

Investigating the biosynthesis of the streptomycete antibiotic pacidamycin

Daniel R. Tromans

Thesis submitted to the University of East Anglia for the
degree of Doctor of Philosophy

Department of Biological Chemistry
John Innes Centre
Norwich

And

School of Chemistry
University of East Anglia
Norwich

August 2013

© This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise the copyright rests with the author and that use of any information derived there from must be in accordance with current UK copyright law. In addition, any quotation or extract must include full attribution.

Date

August 2013

I certify that the work contained in this thesis submitted by me for the degree of PhD is my own original work, except where due reference is made, and has not been submitted by me for a degree at this or any other university.

Signed

Daniel R. Tromans

This PhD was funded by the Norwich Research Park

Acknowledgements

Acknowledgements

The trials, tribulations and experiences of a PhD student could be a book in itself. It is only with the guidance and support of supervisors, colleagues, family and friends that the story is completed. It is for that reason that I dedicate these pages and this thesis to all of them.

I would firstly like to express my deepest gratitude to my PhD supervisors; Dave Lawson, Rebecca Goss and Merv Bibb. I would like to thank you for giving me the opportunity to undertake my PhD and for all of the support and guidance you have given me along the way. I would particularly like to thank Dave for 'taking me in' when the Goss lab moved up to St Andrews.....I have really appreciated it.

I would also like to thank Clare Stevenson and Sabine Grüşchow for all the support and advice they have offered me over the past three years or so. Clare, the help you have given me with my crystallography work has been second to none and of course our 'occasional natter' in Chatt 127 has kept me sane.

For technical support I would like to thank Lionel Hill for all the training and advice you have given me on the LC-MS. Further to this I would like to thank Gerhard Saalbach for carrying out the proteomic analyses that are reported in this work. And I mustn't forget to thank the media kitchen for the service they provide.

As well as thanking the 'key people', I would like to thank all the people I have worked with over the past couple of years, including the members of the Lawson, Goss and Bibb labs and the departments of Biological Chemistry and Molecular Microbiology at the JIC. You have all made my time in Norwich enjoyable. I would particularly like to thank Mike, Abhi, Farzana and Tracey for being their to trouble shoot the research and for good times out of the lab.

I would also like to thank my family: Dad, Aunty Shez, Kate, Jon and Grandparents. You have all been so supportive and have always let me find my own way, even if you didn't always agree with my decisions. And Mum, I know you are looking down on us. I hope I have made you proud.

Acknowledgements

Finally, I would like to thank my fiancée, Laura, for your love and support. I want to thank you for moving here to Norwich; I know it was big upheaval for you too. I also want to thank you for putting up with me when I'm having a 'stress'. I couldn't have done it without you (although I would have had less washing up to do!). I can't wait to see what adventures we will have next.

Abstract

There is an ever increasing need for the development of new antibiotics to fight the emergence of antibacterial resistant strains of pathogens. Developing antimicrobials with 'novel scaffolds' and modes of action is an effective way to combat pathogens that are resistant to compounds currently in clinical use. The pacidamycins are a member of the uridyl peptide class of antibiotics that are produced by the soil dwelling bacterium *Streptomyces coeruleorubidus*. They show specific activity against the pathogen *Pseudomonas aeruginosa*, using a currently unexploited mode of action against a cell wall biosynthetic enzyme target. This thesis reports the investigation into the biosynthesis of pacidamycin, more specifically, into the function of the hypothetical protein genes present in the pacidamycin gene cluster and the biosynthesis of the non-proteinogenic amino acid, (2S, 3S)-diaminobutyric acid (DABA), which is at the core of the pacidamycin structure and other related antimicrobials.

A multidisciplinary approach has been taken in this investigation, utilising biophysical, biochemical and genetic approaches. Protein crystallographic studies have deduced the structure of Pac17, postulated to be a lyase involved in DABA biosynthesis along with structural determination of the protein bound to the proposed substrate aspartate. Site directed mutagenesis of a number of the Pac17 active site amino acids also showed their essentiality for aspartate binding. *In vitro* biochemical approaches to study the enzymatic activity of the DABA biosynthetic proteins were inconclusive, with no activity observed. Genetic disruptions of the genes under investigation revealed the function of *pac13* as a dehydratase, responsible for dehydrating the furan ring of the uridyl nucleoside present in the pacidamycin structure. Further to this, these studies established the essentiality of the DABA biosynthetic genes *pac19* and *pac20* for pacidamycin production in the native producer.

Abbreviations

Å	Armstrong
λ	X-ray wavelength
a, b, c, α , β , γ	Unit cell dimensions
AIM	Autoinduction medium
ASU	Asymmetric unit
ATP	Adenosine triphosphate
bp	Base pairs
Da	Daltons
DABA	(2 <i>s</i> ,3 <i>s</i>)-Diaminobutyric acid
DAP	Diamionpropionic acid
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
dNTP	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
F	Structure factor amplitude
F_{calc}	Calculated structure factor amplitude
F_{obs}	Observed structure factor amplitude
FOM	Figure of merit
g	Gram
h	Hour
IPTG	Isopropyl β -D-1-thiogalactopyranoside
K	Kelvin
kb	Kilobase
kDa	Kilodalton
kV	Kilovolt
LB	Luris Bertani medium
LC-MS	Liquid chromatography – mass spectrometry
LC-MS/MS	Liquid chromatography – mass spectrometry ₂
M	Molar
mg	Milligram
MIC	Minimum inhibition concentration
mL	Millilitre
min	Minute

Abbreviations

μm	Micrometre
μM	Micromolar
mM	Millimolar
ng	Nanogram
NRPS	Non-ribosomal peptide synthetase
nt	Nucleotide
PAGE	Polyacrylamide gel electrophoresis
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Poly ethylene glycol
rpm	Revolutions per minute
r.t	Room temperature
s	Second
SAM	S-adenosylmethionine
SAH	S-adenosylhomocysteine
SDS	Sodium dodecyl sulphate
<i>sp.</i>	Species
TE	Tris-EDTA
T _m	Melting temperature
UEA	University of East Anglia
V	Volt
v/v	Volume to volume
vol	Volumes
v/w	Volume to weight
wt	Wild type
°C	Degrees Celsius

Table of Contents

Title Page	i
Declaration	ii
Acknowledgements	iii
Abstract	v
Abbreviations	vi
Table of Contents	viii
List of Figures	xv
List of Tables	xix
Chapter 1 – Introduction	1
1.1 Natural products	2
1.1.1 General introduction	2
1.1.2 Secondary metabolites	2
1.1.3 Secondary metabolite classification	3
1.1.4 Ribosomally synthesised peptides	4
1.1.5 Non-ribosomal peptides and non-ribosomal peptide biosynthesis	5
1.2 Bacteria	8
1.2.1 The bacterial cell	8
1.2.2 Bacterial infection and resistance	9
1.2.3 <i>Pseudomonas aeruginosa</i>	10
1.2.4 Streptomyces	11
1.2.5 Targets for antimicrobial therapy	12
1.2.6 Peptidoglycan biosynthesis	13
1.2.7 Translocase I	15
1.2.8 Antibacterial cell wall targets	16
1.3 Nucleoside antibiotics	17
1.3.1 General introduction	17
1.3.2 Discovery of the uridyl peptides	18
1.3.3 Uridyl peptide antibacterial activity	19
1.3.4 Uridyl peptides and their postulated inhibition mechanism	20
1.3.5 The pacidamycins	21
1.3.5.1 Introduction	21
1.3.5.2 The pacidamycin gene cluster	22

