant from the Thr-WT strain (Figure B-2). When dilution, absorbance followed a linear trend, whereas fluorescence became saturated. Thus, absorbance was more indicative of pyoverdine levels than fluorescence. However, fluorescence enabled detection of pyoverdine at much lower levels. As such, absorbance and fluorescence were both used in this research.

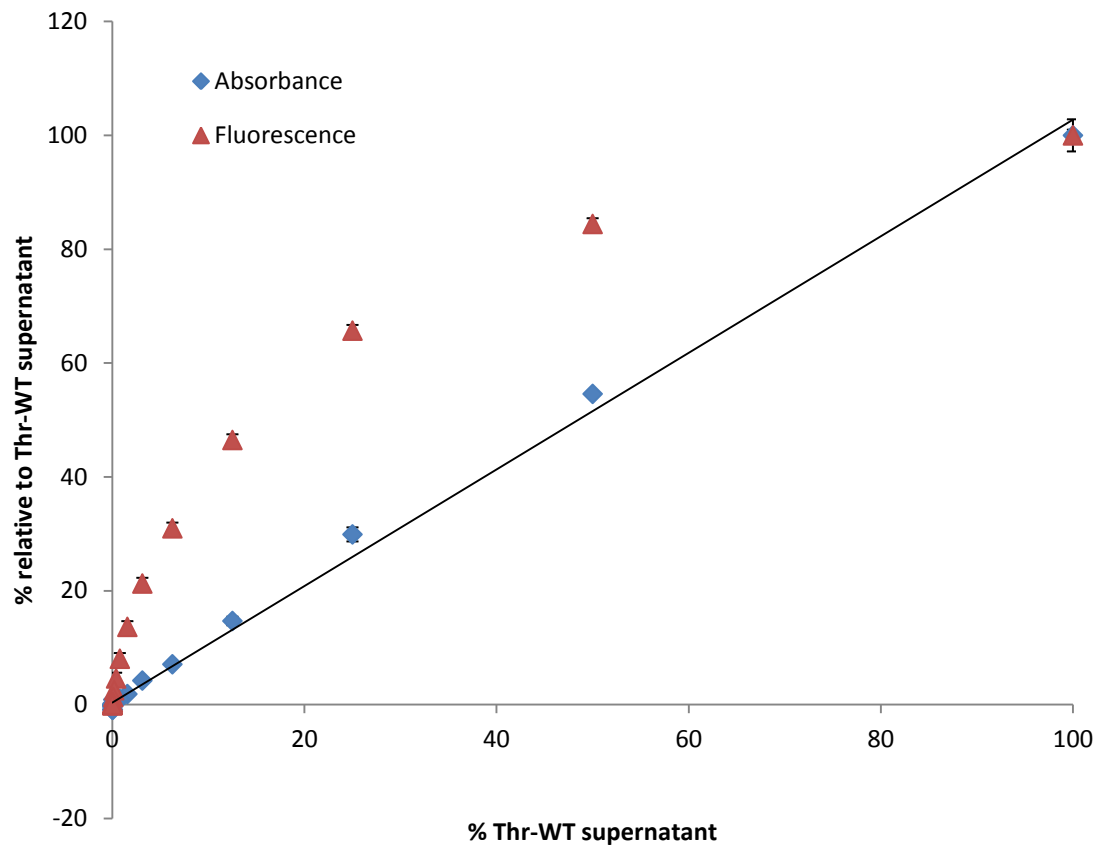


Figure B-2. Absorbance and fluorescence taken from a 2x serial dilution of the supernatant from strain Thr-WT. Line of best fit shown for absorbance measurements. Data are a mean of a triplicate dilution series and error bars represent 1 standard deviation.

### Appendix C. Alignments of T domains from **non-*Pseudomonas* strains**

Sequences of T domains for alignment were taken from the NRPS-PKS database (Ansari et al., 2004) and the enzyme accession numbers are shown in Table C-3. Criteria for choosing the T domains to include were that there should be only bacterial T domains (excluding *Pseudomonas*) and NRPS modules are colinear, or colinear and iterative with the product, i.e. type A or type B pathways. Boundaries for T domains were taken from the PKS/NRPS analysis website used to predict substrate specificity and domain arrangement of NRPS enzymes (Bachmann and Ravel, 2009). Sequences were aligned as per Section 2.10.2 and full alignments are shown in Figure C-3.

Table C-3: Accession numbers of enzymes from which T domain sequences were sourced.

Accession Number	Gene Name	Organism
ADG27358.1	ACMSII	<i>Streptomyces anulatus</i>
ADG27359.1	ACMSIII	<i>Streptomyces anulatus</i>
AAD56240.1	DhbF	<i>Bacillus subtilis</i>
O68006.1	BA1	<i>Bacillus licheniformis</i>
O68007.1	BA2	<i>Bacillus licheniformis</i>
O68008.1	BA3	<i>Bacillus licheniformis</i>
NP_627443.1	CDA1	<i>Streptomyces coelicolor</i> A3(2)
NP_627444.1	CDA2	<i>Streptomyces coelicolor</i> A3(2)
CAD55498.1	CDA3	<i>Streptomyces coelicolor</i> A3(2)
AAB80955.2	fenA	<i>Bacillus subtilis</i>
AAB00093.1	fenB	<i>Bacillus subtilis</i>
AAC36721	fenC	<i>Bacillus subtilis</i>
CAA09819	fenD	<i>Bacillus subtilis</i>
AAB80956	fenE	<i>Bacillus subtilis</i>
BAA00406.1	GrsA	<i>Brevibacillus brevis</i>
AAD04757.1	licA	<i>Bacillus licheniformis</i>
AAD04758.1	licB	<i>Bacillus licheniformis</i>
AAD04759.1	licC	<i>Bacillus licheniformis</i>
CAA49816.1	srfA1	<i>Bacillus subtilis</i>
CAA49817.1	srfA2	<i>Bacillus subtilis</i>
CAA49818.1	srfA3	<i>Bacillus subtilis</i>
AAC45928.1	TycA	<i>Brevibacillus brevis</i>
AAC45929.1	TycB	<i>Brevibacillus brevis</i>
AAC45930.1	TycC	<i>Brevibacillus brevis</i>
Q54959	Pris2	<i>Streptomyces pristinaespiralis</i>
O07944	Pris3	<i>Streptomyces pristinaespiralis</i>
CAC01603.1	apdA	<i>Anabaena</i> sp. 90
CAC01604.1	apdB	<i>Anabaena</i> sp. 90
CAC01606.1	apdD	<i>Anabaena</i> sp. 90
BAB69698.1	ItuA	<i>Bacillus subtilis</i>
BAB69699.1	ItuB	<i>Bacillus subtilis</i>
BAB69700.1	ItuC	<i>Bacillus subtilis</i>
BAA83992.1	McyA	<i>Microcystis aeruginosa</i>
BAA83993.1	McyB	<i>Microcystis aeruginosa</i>
BAA83994.1	McyC	<i>Microcystis aeruginosa</i>
AAF08795.1	Myc1	<i>Bacillus subtilis</i> subsp. spizizenii ATCC 6633
AAF08796.1	Myc2	<i>Bacillus subtilis</i> subsp. spizizenii ATCC 6633
AAF08797.1	Myc3	<i>Bacillus subtilis</i> subsp. spizizenii ATCC 6633
AAO64403.1	NdaA	<i>Nodularia spumigena</i>
AAO64402.1	NdaB	<i>Nodularia spumigena</i>



Figure C-3. ClustalW sequence alignments of T domains from non-*Pseudomonas* NRPS enzymes. Genes are labelled first by a letter representing the type of downstream domain, followed by the gene name according to Table C-3, and finally the module number from which the T domain was sourced.