Table of Contents

1.3.5.3	Additional genes identified to be involved in pacidamycin biosynthesis	24
1.4	Diamino acids and studies into their biosynthesis	25
1.5	Proposed biosynthesis of DABA	26
1.6	Project aims	27
Chapter 2 - Materials and Methods		28
2.1	General	29
2.1.1	General comments	29
2.1.2	General reagents	29
2.1.3	General equipment	30
2.1.4	Medium	31
2.1.5	Antibiotics	32
2.1.6	Buffers and stock solutions	33
2.1.7	Plasmids	34
2.1.8	Cosmids	36
2.1.9	Microorganisms	37
2.2	Basic molecular biology methodology	38
2.2.1	Primer design	38
2.2.2	Polymerase chain reaction	38
2.2.2.1	Primer sequences	38
2.2.2.2	Optimisation of PCR annealing temperature	41
2.2.2.3	Colony PCR using <i>E. coli</i>	41
2.2.2.4	Colony PCR using <i>Streptomyces</i>	42
2.2.3	Restriction, ligation and cloning into the pET vector system	43
2.2.4	Site directed mutagenesis (SDM)	44
2.2.5	Preparation of competent cells	45
2.2.5.1	Preparation of chemically competent cells	45
2.2.5.2	Transformation of chemically competent cells	45
2.2.5.3	Preparation of electrocompetent cells	45
2.2.5.4	Transformation of chemically competent cells	46
2.2.6	DNA purification and analysis	46
2.2.6.1	Alkaline lysis	46

Table of Contents

2.2.6.2	DNA purification	47
2.2.6.3	Gel filtration	47
2.2.6.4	Restriction digestion for plasmid analysis	48
2.2.6.5	Agarose gel electrophoresis	48
2.2.6.6	DNA sequencing	48
2.2.7	Bioassay of pacidamycin activity	49
2.3	Protein methodologies	50
2.3.1	Cell lysis	50
2.3.1.1	Cell lysis using cell lysis buffer	50
2.3.1.2	Cell lysis by sonication	50
2.3.1.3	Cell lysis by cell disruption	50
2.3.2	SDS polyacrylamide gel electrophoresis (SDS-PAGE)	50
2.3.3	Protein production trials	51
2.3.4	Protein production	51
2.3.5	Protein purification by affinity and size exclusion chromatography	51
2.3.5.1	Protein purification using the ÄKTA express FPLC	52
2.3.5.1	Protein purification using the ÄKTA FPLC	52
2.3.6	Bradford assay	53
2.3.7	Dynamic light scattering (DLS) analysis	53
2.3.8	Dialysis	53
2.3.9	His-tag removal	53
2.3.9.1	Removal using thrombin	54
2.3.9.2	His-tag removal using DAPase	54
2.3.10	In vitro protein activity methods	55
2.3.10.1	Determination of Pac17 activity	55
2.3.10.2	Determination of Pac19 activity	56
2.3.11	Protein crystallographic methods	56
2.3.11.1	Protein crystallization	56
2.3.11.2	Protein crystal cryo-protection and data collection	58
2.3.11.3	Data collection	59
2.3.11.4	Data processing	59
2.3.11.5	Molecular replacement	59
2.3.11.6	Phase improvement, model building and refinement	60

Table of Contents

2.3.11.7	Structural validation	61
2.4	Streptomycete genetic methodologies	61
2.4.1	Culturing <i>Streptomyces</i>	61
2.4.2	Culturing <i>Streptomyces sp.</i> on solid medium	61
2.4.2.1	Culturing <i>S. coeruleorubidus</i> in liquid medium	62
2.4.2.2	<i>S. coeruleorubidus</i> feeding experiments in liquid medium	62
2.4.2.3	Culturing <i>S. lividans</i> in liquid medium	62
2.4.3	Generation of genetic mutants in <i>S. coeruleorubidus</i>	62
2.4.4	Conjugation into <i>Streptomyces sp.</i>	66
2.4.5	Heterologous expression in <i>Streptomyces</i>	67
2.4.5.1	FLP-mediated recombination	67
2.4.5.2	Recombination of Sspl fragment into cosmid backbone	68
2.4.5.3	Cosmid integration into a <i>Streptomyces</i> heterologous host	68
2.4.6	Pacidamycin extraction and analysis	69
2.4.6.1	Pacidamycin extraction	69
2.4.6.2	Detection of pacidamycins by LC-MS analysis	69
2.4.6.3	Confirmation of pacidamycin production by LC-MS/MS analysis	70
 Chapter 3 - Cloning, Protein Production and Crystallisation		 71
3.1	Introduction	72
3.2	Results	72
3.2.1	Bioinformatics analysis of the proteins of interest	72
3.2.2	Pac17	78
3.2.2.1	Gene cloning of <i>pac17</i>	78
3.2.2.2	Expression studies of <i>pac17</i>	78
3.2.2.3	Large scale expression and purification of Pac17	78
3.2.2.4	Crystallisation of Pac17	81
3.2.3	Pac18	81
3.2.3.1	Gene cloning of <i>pac18</i>	81
3.2.3.2	Expression studies of <i>pac18</i>	82
3.2.4	Pac19	84
3.2.4.1	Gene cloning of <i>pac19</i>	84
3.2.4.2	Expression studies of <i>pac19</i>	84

Table of Contents

3.2.2.3	Large scale expression and purification of Pac19	8
3.2.2.4	Crystallisation of Pac19	87
3.2.5	Pac20	87
3.2.5.1	Gene cloning of <i>pac20</i>	87
3.2.5.2	Expression studies of <i>pac20</i>	88
3.2.5.3	Large scale expression and purification of Pac20	88
3.2.5.4	Crystallisation of Pac20	91
3.2.6	Pac22	92
3.2.6.1	Large scale expression and purification of Pac22	92
3.2.2.4	Crystallisation of Pac22	94
3.2.7	Gene cloning of the hypothetical protein genes	96
3.2.7.1	Discovery of the function of <i>pac2</i> and a homolog of <i>pac1</i>	97
3.3	Conclusions	98
 Chapter 4 - Structural Investigation of the Pac17 Protein		 99
4.1	Introduction	100
4.2	Results	101
4.2.1	Optimisation of the crystallisation of Pac17	101
4.2.2	Native data collection of a single Pac17 crystal	101
4.2.2.1	Pac17 data processing	102
4.2.2.2	Molecular replacement to solve the Pac17 apo structure	104
4.2.2.3	Model building and refinement of the apo structure	107
4.2.3	Higher resolution dataset of apo Pac17	109
4.2.4	Evaluating Pac17 apo model quality	113
4.2.5	Analysis of the Pac17 apo structure	116
4.2.6	Conservation of residues in the Pac17 structure	120
4.2.7	Crystallisation of Pac17 with its hypothesised substrate	122
4.2.7.1	Solving the structure of co-crystallised Pac17	123
4.2.8	The aspartate bound structure of Pac17	127
4.2.9	Further analysis of the ligand bound structure of Pac17	130
4.2.10	Comparison of Pac17 to AspB	132
4.2.11	Further interrogation of Pac17 substrate	136
4.2.12	Site directed mutagenesis (SDM)	138

Table of Contents

4.3	Conclusions	142
Chapter 5 - Pacidamycin Cluster Gene Disruption Studies		144
5.1	Introduction	145
5.2	Results	146
5.2.1	Gene disruption studies in <i>S. coeruleorubidus</i>	146
5.2.1.1	Gene disruption studies of hypothetical protein genes	147
5.2.1.2	Disruption of <i>pac7</i> and <i>pac13</i> in <i>S. coeruleorubidus</i>	148
5.2.1.3	Liquid growth and LC-MS analysis of <i>S. coeruleorubidus</i> DT-001 and 002	149
5.2.1.4	Further work carried out on <i>S. coeruleorubidus</i> DT-001 and 002	151
5.2.1.5	Gene disruption studies of the DABA biosynthetic genes	152
5.2.1.6	Gene disruption studies in <i>S. coeruleorubidus</i> of DABA genes	152
5.2.1.7	Liquid growth, metabolite extraction and LC-MS of <i>S. coeruleorubidus</i> DT-003, DT-004 and DT-005	156
5.2.1.8	Chemical complementation analysis	158
5.2.2	Gene disruption studies in the heterologous host <i>S. lividans</i> TK24	158
5.2.2.1	Integration of the non-disrupted 2H-5 cosmid into <i>S. lividans</i> TK24	159
5.2.3	New approach to analysis of gene function	161
5.2.3.1	Construction of the disrupted <i>pac17</i> and <i>pac18</i> integration cosmids	162
5.2.3.2	Growth of <i>S. lividans</i> strains DT-007 and DT-008, metabolite extraction and analysis	164
5.2.4	Gene disruption studies of <i>pac19</i> and <i>pac20</i>	165
5.2.4.1	Construction of the disrupted <i>pac19</i> and <i>pac20</i> integration cosmids	167
5.2.3.6	Growth of <i>S. lividans</i> DT-009 and DT-010, metabolite extraction and analysis	169
5.3	Conclusion	172
Chapter 6 - <i>In Vitro</i> Studies of the DABA Biosynthesis Genes		174
6.1	Introduction	175
6.2	Pac17	175
6.2.1	Introduction	175
6.2.2	Results of the investigation into the activity of Pac17	176
6.2.3	GC-MS analysis	178

Table of Contents

6.2.4	Use of cell free extract and incubation with <i>S. coeruleorubidus</i>	179
6.3	Pac19	179
6.3.1	Introduction	179
6.3.2	Results of the investigation into the activity of Pac19	180
6.3.3	Use of cell free extract and incubation with <i>S. coeruleorubidus</i>	181
6.3.4	Coupling of Pac17 and Pac19	181
6.4	Pac20	181
6.5	Discussion	182
Chapter 7 - General Discussion and Conclusions		184
References		191
Appendix 1		206
Publications		210

List of Figures

Chapter 1 - Introduction

Figure 1.1	Examples of structures of secondary metabolites	4
Figure 1.2	Diagram of microbisporicin and nisin	5
Figure 1.3	Adenylation domain (NRPS)	6
Figure 1.4	Peptidyl carrier protein (NRPS)	7
Figure 1.5	Condensation domain (NRPS)	7
Figure 1.6	Synthesis of a non-ribosomal peptide	8
Figure 1.7	Streptomyces life cycle	12
Figure 1.8	Peptidoglycan cell wall	13
Figure 1.9	Initial stages of cell wall biosynthesis	15
Figure 1.10	Cell wall antibacterial targets	16
Figure 1.11	Structure of toyomycin and adenosine	18
Figure 1.13	Uridyl peptides	19
Figure 1.14	Similarities of MraY substrate and uridyl peptides	21
Figure 1.14	Pacidamycin structure	22
Figure 1.15	Pacidamycin gene cluster	23
Figure 1.16	Pacidamycin assembly	25
Figure 1.17	Proposed biosynthesis of DABA	27

Chapter 2 – Materials and Methods

Figure 2.1	Basic solubility phase diagram	57
Figure 2.2	Illustration of the Re-direct protocol	65
Figure 2.3	Illustration of cosmid integration	69

Chapter 3 – Cloning, Protein Production and Crystallisation

Figure 3.1	FOLDINDEX output for Pac17 sequence	73
------------	-------------------------------------	----

List of Figures

Figure 3.2	PSIPRED output for Pac17 sequence	74
Figure 3.3	PHYRE2 output for Pac17 sequence	75
Figure 3.4	XTALPRED output for Pac17 sequence	76
Figure 3.5	Analysis of Pac17 purification	80
Figure 3.6	Pac17 crystal trials	81
Figure 3.7	Analysis of Pac19 purification	86
Figure 3.8	Pac19 crystal trials	87
Figure 3.9	Analysis of Pac20 purification	90
Figure 3.10	Intact mass analysis of Pac20	91
Figure 3.11	Analysis of Pac22 purification	93
Figure 3.12	Mass analysis of Pac22	94
Figure 3.13	SDS-PAGE of Pac22 during treatment with DAPase	95
Figure 3.14	Pac22 protein crystal testing	96
Figure 3.15	Illustration of the function of <i>pac2</i>	98
Chapter 4 – Structural Investigation of the Pac17 Protein		
Figure 4.1	Translationally coupling of <i>pac17-19</i>	100
Figure 4.2	Pac17 crystals	101
Figure 4.3	Pac17 crystal X-ray diffraction pattern	102
Figure 4.4	Self rotation function of Pac17 structure	104
Figure 4.5	Pac17 alignment with 2E9F	105
Figure 4.6	Pac17 self Patterson function	106
Figure 4.7	Graphical representation of Pac17 structure improvement in BUCANEER	107
Figure 4.8	Summary of Pac17 structural elucidation	109
Figure 4.9	Higher resolution Pac17 X-ray diffraction data	110
Figure 4.10	Summary of solving higher resolution Pac17 structure	112
Figure 4.11	MOLPROBITY output for Pac17 structure	114

List of Figures

Figure 4.12	Ramachandran plot for Pac17 model	114
Figure 4.13	Methionine 199 (1)	115
Figure 4.14	Methionine 199 (2)	116
Figure 4.15	PDBSUM output for Pac17 structure	117
Figure 4.16	Pac17 structure images	118
Figure 4.17	Pac17 superimposed onto homologs	120
Figure 4.18	Sequence conservation of Pac17	121
Figure 4.19	Diffraction image from Pac17-Asp crystal X-ray data collection	123
Figure 4.20	Self rotation function for Pac17-Asp structure	124
Figure 4.21	Summary of solving ligand bound Pac17 structure	126
Figure 4.22	MOLPROBILTY output for ligand bound Pac17 structure	127
Figure 4.23	Superimposition of apo- and ligand bound Pac17 structures	128
Figure 4.24	Comparison of atoms between apo- and ligand bound Pac17 structures	129
Figure 4.25	Aspartate ligand images	130
Figure 4.26	Co-ordination of aspartate to active site residues	131
Figure 4.27	Co-ordination of Asn109 with aspartate ligand	132
Figure 4.28	Movement of Ser276 between apo- and ligand bound Pac17 structures	133
Figure 4.29	Postulated enzymatic mechanism of Pac17	135
Figure 4.30	Mutated amino acid images in the Pac17 mutants	142

Chapter 5 - Pacidamycin Cluster Gene Disruption Studies

Figure 5.1	LC-MS profile of <i>S. coeruleorubidus</i> metabolites	146
Figure 5.2	MS ₂ Fragmentation pattern of pacidamycin D	147
Figure 5.3	PCR results for <i>pac7</i> and <i>pac13</i> disruptions	149
Figure 5.4	LC-MS profile of <i>S. coeruleorubidus</i> , <i>pac7</i> and <i>pac13</i> disruptions	150
Figure 5.5	Pacidamycin D with uridyl derivative highlighted	151
Figure 5.6	PCR results for <i>pac17-20</i> disruption <i>S. coeruleorubidus</i>	154

List of Figures

Figure 5.7	PCR results for <i>pac19</i> and <i>pac20</i> disruptions in <i>S. coeruleorubidus</i>	155
Figure 5.8	LC-MS profile of <i>S. coeruleorubidus</i> and <i>pac17-20</i> , <i>pac19</i> and <i>pac20</i> disruptions	157
Figure 5.9	LC-MS profile of <i>S. lividans</i> TK24	159
Figure 5.10	LC-MS and MS ₂ profile of <i>S. lividans</i> heterologously expressing pacidamycin D	160
Figure 5.11	Illustration of new gene disruption strategy	161
Figure 5.12	PCR results for <i>pac17</i> disruption	163
Figure 5.13	PCR results for <i>pac18</i> disruption	164
Figure 5.14	LC-MS profile of negative and positive controls of <i>S. lividans</i> and <i>pac17</i> and <i>18</i> disruptions	165
Figure 5.15	MS ₂ profile of <i>pac18</i> disruption	166
Figure 5.16	PCR results for <i>pac19</i> disruption in <i>S. lividans</i>	168
Figure 5.17	PCR results for <i>pac20</i> disruption in <i>S. lividans</i>	169
Figure 5.18	LC-MS profile of negative and positive controls and <i>pac19</i> and <i>20</i> disruptions	171
Figure 5.19	MS ₂ profile of <i>pac20</i> disruption in <i>S. lividans</i>	172
Chapter 6 – <i>In Vitro</i> Studies of the DABA Biosynthesis Genes		
Figure 6.1	Standard curves of 1 mM aspartate and 1 mM fumarate	177
Figure 6.2	Derivatisation of DABA	179
Figure 6.3	Standard curve of derivitisation of DAP	180
Chapter 7 - General Discussion and Conclusions		
Figure 7.1	Postulated biosynthesis of DABA	186
Figure 7.2	Structure of Pac17	188