## Appendix D. Comparative modelling quality assessment

The quality of 52 homology models of the second C domain of PvdD was assessed using QMEAN6 and QMEANclust. The models are ranked according to the QMEAN6 Z-score in Table D-4, in which a higher Z-score means the model is more similar to high quality crystal structures (Benkert et al., 2011). In Table D-5, the models are ranked according to the QMEANclust score. The model selected for use in this study was RaptorXsa, which was ranked third for QMEAN6 and ninth for QMEANclust.

Table D-4. Ranking of comparative structural models of the Thr-WT C domain by QMEAN6.

QMEAN6		
Model	Z-score	Rank
Modeller_LoopRebuild.pdb	-0.08	1
Modeller_2JGP_1L5A_2XHG.pdb	-0.21	2
RaptorXmsa.pdb	-0.24	3
chunk-TASSER_2JGP.pdb	-0.52	4
Esypred_2jgpa.pdb	-0.63	5
muFold1_2JGP.pdb	-0.66	6
Modeller_2JGP_1L5A.pdb	-0.66	7
M4T_2JGP.pdb	-0.71	8
Modeller_2JGP.pdb	-0.72	9
SparkX_2jpa.pdb	-0.75	10
Modeller_2JGP_2XHG.pdb	-0.83	11
chunk-TASSER_1L5A.pdb	-0.83	12
chunk-TASSER_2XHG.pdb	-0.84	13
RaptorX_2JGP.pdb	-0.86	14
nFOLD_2JGP.pdb	-0.93	15
Modeller_2JGP_1Q9J_2XHG.pdb	-1.01	16
Multicom_msa.pdb	-1.1	17
Raptor2_2JGP.pdb	-1.14	18
TASSER_VMT_1L5A.pdb	-1.18	19
Phyre2_2vsqA.pdb	-1.19	20
SwissModel_2JGP.pdb	-1.2	21
muFold3_2VSQ.pdb	-1.22	22
chunk-TASSER_2VSQ.pdb	-1.22	23
muFold4_2VSQ.pdb	-1.24	24
TASSER_VMT_2XHG.pdb	-1.29	25
muFold2_2JGP.pdb	-1.3	26
popul4_2JGP.pdb	-1.31	27
popul3_2JGP.pdb	-1.36	28
TASSER_VMT_2JGP.pdb	-1.37	29

popul5_2JGP.pdb	-1.37	30
pro_sp3_TASSER_msa5.pdb	-1.43	31
popul1_2JGP.pdb	-1.45	32
RaptorX_2vsqA.pdb	-1.54	33
I_TASSER_2JGP.pdb	-1.6	34
popul2_2JGP.pdb	-1.6	35
pro_sp3_TASSER_msa1.pdb	-1.6	36
muFold5_2XHG.pdb	-1.66	37
pro_sp3_TASSER_msa4.pdb	-1.69	38
pro_sp3_TASSER_msa3.pdb	-1.71	39
TASSER_VMT_2VSQ.pdb	-1.75	40
CPHmodel_2JGP.pdb	-1.75	41
3Djigsaw_2JGP.pdb	-1.8	42
SparkX_115a.pdb	-1.81	43
RaptorX_115a.pdb	-1.83	44
Phyre2_2jgpA.pdb	-1.87	45
pro_sp3_TASSER_msa2.pdb	-2	46
RaptorX_2xhgA.pdb	-2.53	47
Phyre2_2xhgA.pdb	-2.84	48
Modeller_2XHG.pdb	-2.95	49
Modeller_1L5A.pdb	-3.17	50
FUG_2jgpa.pdb	-3.23	51
FUG_115a.pdb	-3.95	52

Table D-5. Ranking of comparative structural models of the Thr-WT C domain by QMEANclust.

QMEANclust		
Model	QMEANclust	Rank
muFold1_2JGP.pdb	0.871	1
muFold2_2JGP.pdb	0.87	2
muFold3_2VSQ.pdb	0.87	3
Multicom_msa.pdb	0.867	4
muFold4_2VSQ.pdb	0.866	5
pro_sp3_TASSER_msa4.pdb	0.865	6
M4T_2JGP.pdb	0.863	7
nFOLD_2JGP.pdb	0.862	8
RaptorXmsa.pdb	0.862	9
I_TASSER_2JGP.pdb	0.862	10
pro_sp3_TASSER_msa3.pdb	0.861	11
TASSER_VMT_1L5A.pdb	0.861	12
pro_sp3_TASSER_msa2.pdb	0.861	13
pro_sp3_TASSER_msa1.pdb	0.861	14
TASSER_VMT_2VSQ.pdb	0.859	15
Esypred_2jgpa.pdb	0.859	16
pro_sp3_TASSER_msa5.pdb	0.858	17
chunk-TASSER_2VSQ.pdb	0.856	18
Raptor2_2JGP.pdb	0.856	19
FUG_2jgpa.pdb	0.855	20
SparkX_2jpa.pdb	0.853	21
muFold5_2XHG.pdb	0.853	22
Modeller_2JGP.pdb	0.851	23
chunk-TASSER_2XHG.pdb	0.851	24
SwissModel_2JGP.pdb	0.851	25
RaptorX_2JGP.pdb	0.851	26
Modeller_2JGP_1L5A_2XHG.pdb	0.85	27
Modeller_LoopRebuild.pdb	0.85	28
Modeller_2JGP_1L5A.pdb	0.846	29
chunk-TASSER_2JGP.pdb	0.846	30
popul3_2JGP.pdb	0.839	31
Modeller_2JGP_2XHG.pdb	0.837	32
popul4_2JGP.pdb	0.834	33
popul5_2JGP.pdb	0.834	34
chunk-TASSER_1L5A.pdb	0.833	35
Phyre2_2jgpa.pdb	0.831	36
TASSER_VMT_2JGP.pdb	0.831	37
TASSER_VMT_2XHG.pdb	0.829	38
Modeller_2JGP_1Q9J_2XHG.pdb	0.824	39
CPHmodel_2JGP.pdb	0.823	40
RaptorX_2vsqa.pdb	0.668	41

popul2_2JGP.pdb	0.648	42
Phyre2_2vsqA.pdb	0.648	43
popul1_2JGP.pdb	0.645	44
3Djigsaw_2JGP.pdb	0.607	45
RaptorX_2xhgA.pdb	0.541	46
Phyre2_2xhgA.pdb	0.534	47
RaptorX_115a.pdb	0.512	48
SparkX_115a.pdb	0.508	49
Modeller_2XHG.pdb	0.5	50
FUG_115a.pdb	0.489	51
Modeller_1L5A.pdb	0.442	52

## Appendix E. Pyoverdine production by strains containing the directed evolution construct

The pUCP22 based plasmid created in Chapter 6 was tested to determine the concentration of L-arabinose to use for induction. This was first tested by measuring the fluorescence of the restriction site control strain containing the C domain ligated back into pDEC in comparison to the wild type and *pvdD* deletion strains (Figure E-4). This showed increased pyoverdine production compared to the wild type strain at low concentrations of L-arabinose, and similar levels at a concentration of 0.1 % (w/v) L-arabinose. When testing the fluorescence of other strains from Chapter 6 (Figure E-5), there was increased fluorescence for the strain J12-D3 that showed low levels of fluorescence. This increased fluorescence was greatest at 0.1 % (w/v) L-arabinose. Consequently, it was decided to use this concentration of L-arabinose for induction of pyoverdine when using this plasmid.