List of Tables

Chapter 2 – Materials and Methods

Table 2.1	Medium recipes	31
Table 2.2	Antibiotics	32
Table 2.3	General reagent recipes	33
Table 2.4	Plasmids	34
Table 2.5	Cosmids	36
Table 2.6	Microorganisms	37
Table 2.7	Primer sequences	38

Chapter 3 – Cloning, Protein Production and Crystallisation

Table 3.1	Summary of bioinformatics analysis of proteins in this study	77
Table 3.2	Cloning parameters for the hypothetical proteins	97

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.1	Statistics for X-ray data collection of Pac17	103
Table 4.2	Statistics for X-ray higher resolution data collection of Pac17	111
Table 4.3	Final structural parameters for Pac17 structure	113
Table 4.4	Pac17 structural homologs comparison	199
Table 4.5	Data collection statistics for Pac17 co-crystallised with aspartate	124
Table 4.6	Final structural parameters of ligand bound Pac17 structure	126
Table 4.7	Comparison of active site residues in Pac17 and AspB	134
Table 4.8	Data collection statistics for Pac17 co-crystallised with L-glutamine and L-glutamate	136
Table 4.9	Primers used for site directed mutagenesis of Pac17	138
Table 4.10	Statistics for data collection of Asn109 and Met279 Pac17 mutants	139
Table 4.11	Final structural parameters for Asn109 and Met176 Pac17 mutants	140
Table 4.12	Statistics for data collection of Lys282 and As284 Pac17 mutants	140

List of Tables

Table 4.13 Final structural parameter for Lys279 and Asn284 Pac17 mutants 141

Chapter 6 -In Vitro Studies of the DABA Biosynthesis Genes

Table 6.1 Conditions used in the Pac17 *in vitro* activity study 178

Introduction

1.1 Natural products

1.1.1 General introduction

Natural products can be subdivided into three classes: primary and secondary metabolites and high molecular weight polymeric material. Primary metabolites are compounds essential for cell survival. Secondary metabolites are not essential for cell survival but benefit the organism in some way, often offering a selective advantage to the producing organism (Stone and Williams, 1992). High molecular weight polymeric material includes cellulose and chitin. These are generally responsible for the structural integrity of the organism. Primary metabolites are highly structurally conserved, however, secondary metabolites often have many unique and unusual features. The regulation of primary and secondary metabolite production is tightly controlled as gene expression and natural product production is energetically expensive. As a result, an organism will generally produce secondary metabolites at a later stage of its life cycle or at a time when competition for survival is high.

1.1.2 Secondary metabolites

For millennia secondary metabolites have been exploited by humans in the form of natural remedies and poisons (Newman et al., 2000). Even so, it was not until the 1920's - 40's, with the research of Fleming, Florey and Chain, and the eventual mass production of penicillin that the 'secondary metabolite boom' began. Since then, many diverse classes of secondary metabolites have been isolated from an assortment of bacteria, fungi and plants, and many have formed the basis for a wide selection of clinically important drugs, including anti-bacterials, anti-fungals, anti-oxidants and anti-inflammatories (Clardy and Walsh, 2004). The majority of secondary metabolites discovered and utilised in the clinic have been isolated from actinobacteria, particularly from the streptomycete family (Newman and Cragg, 2012). The first antibiotic identified from this family was actinomycin from *Streptomyces antibioticus* which was found to inhibit transcription. This was also the first secondary metabolite found to have anticancer activity (Waksman and Woodruff, 1941, Hollstein, 1974).

To date more than 20,000 secondary metabolites have been identified as having some form of bioactivity (Marinelli, 2009), however, due to a decrease in novel structures and chemical classes being discovered, the techniques used in natural product discovery have

become more and more elaborate. A need for new potent compounds that can fight an array of infections and ailments has resulted in exploration of the seas and the most extreme natural environments (Desbois et al., 2009).

1.1.3 Secondary metabolite classification

Secondary metabolites can be separated into a number of chemical classes. These include the polyketides, the terpenoids, the phenylpropanoids, the alkaloids and the ribosomal and non-ribosomal peptides. The polyketides (Figure 1.1) are formed from the decarboxylative condensation of malonate or methyl/ethyl malonate building blocks obtained from acetyl co-enzyme A; their assembly is mediated by polyketide synthases in a similar fashion to fatty acid synthesis by the fatty acid synthase. The terpenoids (Figure 1.1) include the steroids and are constructed from the assembly of five carbon isoprene units. The majority of these compounds contain characteristic cyclic structures. Phenylpropanoids (Figure 1.1) as their name suggests contain aromatic ring structures with a three carbon chain attached to the ring (Hanson, 2003). These secondary metabolites are commonly synthesised in plants and are often found as a component of essential oils (Desbois et al., 2009). The alkaloids (example Figure 1.1) are a chemical class characterised by the presence of a base containing a nitrogen and, again, are synthesised by plants. Many alkaloids have neuro-active properties, a property which has been frequently exploited. Finally the ribosomal and non-ribosomal peptide class (Figure 1.1) consists of the ribosomal peptides, where the peptide is directly encoded in the DNA and produced by the ribosome, and the non-ribosomal peptides that are synthesised by non-ribosomal peptide synthetases, which, like the polyketide synthases are multifunctional enzymes (Hanson, 2003).

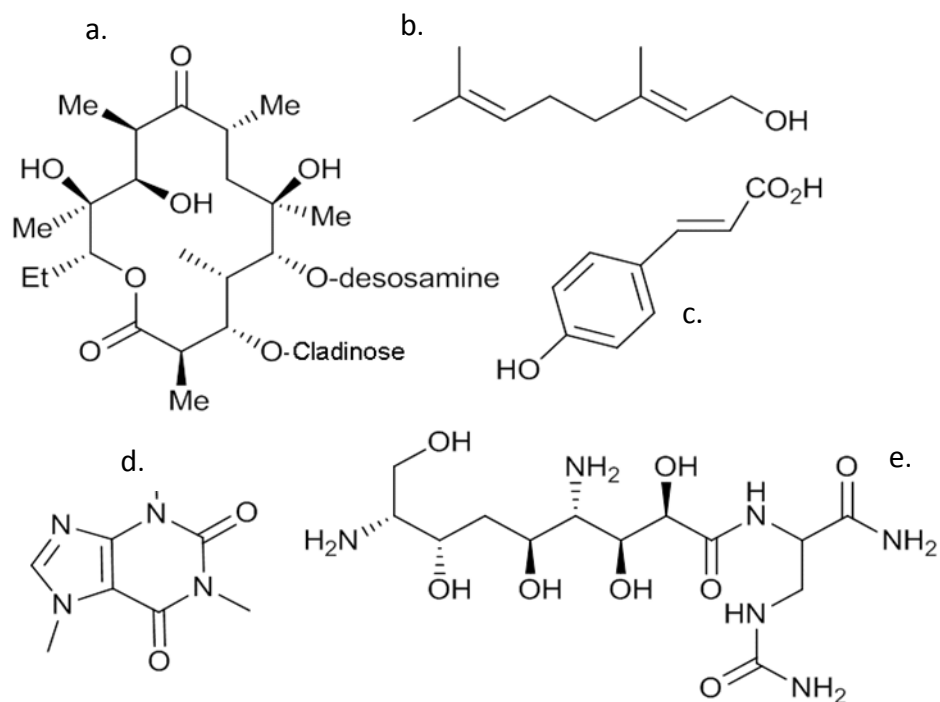


Figure 1.1 Examples of structures of the different secondary metabolite classes, (a) erythromycin is a macrolide polyketide antibiotic, (b) geraniol is a mono-terpenoid and (c) 4-hydroxycinnamic acid is one of the simpler plant phenylpropanoids. A commonly consumed alkaloid is (d) caffeine which is a type of purine alkaloid and (e) zwittermicin A is an antibacterial and antifungal non-ribosomal peptide used in agriculture.

1.1.4 Ribosomally synthesised peptides

Ribosomally synthesised peptides are synthesised by the same machinery as proteins required for primary metabolism. After the synthesis of the linear peptide, it can undergo additional modification (post translational modification) such as hydroxylation, decarboxylation, proteolysis and the formation of disulphide bridges. An example of ribosomally synthesised secondary metabolites are the lantibiotics (Asaduzzaman and Sonomoto, 2009). These natural products are characterised by the presence of lanthionine and 3-methylanthionine bridges. Examples of lantibiotics include nisin (commonly used in food preservation) and the potent antimicrobial microbisporicin (Sang and Blecha, 2008, Castiglione et al., 2008, Foulston and Bibb, 2010).

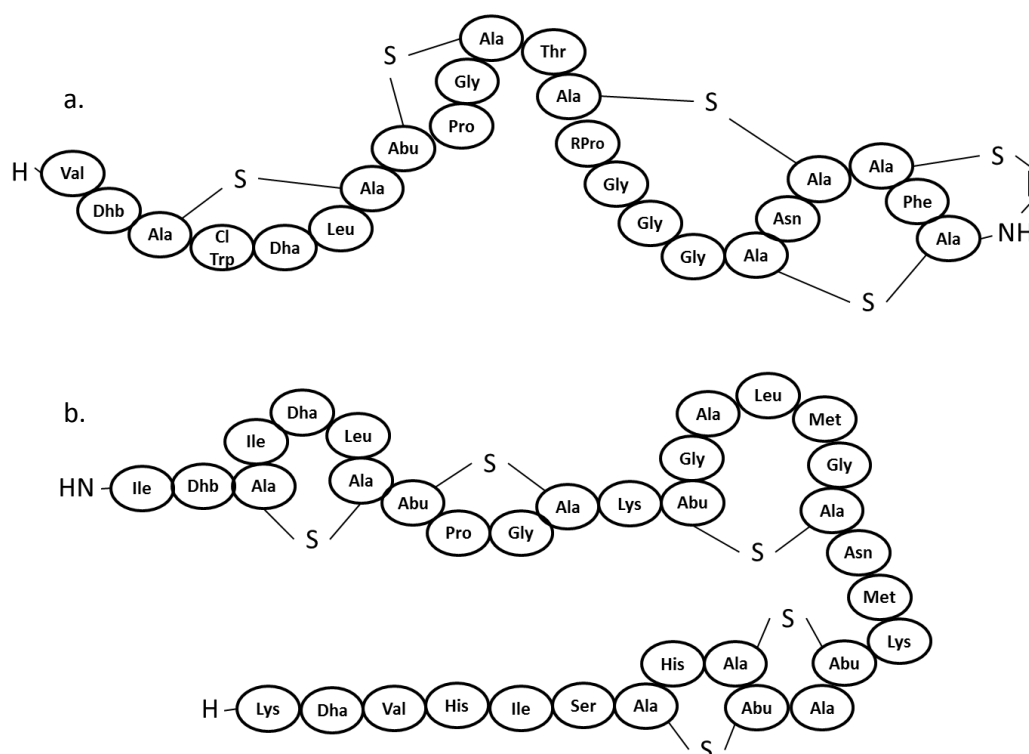


Figure 1.2 Diagrammatic representation of (a) microbisporicin and (b) nisin, two lantibiotics. The peptides are made ribosomally and are modified post translationally to form lanthionine bridges. Certain amino acids are modified to form derivatives such as 2,3-didehydroalanine (Dha), 2,3-didehydrobutyrine(Dhb), 2-amino butyric acid (Abu), chlorinated tryptophan (Cl-Trp) and 3,4-dihydroxyproline.

1.1.5 Non-ribosomal peptides and non-ribosomal peptide biosynthesis

The non-ribosomal peptides are a diverse class of secondary metabolites, with many having valuable anti-bacterial and anti-fungal properties (Schwarzer et al., 2003). In nature, non-ribosomal peptides are synthesised by complexes known as non-ribosomal peptide synthetases (NRPS's). NRPS's are similar in organisation to the polyketide synthases in that they are multi-enzyme complexes containing modules and domains (Hanson, 2003, Challis and Naismith, 2004). Each module within the NRPS results in the addition of an amino acid to the growing peptide. A module contains a number of individual domains. The number of domains within the module will vary, from comprising of the 'core' domains alone (an adenylation domain, condensation domain and peptidyl carrier protein (PCP)) to also including many 'tailoring enzymes' whose functions are to catalyse different modifications to the peptide. Some of these enzymes include cyclases, ketoreductases,

Chapter 1 - Introduction

methyltransferases, epimerases and glycosyltransferases (Dieckmann et al., 1995, Hanson, 2003).

The adenylation domain's role is to control specificity of amino acid selection and to 'activate' the amino acid (Marahiel et al., 1997, Dewick, 2003). The amino acid is activated by converting it to an aminoacyl adenylate, using adenosine triphosphate (ATP). The activated amino acid can then be transferred to the co-factor of the PCP domain, forming an aminoacyl thioester (Dieckmann et al., 1995, Dewick, 2003). The ability of the adenylation domain to be specific yet diverse (in terms of what it may activate) has allowed for a huge catalogue of suitable substrates, including non-proteinogenic amino acids, a characteristic of the product that has made them very useful to humans (Marahiel et al., 1997, Lautru and Challis, 2004).

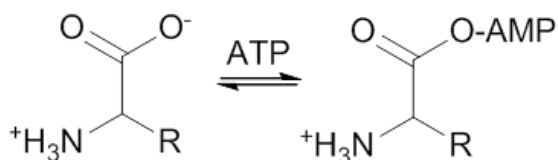


Figure 1.3 The adenylation domain 'activates' a specific amino acid, converting it to its aminoacyl adenylate derivative.

The peptidyl carrier protein of the NRPS has a role analogous to that of the acyl carrier protein (ACP) of the polyketide synthase. The PCP originally exists in its inactive or apo-form and must be 'primed' to its active or holo-form by addition of its co-factor 4'-phosphopantethiene (4-PP) (Lambalot et al., 1996). The co-factor is transferred to a conserved serine residue of the PCP by a 4'-phosphopantetheinyl transferase (Lambalot et al., 1996).

Chapter 1 - Introduction

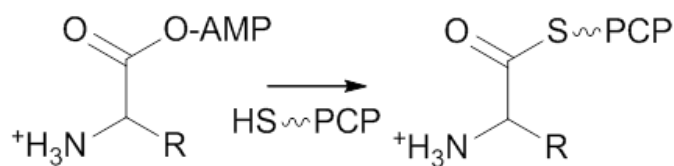


Figure 1.4 The activated amino acid is uploaded as a thioester onto the 4'-phosphopantetheinyl appendage of the PCP, which will act to chaperone the amino acid to the various enzymatic domains for subsequent incorporation into the peptide and modification with the aminoacyl.

The final 'core' domain of each module is the condensation domain. The condensation domain is responsible for the peptide bond formation between the peptide chain and the activated amino acid (Stachelhaus et al., 1998, Bergendahl et al., 2002). Studies have shown that the condensation domain also has substrate specificity and is able to distinguish between enantiomers and differentiate between different side chains (Belshaw et al., 1999). It is worth mentioning that the first module of a NRPS often lacks the condensation domain as no peptide bond needs to be formed, this module is normally named the 'initiation module' (Hanson, 2003).

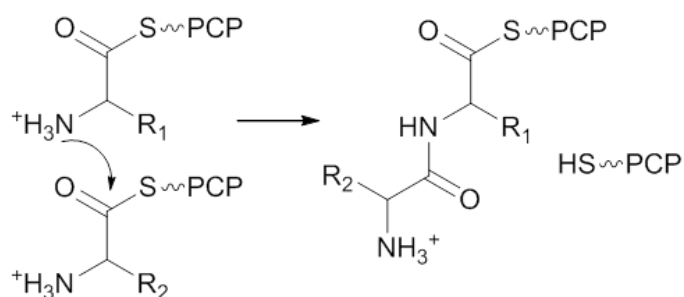


Figure 1.5 The condensation domain mediates amide bond formation between the activated amino acid and the growing peptide chain tethered to two neighboring PCPs.