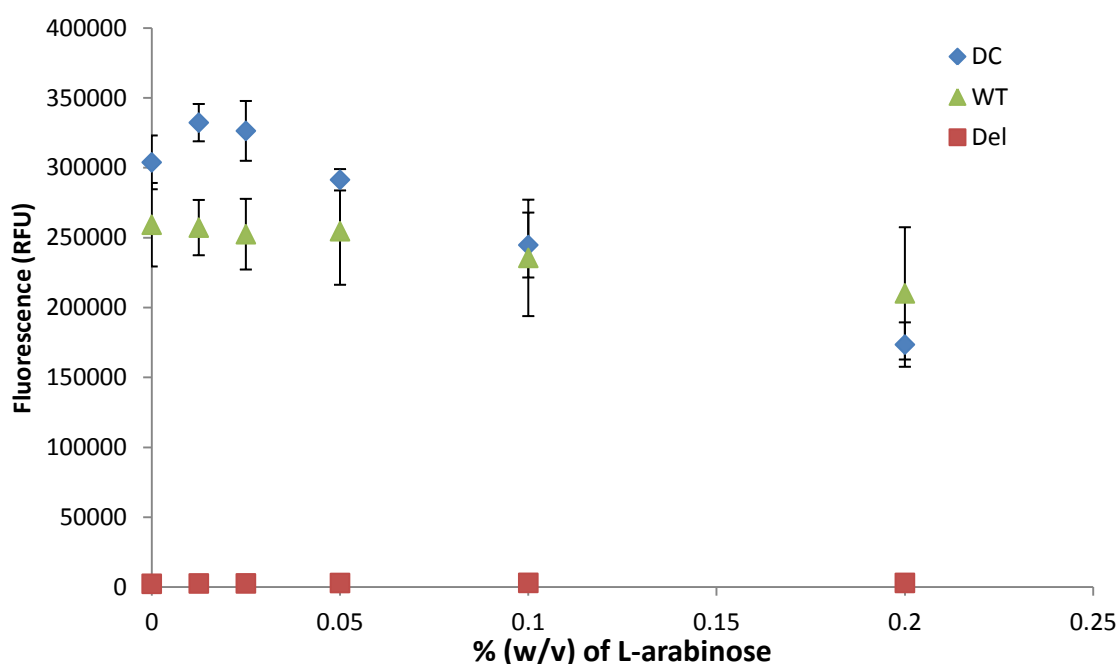


Figure E-4. L-arabinose induction of pyoverdine synthesis for the pDEC restriction site control strain. Pyoverdine production measured using fluorescence. DC refers to the restriction site control strain, WT refers to the *P. aeruginosa* PAO1 strain, and Del to the *pvdD* deletion strain. Data are the mean of 6 independent replicates, and error bars indicate 1 standard deviation

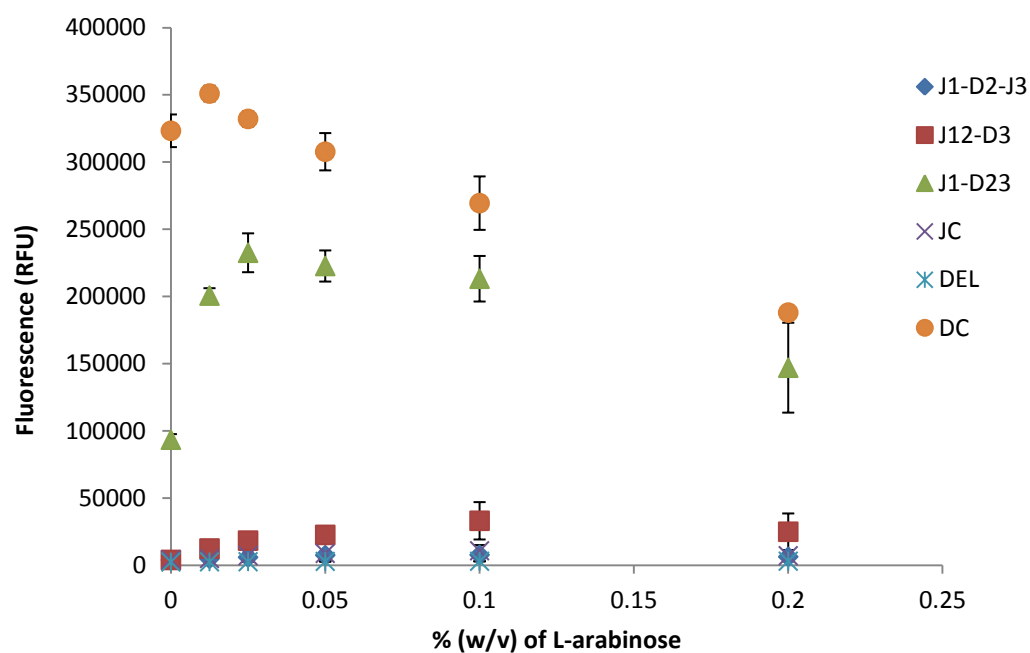


Figure E-5. L-arabinose induction of pyoverdine synthesis for the chimeric C domain strains from Chapter 6. Pyoverdine production measured using fluorescence. Nomenclature for DC<sub>Thr</sub> and Del is noted in Figure E-4 and nomenclature for remaining strains in Figure 6-6 . Data are the mean of 3 independent replicates, and error bars indicate 1 standard deviation