Studies by Stein et al. (1996) have suggested that the affinity of the 4'-PP cofactor of the PCP alters depending on which state it is in (Figure 1.6) (Stein et al., 1996). Without a loaded amino acid, the 4'-PP has a higher affinity for the adenylation domain. When loaded with an amino acid, its affinity changes, being higher for the condensation domain, where a peptide bond is formed. The peptidyl-loaded PCP then changes its affinity again for the donor site of the condensation domain of the next module (Marahiel et al., 1997).

Chapter 1 - Introduction

The final module of the NRPS tends to be a thioesterase or TE domain. This domain allows the release of the peptide by cleaving the thioester bond between the peptide chain and the 4'-PP cofactor of the PCP domain (Hanson, 2003). After the release of the peptide product, many go through macrocyclisation to leave the peptide in its final structural state (Marahiel et al., 1997).

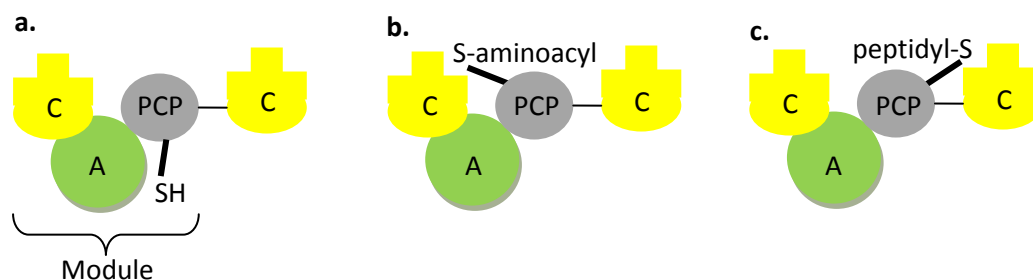


Figure 1.6 Diagrammatic representation of the synthesis of the peptide by the core domains of the NRPS (adapted from Marahiel et al. 2003). Originally the 4'-PP cofactor has highest affinity for the adenylation domain (a), and is loaded with the amino acid. Once loaded, its affinity changes for the acceptor site of the module's condensation domain (b) and a peptide bond is formed between the growing peptide and the activated amino acid. Finally, the affinity of the PCP loaded with the growing peptide changes for the donor site of the condensation domain of the next module (c), at which point the process begins again.

Some non-ribosomal peptides also contain acetate and propionate units (the building blocks of the polyketides) and are known as non-ribosomal peptide/polyketide hybrids (Newman et al., 2000).

1.2 Bacteria

1.2.1 The bacterial cell

Bacteria and archaea are the two most abundant kingdoms on earth. Bacteria form the basis of the majority of the World's ecosystems and therefore have an essential role for the continuity of life.

The bacterial cell consists of the cytoplasm which contains most of the cellular material, such as plasmids, ribosomes and other essential cell features (the bacterial cell not being as ordered as a eukaryotic cell) (Pollard and Earnshaw, 2008). The cytoplasm of the cell is contained by the cytoplasmic membrane. Bacterial cell walls differ between Gram positive and Gram negative bacteria. Gram negative bacteria have a thinner peptidoglycan layer

Chapter 1 - Introduction

than that of Gram positive bacteria, but they also have an additional outer layer consisting of an assortment of lipopolysaccharides and proteins which has been found to be an effective barrier against the uptake of secondary metabolites (Kimura and Bugg, 2003, Madigan and Martinko, 2005).

1.2.2 Bacterial infection and resistance

The majority of bacterial species are harmless to humans and other eukaryotic organisms, many are even beneficial, for example the microflora in the gut which aids digestion (Sears, 2006). However, pathogenic bacteria have plagued the health of humans and animals alike for millennia and include members of the *Mycobacterium*, *Pseudomonas*, *Streptococcus*, *Staphylococcus* and *Escherichia* genera (Heise, 1982, Saiman, 2004).

The era of antibiotics has played an important role in medicine. The discovery of antibiotics has allowed for successful treatment of infections that were previously untreatable and often led to death (Cirz et al., 2005, Payne et al., 2007). Today, due to overuse and misuse of these compounds, many common pathogens have developed resistance to common antibiotics (Overbye and Barrett, 2005). Some examples of bacterial resistance emergence have been astonishingly rapid, for example, the introduction into the clinic of derivatives of penicillin, such as methicillin to circumvent bacterial resistance to penicillin resulted in the isolation of methicillin resistant *Staphylococcus aureus* (MRSA) within a year. The resistance was caused by the bacteria's ability to acquire a hydrolytic enzyme that is able to breakdown the antibiotic (Friedmann, 1948). The emergence of resistance to other antibiotics has been less swift, for example, it took almost 30 years for pathogenic bacteria to acquire resistance against vancomycin. This was due to the bacteria needing to acquire a cassette of five genes to become resistant (Fan et al., 1994, Walsh, 2003a). A method to combat the emergence of antibacterial resistance is to discover and develop new secondary metabolite classes and clinical targets, however, recently very few new structural classes of secondary metabolites have been identified, with only four new structural classes being discovered in the past fifty years (Fischbach and Walsh, 2009).

The Infectious Disease Society of America (IDSA) reported in 2004 that two million patients developed bacterial infections, 90,000 of which were fatal. Of the two million cases, more than 70% of the pathogens found to be causing the infections had resistance to one or more of the most commonly used antibiotics (Overbye and Barrett, 2005).

Chapter 1 - Introduction

Bacteria can acquire resistance in a number of ways. Firstly, resistance may develop passively from existing innate responses. Secondly, it can develop actively, normally by the acquisition of genetic material by horizontal transfer (Wright and Sutherland, 2007). A number of different mechanisms for antibacterial resistance have been identified. These mechanisms include enzymatic modification of the drug, mutation of the drug target, sequestration of the drug, preventing drug entry and the active transport of the drug from the cell interior via efflux pumps (Wright, 2005, Walsh, 2003b). Other studies into antibacterial resistance in pathogens have shown that prolonged exposure of these organisms to certain anti-bacterials can accelerate the development of resistance, primarily by horizontal transfer and via mutation caused by the SOS response (Wright and Sutherland, 2007).

1.2.3 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is an aerobic Gram negative bacterium (Lyczak et al., 2000, Driscoll et al., 2007). It can colonise a wide range of environments including aquatic habitats, the phyllosphere of plants and the rhizosphere of soil (Lyczak et al., 2000, Kerr and Snelling, 2009). The ability of this pathogen to form biofilms helps it to survive in these environmental niches (Kerr and Snelling, 2009). The common occurrence of *P. aeruginosa* in nature, particularly water, and its ability to utilise a variety of energy sources, gives it an opportunistic position to cause infection in humans (Lyczak et al., 2000). These characteristics cause problems in eradicating the pathogen within hospitals using general cleaning procedures. The ability of *P. aeruginosa* to form biofilms within the host, for example, within the lung cavity, makes the treatment against this pathogen more difficult (Kerr and Snelling, 2009). Furthermore, *P. aeruginosa* has been found to develop resistance more efficiently than many other Gram negative and Gram positive pathogens. Its most common resistance being against β -lactams and antibiotics such as tetracycline and chloramphenicol, making this pathogen particularly problematic (Li et al., 1994, Lambert, 2002, Livermore, 2002, Walsh et al., 2005).

It needs to be iterated that infection in healthy individuals by this pathogen is rare. Infection commonly occurs in burn patients, individuals with cystic fibrosis (lung condition) and others with a compromised immune response (Lyczak et al., 2000). Even so, *P. aeruginosa* was the etiological infection (original infection) of 18.1% of hospital acquired

Chapter 1 - Introduction

pneumonia, 3.4% of bloodstream infections, 16.3% of urinary tract infections and 9.5% of surgical site infections in USA hospitals in 2003, according to IDSA reports (NNIS, 2004).

In recent years, a number of antibiotic therapies have been developed to fight pseudomonal infections (Doering and Pier, 2008). Among these are the lipopolysaccharide, surface polysaccharide and polysaccharide-protein conjugate antigen vaccines (Doering and Pier, 2008). However, none have gained market approval and very few have reached clinical trials (Doering and Pier, 2008).

P. aeruginosa has a broad range of virulence mechanisms. Among these are secreting toxins and forming biofilms (Driscoll et al., 2007). In recent years, studies into resistance by this organism have alarmed scientists and medical staff alike. A study by the National Nosocomial Infections Surveillance group revealed that resistance of the pathogen to commonly used anti-bacterials such as flouroquinolones, imipenem and third generation cephalosporins had increased to 29.5%, 21.1% and 31.9%, respectively, in 2003 (NNIS, 2004). The virulence of *P. aeruginosa* has been found to be both multi-factorial and combinatorial. A study in 2006 showed that the virulence genes in one strain of the pathogen may not be present in another strain (Lee et al., 2006). The study also suggested that the presence of pathogenicity islands (clusters of virulence related genes), can be acquired via horizontal transfer (Lee et al., 2006). Antibacterial resistance by *P. aeruginosa* can consist of a number of mechanisms. The pathogen has been found to have a variety of efflux pumps, allowing it to pump the antibacterial agent from the cytosol to the cell exterior (Lee et al., 2006). Furthermore, inactivating enzymes are used to deactivate the activity of the antibacterial agent (Lee et al., 2006).

1.2.4 Streptomycetes

Streptomycetes are soil dwelling Gram positive bacteria which have a characteristically high GC content ($\geq 70\%$) in their genomes. The life cycle of *Streptomyces sp.* is similar to that of fungi whereby a free spore, when located on favorable medium, germinates. Spore germination is then followed by the formation of vegetative mycelium (mycelium grows into solid phase) and the eventual formation of aerial hyphae from the sporophores formed by the vegetative mycelium. The aerial hyphae become spiraled, finally forming strands of mature spores, thus allowing the life cycle to begin again. As previously stated, the streptomycete family are prolific producers of secondary metabolites. Secondary

Chapter 1 - Introduction

metabolism in *Streptomyces sp.* is tightly regulated, with the formation of secondary metabolites occurring during the generation of aerial hyphae (Kieser et al., 2000, Bibb, 2005).

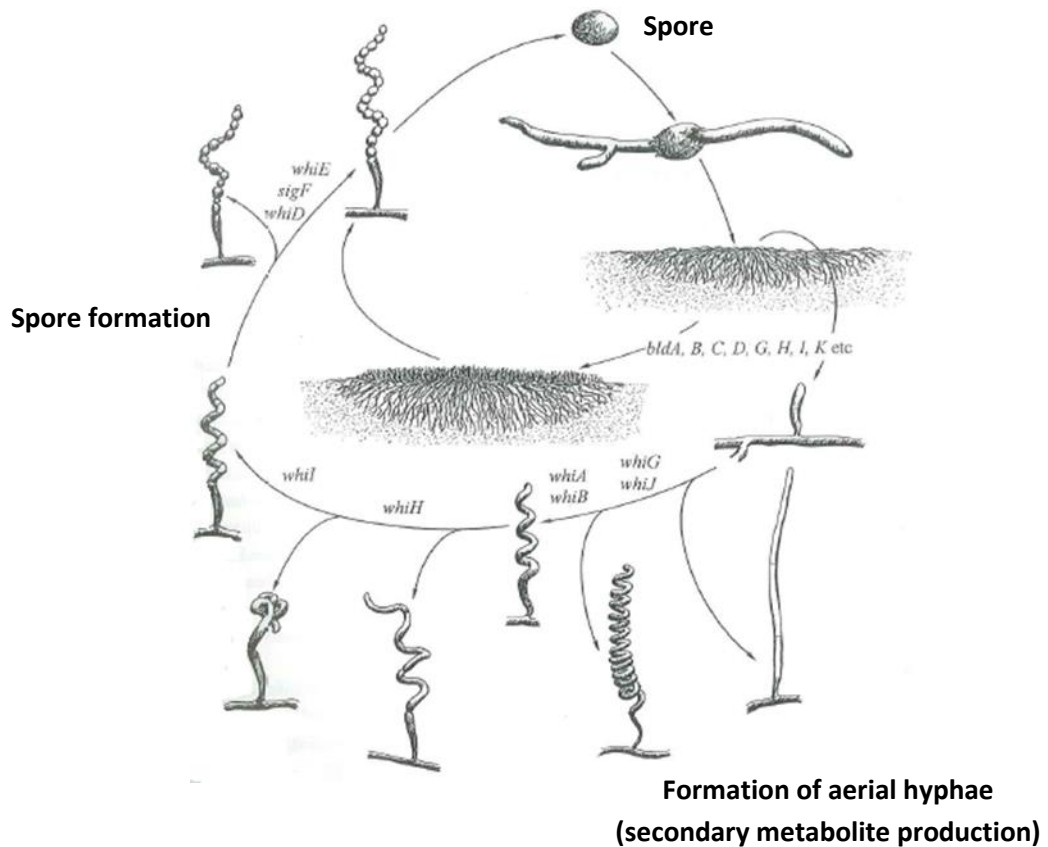


Figure 1.7 The life cycle of *Streptomyces coelicolor* (most streptomycetes have similar life cycles) (adapted from Kieser et al. 2000). The point in the life cycle where formation of aerial hyphae occurs (and production of secondary metabolites) has been labeled.

1.2.5 Targets for antimicrobial therapy

There are a number of antimicrobial targets in bacteria that are exploited for medicinal purposes. These include bacterial cell wall biosynthesis, the control of DNA topology and protein and nucleic acid synthesis (Fischbach and Walsh, 2009). A favoured target for antimicrobial action is the bacterial cell wall due to there being no homologous structure in humans and therefore targeting the cell wall biomachinery of a bacterial cell is less likely to have undesirable side-effects on the host.

Chapter 1 - Introduction

1.2.6 Peptidoglycan biosynthesis

The peptidoglycan layer of the cell wall of bacteria is an essential part of the cell. This mesh of amino sugars and pentapeptides protects the bacterial cell from bursting open by osmotic lysis. The layer also gives the cell its structural integrity, allowing for the shape of the cell to be maintained (Bouhss et al., 2004).

The peptidoglycan layer of the cell wall consists of two amino sugars: *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) linked by a β -1,4 linkage and crosslinked by a pentapeptide, giving a mesh-like structure (Bugg and Walsh, 1992). The pentapeptide sequence differs depending on whether the bacterium is a Gram positive or Gram negative species. The general consensus is that, for Gram negative bacteria such as *P. aeruginosa*, the pentapeptide sequence is: L-Alanine- γ -D-Glutamate-*meso*-diaminopimelic acid-D-Alanine-D-Alanine (L-Ala- γ -D-Glu-*m*-DAPA-D-Ala-D-Ala). The cross linking is achieved by the L-Ala residue of the pentapeptide bonding to the C3 of MurNAc and the penultimate D-Ala of the same pentapeptide becoming linked by an enamide bond to the *m*-DAPA residue of an adjacent pentapeptide (Bugg and Walsh, 1992).

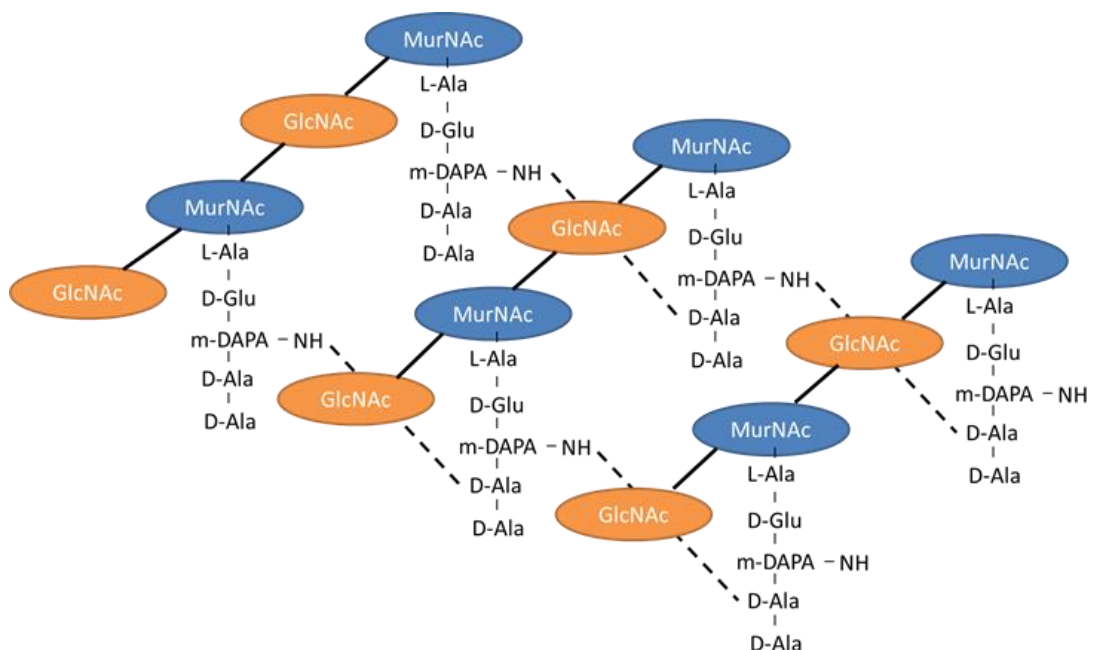


Figure 1.8 The organisation of the peptidoglycan layer of the cell wall of Gram negative bacteria. The glycan β -1,4 linkage is shown by the solid line, the peptide cross-link between the amino group of DAPA and D-Alanine shown by a broken line.