## References

- Ackerley, D.F., and Lamont, I.L. (2004). Characterization and Genetic Manipulation of Peptide Synthetases in *Pseudomonas aeruginosa* PAO1 in Order to Generate Novel Pyoverdines. *Chem Biol.* 11, 971–980.
- Ackerley, D.F., Caradoc-Davies, T.T., and Lamont, I.L. (2003). Substrate Specificity of the Nonribosomal Peptide Synthetase PvdD from *Pseudomonas aeruginosa*. *J Bacteriol.* 185, 2848–2855.
- Albrethsen, J. (2011). The first decade of MALDI protein profiling: A lesson in translational biomarker research. *J Proteomics.* 74, 765–773.
- Ansari, M.Z., Yadav, G., Gokhale, R.S., and Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* 32, W405–W413.
- Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics.* 22, 195–201.
- Bachmann, B.O., and Ravel, J. (2009). Chapter 8 Methods for In Silico Prediction of Microbial Polyketide and Nonribosomal Peptide Biosynthetic Pathways from DNA Sequence Data. In *Methods in Enzymology*, David A. Hopwood, ed. (Academic Press), pp. 181–217.
- Baltz, R.H. (2012). Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.*
- Bates, P.A., Kelley, L.A., MacCallum, R.M., and Sternberg, M.J.E. (2001). Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins.* 45, 39–46.
- Baynham, P.J., Ramsey, D.M., Gvozdyev, B.V., Cordonnier, E.M., and Wozniak, D.J. (2006). The *Pseudomonas aeruginosa* Ribbon-Helix-Helix DNA-Binding Protein AlgZ (AmrZ) Controls Twitching Motility and Biogenesis of Type IV Pili. *J Bacteriol.* 188, 132–140.
- Beavis, R.C., Chaudhary, T., and Chait, B.T. (1992).  $\alpha$ -Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted laser desorption mass spectrometry. *Organic Mass Spectrometry.* 27, 156–158.
- Beld, J., Sonnenschein, E.C., Vickery, C.R., Noel, J.P., and Burkart, M.D. (2013). The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* 31, 61–108.
- Belshaw, P.J., Walsh, C.T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science.* 284, 486–489.
- Benkert, P., Tosatto, S.C.E., and Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins.* 71, 261–277.

Benkert, P., Künzli, M., and Schwede, T. (2009a). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37, W510–514.

Benkert, P., Schwede, T., and Tosatto, S.C. (2009b). QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. *BMC Struct. Biol.* 9, 35.

Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 27, 343–350.

Bergendahl, V., Linne, U., and Marahiel, M.A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur J Biochem.* 269, 620–629.

Bloudoff, K., Rodionov, D., and Schmeing, T.M. (2013). Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J Mol Biol.* 425, 3137–3150.

Briskot, G., Taraz, K., and Budzikiewicz, H. (1989). Bacterial Constituents, XXXVII. Pyoverdinin-Type Siderophores from *Pseudomonas aeruginosa*. *Liebigs Annalen Der Chemie* 1989, 375–384.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T., and Stubbs, M.T. (2002). Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE. *Structure.* 10, 301–310.

Budzikiewicz, H., Schäfer, M., Fernández, D.U., Matthijs, S., and Cornelis, P. (2007). Characterization of the chromophores of pyoverdins and related siderophores by electrospray tandem mass spectrometry. *Biometals.* 20, 135–144.

Bush, K. (2012). Improving known classes of antibiotics: an optimistic approach for the future. *Curr Opin Pharmacol.* 12, 527–534.

Butler, M.S. (2005). Natural products to drugs: natural product derived compounds in clinical trials. *Nat Prod Rep.* 22, 162.

Buyer, J.S., and Leong, J. (1986). Iron transport-mediated antagonism between plant growth-promoting and plant-deleterious *Pseudomonas* strains. *J. Biol. Chem.* 261, 791–794.

Caboche, S., Leclère, V., Pupin, M., Kuchеров, G., and Jacques, P. (2010). Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. *J Bacteriol.* 192, 5143–5150.

Cavasotto, C.N., and Phatak, S.S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discov Today.* 14, 676–683.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol.* 7, 211–224.

Chen, C.-Y., Georgiev, I., Anderson, A.C., and Donald, B.R. (2009). Computational structure-based redesign of enzyme activity. *PNAS.*

- Chiocchini, C., Linne, U., and Stachelhaus, T. (2006). *In Vivo* Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex. *Chem Biol.* 13, 899–908.
- Clugston, S.L., Sieber, S.A., Marahiel, M.A., and Walsh, C.T. (2003). Chirality of Peptide Bond-Forming Condensation Domains in Nonribosomal Peptide Synthetases: The C5 Domain of Tyrocidine Synthetase Is a  $^D$ C<sub>L</sub> Catalyst†. *Biochemistry.* 42, 12095–12104.
- Conti, E., Franks, N.P., and Brick, P. (1996). Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure.* 4, 287–298.
- Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* 16, 4174–4183.
- Cornelis, P. (2010). Iron uptake and metabolism in pseudomonads. *Appl Microbiol Biotechnol.* 86, 1637–1645.
- Cozzetto, D., Kryshtafovych, A., and Tramontano, A. (2009). Evaluation of CASP8 model quality predictions. *Proteins.* 77, 157–166.
- De Crecy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechon, A., Bamas-Jacques, N., Crouzet, J., and Thibaut, D. (1997). Pristinamycin I biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural peptide synthetase genes. *J Bacteriol.* 179, 705–713.
- De Crécy-Lagard, V., Marlière, P., and Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C. R. Acad. Sci. III.* 318, 927–936.
- Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A Sequence Logo Generator. *Genome Res.* 14, 1188–1190.
- Cude, W.N., Mooney, J., Tavanaei, A.A., Hadden, M.K., Frank, A.M., Gulvik, C.A., May, A.L., and Buchan, A. (2012). Production of the Antimicrobial Secondary Metabolite Indigoidine Contributes to Competitive Surface Colonization by the Marine *Roseobacter Phaeobacter* sp. Strain Y4I. *Appl. Environ. Microbiol.* 78, 4771–4780.
- Demange, P., Wendenbaum, S., Linget, C., Mertz, C., Cung, M.T., Dell, A., and Abdallah, M.A. (1990). Bacterial siderophores: structure and NMR assignment of pyoverdins Pa, siderophores of *Pseudomonas aeruginosa* ATCC 15692. *Biol Metals.* 3, 155–170.
- Dieckmann, R., Lee, Y.-O., van Liempt, H., von Döhren, H., and Kleinkauf, H. (1995). Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases. *FEBS Lett.* 357, 212–216.
- Dietrich, J.A., McKee, A.E., and Keasling, J.D. (2010). High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu Rev Biochem.* 79, 563–590.
- Doekel, S., and Marahiel, M.A. (2000). Dipeptide formation on engineered hybrid peptide synthetases. *Chem Biol.* 7, 373–384.

- Doekel, S., Gal, M.F.C.L., Gu, J.Q., Chu, M., Baltz, R.H., and Brian, P. (2008). Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology*. 154, 2872–2880.
- Von Döhren, H., Keller, U., Vater, J., and Zocher, R. (1997). Multifunctional Peptide Synthetases. *Chem. Rev.* 97, 2675–2706.
- Von Döhren, H., Dieckmann, R., and Pavela-Vrancic, M. (1999). The nonribosomal code. *Chem Biol.* 6, R273–R279.
- Drake, E.J., and Gulick, A.M. (2011). Structural characterization and high throughput screening of inhibitors of PvdQ, an NTN hydrolase involved in pyoverdine synthesis. *ACS Chem Biol.* 6, 1277–1286.
- Du, L., and Lou, L. (2010). PKS and NRPS release mechanisms. *Nat Prod Rep.* 27, 255.
- Duerfahrt, T., Doekel, S., Sonke, T., Quaedflieg, P.J.L.M., and Marahiel, M.A. (2003). Construction of hybrid peptide synthetases for the production of  $\alpha$ -l-aspartyl-l-phenylalanine, a precursor for the high-intensity sweetener aspartame. *Eur J Biochem.* 270, 4555–4563.
- Ehmann, D.E., Trauger, J.W., Stachelhaus, T., and Walsh, C.T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem Biol.* 7, 765–772.
- Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002). Exploitation of the Selectivity-Confering Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics†. *Biochemistry*. 41, 9718–9726.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M. s., Eramian, D., Shen, M., Pieper, U., and Sali, A. (2001). Comparative Protein Structure Modeling Using MODELLER. In *Curr Protoc Protein Sci.*, (John Wiley & Sons, Inc.).
- Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011). Directed Evolution of the Nonribosomal Peptide Synthetase AdmK Generates New Andrimid Derivatives *In Vivo*. *Chem Biol.* 18, 601–607.
- Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008). Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol Pharm.* 5, 191–211.
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*. 39, 783–791.
- Ferra, F. de, Rodriguez, F., Tortora, O., Tosi, C., and Grandi, G. (1997). Engineering of Peptide Synthetases KEY ROLE OF THE THIOESTERASE-LIKE DOMAIN FOR EFFICIENT PRODUCTION OF RECOMBINANT PEPTIDES. *J. Biol. Chem.* 272, 25304–25309.
- Fischbach, M.A., Lai, J.R., Roche, E.D., Walsh, C.T., and Liu, D.R. (2007). Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *PNAS*. 104, 11951–11956.

- Fischbach, M.A., Walsh, C.T., and Clardy, J. (2008). The evolution of gene collectives: How natural selection drives chemical innovation. *PNAS*. 105, 4601–4608.
- Frueh, D.P., Arthanari, H., Koglin, A., Vosburg, D.A., Bennett, A.E., Walsh, C.T., and Wagner, G. (2008). Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature*. 454, 903–906.
- Fuchs, R., and Budzikiewicz, H. (2000). Structural studies of pyoverdins with cyclopeptidic substructures by electrospray ionization and collision induced fragmentation 1,2. *Spectroscopy*. 14, 229–246.
- Geoffroy, V.A., Fetherston, J.D., and Perry, R.D. (2000). *Yersinia pestis* YbtU and YbtT Are Involved in Synthesis of the Siderophore Yersiniabactin but Have Different Effects on Regulation. *Infect Immun*. 68, 4452–4461.
- Gevers, W., Kleinkauf, H., and Lipmann, F. (1968). The activation of amino acids for biosynthesis of gramicidin S. *PNAS*. 60, 269–276.
- Gevers, W., Kleinkauf, H., and Lipmann, F. (1969). Peptidyl transfers in gramicidin S biosynthesis from enzyme-bound thioester intermediates. *PNAS*. 63, 1335–1342.
- Giessen, T.W., and Marahiel, M.A. (2012). Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett*. 586, 2065–2075.
- Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003). 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*. 19, 1015–1018.
- Guex, N., and Peitsch, M.. (1996). Swiss-PdbViewer: A Fast and Easy-to-use PDB Viewer for Macintosh and PC. *Protein Data Bank Quarterly Newsletter*. 77, 7.
- Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 18, 2714–2723.
- Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter. *J Bacteriol*. 177, 4121–4130.
- Haese, A., Pieper, R., von Ostrowski, T., and Zocher, R. (1994). Bacterial expression of catalytically active fragments of the multifunctional enzyme enniatin synthetase. *J. Mol. Biol*. 243, 116–122.
- Hahn, M., and Stachelhaus, T. (2004). Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *PNAS*. 101, 15585–15590.
- Hahn, M., and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *PNAS*. 103, 275–280.
- Han, J.W., Kim, E.Y., Lee, J.M., Kim, Y.S., Bang, E., and Kim, B.S. (2012). Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnol Lett*. 34, 1327–1334.

- Hannauer, M., Yeterian, E., Martin, L.W., Lamont, I.L., and Schalk, I.J. (2010). An efflux pump is involved in secretion of newly synthesized siderophore by *Pseudomonas aeruginosa*. FEBS Lett. 584, 4751–4755.
- Hannauer, M., Schäfer, M., Hoegy, F., Gizzi, P., Wehrung, P., Mislin, G.L.A., Budzikiewicz, H., and Schalk, I.J. (2012). Biosynthesis of the pyoverdine siderophore of *Pseudomonas aeruginosa* involves precursors with a myristic or a myristoleic acid chain. FEBS Lett. 586, 96–101.
- Hider, R.C., and Kong, X. (2010). Chemistry and biology of siderophores. Nat. Prod. Rep. 27, 637–657.
- Hillisch, A., Pineda, L.F., and Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. Drug Discov Today. 9, 659–669.
- Hoang, T.T., Kutchma, A.J., Becher, A., and Schweizer, H.P. (2000). Integration-Proficient Plasmids for *Pseudomonas aeruginosa*: Site-Specific Integration and Use for Engineering of Reporter and Expression Strains. Plasmid. 43, 59–72.
- Hur, G.H., Vickery, C.R., and Burkart, M.D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. Nat. Prod. Rep. 29, 1074–1098.
- Joardar, V., Lindeberg, M., Jackson, R.W., Selengut, J., Dodson, R., Brinkac, L.M., Daugherty, S.C., DeBoy, R., Durkin, A.S., Giglio, M.G., et al. (2005). Whole-Genome Sequence Analysis of *Pseudomonas syringae* pv. phaseolicola 1448A Reveals Divergence among Pathovars in Genes Involved in Virulence and Transposition. J Bacteriol. 187, 6488–6498.
- Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8, 275–282.
- Jones, D.T., Bryson, K., Coleman, A., McGuffin, L.J., Sadowski, M.I., Sodhi, J.S., and Ward, J.J. (2005). Prediction of novel and analogous folds using fragment assembly and fold recognition. Proteins. 61, 143–151.
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. Nat. Protocols. 7, 1511–1522.
- Keating, T.A., and Walsh, C.T. (1999). Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. Curr Opin Chem Biol. 598–606.
- Keating, T.A., Marshall, C.G., Walsh, C.T., and Keating, A.E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat Struct Mol Biol. 9, 522–526.
- Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protocols. 4, 363–371.
- King, E.O., Ward, M.K., and Raney, D.E. (1954). Two simple media for the demonstration of pyocyanin and fluorescein. J. Lab. Clin. Med. 44, 301–307.

- Kirschning, A., and Hahn, F. (2012). Merging Chemical Synthesis and Biosynthesis: A New Chapter in the Total Synthesis of Natural Products and Natural Product Libraries. *Angewandte Chemie International Edition*. 51, 4012–4022.
- Kleinkauf, H., Gevers, W., and Lipmann, F. (1969). Interrelation between activation and polymerization in gramicidin S biosynthesis. *PNAS. U S A* 62, 226–233.
- Koglin, A., and Walsh, C.T. (2009). Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat Prod Rep*. 26, 987.
- Koglin, A., Mofid, M.R., Löhr, F., Schäfer, B., Rogov, V.V., Blum, M.-M., Mittag, T., Marahiel, M.A., Bernhard, F., and Dötsch, V. (2006). Conformational Switches Modulate Protein Interactions in Peptide Antibiotic Synthetases. *Science*. 312, 273–276.
- Konz, D., Doekel, S., and Marahiel, M.A. (1999). Molecular and Biochemical Characterization of the Protein Template Controlling Biosynthesis of the Lipopeptide Lichenysin. *J Bacteriol*. 181, 133–140.
- Kraas, F.I., Giessen, T.W., and Marahiel, M.A. (2012). Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett*. 586, 283–288.
- Kryshtafovych, A., Fidelis, K., and Tramontano, A. (2011). Evaluation of model quality predictions in CASP9. *Proteins*. 79, 91–106.
- Lai, J.R., Fischbach, M.A., Liu, D.R., and Walsh, C.T. (2006). A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. *PNAS*. 103, 5314–5319.
- Lambalot, R.H., Gehring, A.M., Flugel, R.S., Zuber, P., LaCelle, M., Marahiel, M.A., Reid, R., Khosla, C., and Walsh, C.T. (1996). A new enzyme superfamily — the phosphopantetheinyl transferases. *Chem Biol*. 3, 923–936.
- Lambert, C., Léonard, N., Bolle, X.D., and Depiereux, E. (2002). ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics*. 18, 1250–1256.
- Larsson, P., Wallner, B., Lindahl, E., and Elofsson, A. (2008). Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci*. 17, 990–1002.
- Lawen, A., and Traber, R. (1993). Substrate specificities of cyclosporin synthetase and peptolide SDZ 214-103 synthetase. Comparison of the substrate specificities of the related multifunctional polypeptides. *J. Biol. Chem*. 268, 20452–20465.
- Leemhuis, H., Kelly, R.M., and Dijkhuizen, L. (2009). Directed evolution of enzymes: Library screening strategies. *IUBMB Life*. 61, 222–228.
- Lesk, A.M., and Chothia, C.H. (1986). The Response of Protein Structures to Amino-Acid Sequence Changes. *Phil. Trans. R. Soc. Lond. A* 317, 345–356.
- Linne, U., and Marahiel, M.A. (2000). Control of Directionality in Nonribosomal Peptide Synthesis: Role of the Condensation Domain in Preventing Misinitiation and Timing of Epimerization†. *Biochemistry*. 39, 10439–10447.

- Linne, U., Doekel, S., and Marahiel, M.A. (2001). Portability of Epimerization Domain and Role of Peptidyl Carrier Protein on Epimerization Activity in Nonribosomal Peptide Synthetases†. *Biochemistry*. 40, 15824–15834.
- Lipmann, F., Roskoski, R., Kleinkauf, H., and Gevers, W. (1970). Isolation of enzyme-bound peptide intermediates in tyrocidine biosynthesis. *Biochemistry*. 9, 4846–4851.
- Liu, Y., Zheng, T., and Bruner, S.D. (2011). Structural Basis for Phosphopantetheinyl Carrier Domain Interactions in the Terminal Module of Nonribosomal Peptide Synthetases. *Chem Biol*. 18, 1482–1488.
- Marahiel, M.A. (2009). Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *Journal of Peptide Science*. 15, 799–807.
- Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev*. 97, 2651–2674.
- May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. (2002). Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *PNAS*. 99, 12120–12125.
- McMorran, B.J., Kumara, H.M.C.S., Sullivan, K., and Lamont, I.L. (2001). Involvement of a transformylase enzyme in siderophore synthesis in *Pseudomonas aeruginosa*. *Microbiology*. 147, 1517–1524.
- Mercer, A.C., and Burkart, M.D. (2007). The ubiquitous carrier protein—a window to metabolite biosynthesis. *Nat Prod Rep*. 24, 750.
- Merriman, T.R., Merriman, M.E., and Lamont, I.L. (1995). Nucleotide sequence of pvdD, a pyoverdine biosynthetic gene from *Pseudomonas aeruginosa*: PvdD has similarity to peptide synthetases. *J Bacteriol*. 177, 252–258.
- Meyer, J.-M., Stintzi, A., Vos, D.D., Cornelis, P., Tappe, R., Taraz, K., and Budzikiewicz, H. (1997). Use of Siderophores to Type Pseudomonads: The Three *Pseudomonas Aeruginosa* Pyoverdine Systems. *Microbiology*. 143, 35–43.
- Meyer, J.-M., Stintzi, A., Coulanges, V., Shivaji, S., Voss, J.A., Taraz, K., and Budzikiewicz, H. (1998). Siderotyping of fluorescent pseudomonads: characterization of pyoverdines of *Pseudomonas fluorescens* and *Pseudomonas putida* strains from Antarctica. *Microbiology*. 144, 3119–3126.
- Meyer, J.-M., Stintzi, A., and Poole, K. (1999). The ferripyoverdine receptor FpvA of *Pseudomonas aeruginosa* PAO1 recognizes the ferripyoverdines of *P. aeruginosa* PAO1 and *P. fluorescens* ATCC 13525. *FEMS Microbiol Lett*. 170, 145–150.
- Meyer, J.-M., Geoffroy, V.A., Baysse, C., Cornelis, P., Barelmann, I., Taraz, K., and Budzikiewicz, H. (2002). Siderophore-Mediated Iron Uptake in Fluorescent *Pseudomonas*: Characterization of the Pyoverdine-Receptor Binding Site of Three Cross-Reacting Pyoverdines. *Arch Biochem Biophys*. 397, 179–183.
- Meyer, J.-M., Gruffaz, C., Raharinosy, V., Bezverbnaya, I., Schäfer, M., and Budzikiewicz, H. (2008). Siderotyping of fluorescent *Pseudomonas*: molecular mass determination by mass spectrometry as a powerful pyoverdine siderotyping method. *Biometals*. 21, 259–271.

- Moon, C., Zhang, X.-X., Matthijs, S., Schafer, M., Budzikiewicz, H., and Rainey, P. (2008). Genomic, genetic and structural analysis of pyoverdine-mediated iron acquisition in the plant growth-promoting bacterium *Pseudomonas fluorescens* SBW25. *BMC Microbiol.* 8, 7.
- Mootz, H.D., and Marahiel, M.A. (1997). The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J Bacteriol.* 179, 6843–6850.
- Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2000). Construction of hybrid peptide synthetases by module and domain fusions. *PNAS.* 97, 5848–5853.
- Mootz, H.D., Finking, R., and Marahiel, M.A. (2001). 4'-Phosphopantetheine Transfer in Primary and Secondary Metabolism of *Bacillus subtilis*. *J. Biol. Chem.* 276, 37289–37298.
- Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002). Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *ChemBioChem.* 3, 490–504.
- Morley, K.L., and Kazlauskas, R.J. (2005). Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* 23, 231–237.
- Mossialos, D., Ochsner, U., Baysse, C., Chablain, P., Pirnay, J., Koedam, N., Budzikiewicz, H., Fernández, D.U., Schäfer, M., Ravel, J., et al. (2002). Identification of new, conserved, non-ribosomal peptide synthetases from fluorescent pseudomonads involved in the biosynthesis of the siderophore pyoverdine. *Mol Microbiol.* 45, 1673–1685.
- Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 48, 443–453.
- Nelson, K.E., Weinl, C., Paulsen, I.T., Dodson, R.J., Hilbert, H., Martins dos Santos, V.A.P., Fouts, D.E., Gill, S.R., Pop, M., Holmes, M., et al. (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol.* 4, 799–808.
- Newman, D.J., and Cragg, G.M. (2012). Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* 75, 311–335.
- Nguyen, K.T., Ritz, D., Gu, J.-Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R.H. (2006). Combinatorial biosynthesis of novel antibiotics related to daptomycin. *PNAS.* 103, 17462–17467.
- Nguyen, K.T., He, X., Alexander, D.C., Li, C., Gu, J.-Q., Mascio, C., Van Praagh, A., Mortin, L., Chu, M., Silverman, J.A., et al. (2010). Genetically Engineered Lipopeptide Antibiotics Related to A54145 and Daptomycin with Improved Properties. *Antimicrob Agents Chemother.* 54, 1404–1413.
- Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T.N. (2010). CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.* 38, W576–W581.

O'Connell, K.M.G., Hodgkinson, J.T., Sore, H.F., Welch, M., Salmond, G.P.C., and Spring, D.R. (2013). Combating Multidrug-Resistant Bacteria: Current Strategies for the Discovery of Novel Antibacterials. Angewandte Chemie International Edition.

Offman, M.N., Fitzjohn, P.W., and Bates, P.A. (2006). Developing a move-set for protein model refinement. *Bioinformatics*. 22, 1838–1845.

Owen, J. (2010). Characterisation, manipulation and directed evolution of non-ribosomal peptide synthetase enzymes: A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of doctor of philosophy in biotechnology.

Owen, J.G., and Ackerley, D.F. (2011). Characterization of pyoverdine and achromobactin in *Pseudomonas syringae* pv. phaseolicola 1448a. *BMC Microbiol.* 11, 218.

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

Peypoux, F., Bonmatin, J.-M., Labbé, H., Das, B.C., Ptak, M., and Michel, G. (1991). Isolation and characterization of a new variant of surfactin, the [Val7]surfactin. *Eur J Biochem.* 202, 101–106.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* 33, 5799–5808.

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D.H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol.* 7, 78.

Ravel, J., and Cornelis, P. (2003). Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends Microbiol.* 11, 195–200.

Reverchon, S., Rouanet, C., Expert, D., and Nasser, W. (2002). Characterization of Indigoidine Biosynthetic Genes in *Erwinia chrysanthemi* and Role of This Blue Pigment in Pathogenicity. *J Bacteriol.* 184, 654–665.

Roche, E.D., and Walsh, C.T. (2003). Dissection of the EntF Condensation Domain Boundary and Active Site Residues in Nonribosomal Peptide Synthesis†. *Biochemistry.* 42, 1334–1344.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* 11, 601.

Roongsawang, N., Lim, S.P., Washio, K., Takano, K., Kanaya, S., and Morikawa, M. (2005). Phylogenetic analysis of condensation domains in the nonribosomal peptide synthetases. *FEMS Microbiol Lett.* 252, 143–151.

Rottig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* 39, W362–W367.

- Rykunov, D., Steinberger, E., Madrid-Aliste, C.J., and Fiser, A. (2009). Improved scoring function for comparative modeling using the M4T method. *J Struct Funct Genomics*. 10, 95–99.
- Samel, S.A., Wagner, B., Marahiel, M.A., and Essen, L.-O. (2006). The Thioesterase Domain of the Fengycin Biosynthesis Cluster: A Structural Base for the Macrocyclization of a Non-ribosomal Lipopeptide. *J Mol Biol*. 359, 876–889.
- Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A., and Essen, L.-O. (2007). Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. *Structure*. 15, 781–792.
- Schalk, I.J. (2008). Metal trafficking via siderophores in Gram-negative bacteria: Specificities and characteristics of the pyoverdine pathway. *Journal of Inorganic Biochemistry*. 102, 1159–1169.
- Schalk, I.J., and Guillon, L. (2013). Pyoverdine biosynthesis and secretion in *Pseudomonas aeruginosa*: implications for metal homeostasis. *Environ Microbiol*. 15, 1661–1673.
- Schlumbohm, W., Stein, T., Ullrich, C., Vater, J., Krause, M., Marahiel, M.A., Kruft, V., and Wittmann-Liebold, B. (1991). An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem*. 266, 23135–23141.
- Schneider, A., and Marahiel, M.A. (1998). Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch Microbiol*. 169, 404–410.
- Schneider, A., Stachelhaus, T., and Marahiel, M.A. (1998). Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Mol. Gen. Genet*. 257, 308–318.
- Schwarzer, D., Mootz, H.D., and Marahiel, M.A. (2001). Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases. *Chem Biol*. 8, 997–1010.
- Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *PNAS*. 99, 14083–14088.
- Shaw-Reid, C.A., Kelleher, N.L., Losey, H.C., Gehring, A.M., Berg, C., and Walsh, C.T. (1999). Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization. *Chem Biol*. 6, 385–400.
- Shi, J., Blundell, T.L., and Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*. 310, 243–257.
- Shiau, C.-Y., Baldwin, J.E., Byford, M.F., and Schofield, C.J. (1995).  $\delta$ -l-( $\alpha$ -Aminoadipoyl)-l-cysteinyl-d-valine synthetase: isolation of l-cysteinyl-d-valine, a “shunt”

product, and implications for the order of peptide bond formation. *FEBS Lett.* 373, 303–306.

Shortle, D., Simons, K.T., and Baker, D. (1998). Clustering of low-energy conformations near the native structures of small proteins. *PNAS.* 95, 11158–11162.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics. *Chem. Rev.* 105, 715–738.

Smith, E.E., Sims, E.H., Spencer, D.H., Kaul, R., and Olson, M.V. (2005). Evidence for Diversifying Selection at the Pyoverdine Locus of *Pseudomonas Aeruginosa*. *J Bacteriol.* 187, 2138–2147.

Stachelhaus, T., and Marahiel, M.A. (1995). Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme GrsA. *J. Biol. Chem.* 270, 6163–6169.

Stachelhaus, T., and Walsh, C.T. (2000). Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase†. *Biochemistry.* 39, 5775–5787.

Stachelhaus, T., Schneider, A., and Marahiel, M.A. (1995). Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science.* 269, 69–72.

Stachelhaus, T., Hüser, A., and Marahiel, M.A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem Biol.* 3, 913–921.

Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A. (1998). Peptide Bond Formation in Nonribosomal Peptide Biosynthesis CATALYTIC ROLE OF THE CONDENSATION DOMAIN. *J. Biol. Chem.* 273, 22773–22781.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol.* 6, 493–505.

Stein, D.B., Linne, U., and Marahiel, M.A. (2005). Utility of epimerization domains for the redesign of nonribosomal peptide synthetases. *FEBS Journal.* 272, 4506–4520.

Stein, T., Vater, J., Kruft, V., Wittmann-Liebold, B., Franke, P., Panico, M., Dowell, R.M., and Morris, H.R. (1994). Detection of 4'-phosphopantetheine at the thioester binding site for l-valine of gramicidin S synthetase 2. *FEBS Lett.* 340, 39–44.

Stein, T., Kluge, B., Vater, J., Franke, P., Otto, A., and Wittmann-Liebold, B. (1995). Gramicidin S synthetase 1 (phenylalanine racemase), a prototype of amino acid racemases containing the cofactor 4'-phosphopantetheine. *Biochemistry.* 34, 4633–4642.

Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H.R. (1996). The Multiple Carrier Model of Nonribosomal Peptide Biosynthesis at Modular Multienzymatic Templates. *J. Biol. Chem.* 271, 15428–15435.

- Stevens, B.W., Lilien, R.H., Georgiev, I., Donald, B.R., and Anderson, A.C. (2006). Redesigning the PheA Domain of Gramicidin Synthetase Leads to a New Understanding of the Enzyme's Mechanism and Selectivity†. *Biochemistry*. 45, 15495–15504.
- Stindl, A., and Keller, U. (1994). Epimerization of the D-Valine Portion in the Biosynthesis of Actinomycin D. *Biochemistry* 33, 9358–9364.
- Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*. 406, 959–964.
- Strieker, M., Tanović, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol*. 20, 234–240.
- Suo, Z. (2005). Thioesterase Portability and Peptidyl Carrier Protein Swapping in Yersiniabactin Synthetase from *Yersinia pestis*†. *Biochemistry*. 44, 4926–4938.
- Szájli, E., Fehér, T., and Medzihradszky, K.F. (2008). Investigating the Quantitative Nature of MALDI-TOF MS. *Mol Cell Proteomics*. 7, 2410–2418.
- Takahashi, H., Kumagai, T., Kitani, K., Mori, M., Matoba, Y., and Sugiyama, M. (2007). Cloning and Characterization of a *Streptomyces* Single Module Type Non-ribosomal Peptide Synthetase Catalyzing a Blue Pigment Synthesis. *J Biol Chem*. 282, 9073 –9081.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*. 28, 2731–2739.
- Tanovic, A., Samel, S.A., Essen, L.-O., and Marahiel, M.A. (2008). Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science*. 321, 659 –663.
- Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.-W., Smith, C.P., and Micklefield, J. (2012). Introduction of a Non-Natural Amino Acid into a Nonribosomal Peptide Antibiotic by Modification of Adenylation Domain Specificity. *Angewandte Chemie International Edition*. 51, 7181–7184.
- Trauger, J.W., Kohli, R.M., Mootz, H.D., Marahiel, M.A., and Walsh, C.T. (2000). Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature*. 407, 215–218.
- Tseng, C.C., Bruner, S.D., Kohli, R.M., Marahiel, M.A., Walsh, C.T., and Sieber, S.A. (2002). Characterization of the Surfactin Synthetase C-Terminal Thioesterase Domain as a Cyclic Depsipeptide Synthase†. *Biochemistry*. 41, 13350–13359.
- Turgay, K., Krause, M., and Marahiel, M.A. (1992). Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol*. 6, 529–546.
- Uguru, G.C., Milne, C., Borg, M., Flett, F., Smith, C.P., and Micklefield, J. (2004). Active-Site Modifications of Adenylation Domains Lead to Hydrolysis of Upstream Nonribosomal Peptidyl Thioester Intermediates. *J. Am. Chem. Soc*. 126, 5032–5033.

- Villiers, B., and Hollfelder, F. (2011). Directed Evolution of a Gatekeeper Domain in Nonribosomal Peptide Synthesis. *Chem Biol.* 18, 1290–1299.
- Villiers, B., and Hollfelder, F. (2009). Mapping the Limits of Substrate Specificity of the Adenylation Domain of TycA. *ChemBioChem.* 10, 671–682.
- Visca, P., Imperi, F., and Lamont, I.L. (2007). Pyoverdine siderophores: from biogenesis to biosignificance. *Trends Microbiol.* 15, 22–30.
- Walsh, C. (2003). Where will new antibiotics come from? *Nat Rev Micro.* 1, 65–70.
- Walsh, C.T., and Wencewicz, T.A. (2012). Flavoenzymes: Versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* 30, 175–200.
- Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol.* 5, 525–534.
- Wang, Z., Eickholt, J., and Cheng, J. (2010). MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics.* 26, 882–888.
- Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure.* 8, 407–418.
- Weissman, K.J., and Müller, R. (2008). Crystal Structure of a Molecular Assembly Line. *Angewandte Chemie International Edition.* 47, 8344–8346.
- West, S.E.H., Schweizer, H.P., Dall, C., Sample, A.K., and Runyen-Janecky, L.J. (1994). Construction of improved *Escherichia-Pseudomonas* shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in *Pseudomonas aeruginosa*. *Gene.* 148, 81–86.
- Winsor, G.L., Lam, D.K.W., Fleming, L., Lo, R., Whiteside, M.D., Yu, N.Y., Hancock, R.E.W., and Brinkman, F.S.L. (2011). *Pseudomonas* Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes. *Nucleic Acids Res.* 39, D596–600.
- Yang, Y., Faraggi, E., Zhao, H., and Zhou, Y. (2011). Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics.* 27, 2076–2082.
- Yeh, E., Kohli, R.M., Bruner, S.D., and Walsh, C.T. (2004). Type II Thioesterase Restores Activity of a NRPS Module Stalled with an Aminoacyl-S-enzyme that Cannot Be Elongated. *ChemBioChem.* 5, 1290–1293.
- Yeterian, E., Martin, L.W., Guillon, L., Journet, L., Lamont, I.L., and Schalk, I.J. (2009). Synthesis of the siderophore pyoverdine in *Pseudomonas aeruginosa* involves a periplasmic maturation. *Amino Acids.* 38, 1447–1459.

- Yonus, H., Neumann, P., Zimmermann, S., May, J.J., Marahiel, M.A., and Stubbs, M.T. (2008). Crystal Structure of DltA IMPLICATIONS FOR THE REACTION MECHANISM OF NON-RIBOSOMAL PEPTIDE SYNTHETASE ADENYLATION DOMAINS. *J. Biol. Chem.* 283, 32484–32491.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 9, 40.
- Zhang, J., Wang, Q., Barz, B., He, Z., Kosztin, I., Shang, Y., and Xu, D. (2010). MUFOLD: A new solution for protein 3D structure prediction. *Proteins.* 78, 1137–1152.
- Zhang, K., Nelson, K.M., Bhuripanyo, K., Grimes, K.D., Zhao, B., Aldrich, C.C., and Yin, J. (2013). Engineering the Substrate Specificity of the DhbE Adenylation Domain by Yeast Cell Surface Display. *Chem Biol.* 20, 92–101.
- Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotech* 16, 258–261.
- Zhou, H., and Skolnick, J. (2007). *Ab Initio* Protein Structure Prediction Using Chunk-TASSER. *Biophys J.* 93, 1510–1518.
- Zhou, H., and Skolnick, J. (2009). Protein Structure Prediction by Pro-Sp3-TASSER. *Biophys J.* 96, 2119–2127.
- Zhou, H., and Skolnick, J. (2011). Template-based protein structure modeling using TASSER-VMT. *Proteins.*
- Zhou, Z., Lai, J.R., and Walsh, C.T. (2006). Interdomain Communication between the Thiolation and Thioesterase Domains of EntF Explored by Combinatorial Mutagenesis and Selection. *Chem Biol.* 13, 869–879.
- Zhou, Z., Lai, J.R., and Walsh, C.T. (2007). Directed evolution of aryl carrier proteins in the enterobactin synthetase. *PNAS.* 104, 11621–11626.