Chapter 1 - Introduction

Peptidoglycan biosynthesis involves various enzymes and takes place in two locations in the cell. Biosynthesis begins in the cytoplasm and finishes by forming the peptidoglycan layer on the outer edge of the periplasmic space. Biosynthesis begins with the production of the core components in the cytoplasm; UDP-GlcNAc and UDP-MurNAc. UDP-MurNAc is synthesised from the addition of an enolpyruvyl group to the 3'-hydroxy of UDP-GlcNAc by MurA, a cytoplasmic transferase. The pentapeptide is synthesised by the addition of amino acids to UDP-MurNAc, by a series of Mur ligases, at the expense of ATP (Bugg and Walsh, 1992). MurC, MurD and MurE add L-Ala, D-Glu and *m*-DAPA respectively, with MurF ligase linking the final two amino acids (already a dipeptide by the actions of a D-Ala-D-Ala ligase) to the growing chain to produce UDP-MurNAc-L-Ala-γ-D-Glu-*m*-DAPA-D-Ala-D-Ala (Ward, 1984, Bugg and Walsh, 1992). At this stage, UDP-MurNAc-pentapeptide needs to be transported across the cytoplasmic membrane, which is done by means of a undecaprenyl phosphate lipid carrier. UDP-MurNAc-pentapeptide is transferred to the lipid carrier by the transferase; translocase I, also known as MraY, forming lipid intermediate I (Struve et al., 1966, Bugg and Walsh, 1992). Another transferase, translocase II (MurG) catalyses the addition of the GlcNAc aminosugar to UDP-MurNAc-pentapeptide, which results in the formation of lipid intermediate II (Bugg and Walsh, 1992, Bupp and Vanheijenoort, 1993). At this stage, the intermediate is flipped, by an unknown mechanism, from the cytoplasm to the external surface of the cell (Marahiel et al., 1997). On the cell surface, lipid intermediate II is transglycosylated by members of the penicillin-binding family, leading to the polymerisation of the sugar backbone of the peptidoglycan layer (Bupp and Vanheijenoort, 1993). The mesh structure of the peptidoglycan layer is completed by peptide bond formation between the carbonyl group of the penultimate D-Ala residue of one pentapeptide with the terminal amino group of the *m*-DAPA residue of an adjacent pentapeptide (Bugg and Walsh, 1992).

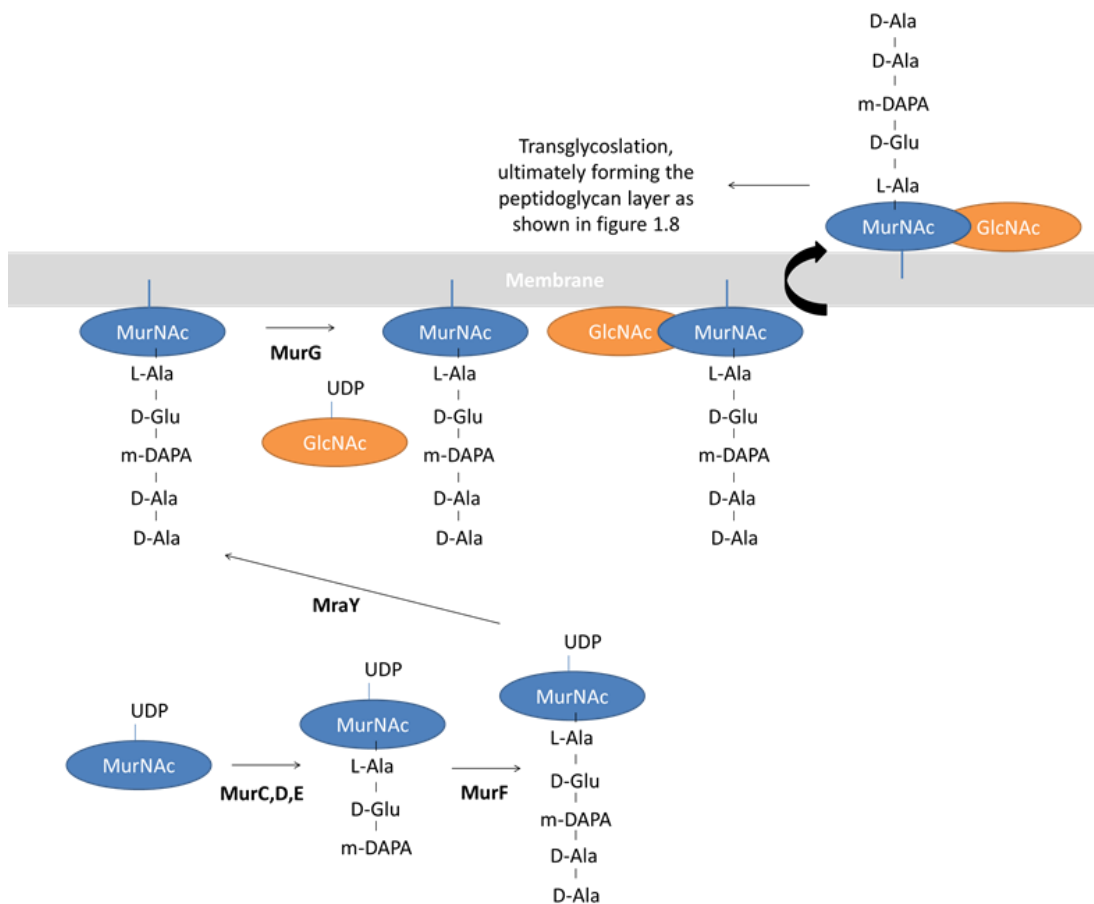


Figure 1.9 Outline of the initial stages of cell wall biosynthesis (prior to the transglycosylation).

1.2.7 Translocase I

As previously described, translocase I (MraY) is an enzyme involved in the transfer of UDP-MurNAc-pentapeptide to the undecaprenyl phosphate lipid carrier to form lipid intermediate I during peptidoglycan biosynthesis (Bugg and Walsh, 1992). For many years translocase I has been identified as a possible target for novel antibiotics for a number of reasons (Brandish et al., 1996). Firstly, it is known that translocase I plays an important role in the biosynthesis of the cell wall, a component of the bacterial cell which, if not present, would result in viability issues for the cell (Brandish et al., 1996). Secondly, the location of translocase I in the bacterial cell means accessibility by antimicrobial agents is relatively easy (Bouhss et al., 2004). Furthermore, there is no similar homolog within eukaryotes, suggesting activity against this enzyme would specifically target bacterial cells, without affecting any processes within the human patient (Bouhss et al., 2004). Nevertheless, to date this enzyme is clinically unexploited, despite the discovery of antibiotics, including the

Chapter 1 - Introduction

mureidomycins, liposidomycins and pacidamycins that have been found to have activity against this target (Inukai et al., 1993, Winn et al., 2010).

Studies of translocase I have elucidated its structures within *Escherichia coli* (*E. coli*) and *Staphylococcus aureus*. Both structures show strong homology, suggesting the structure of translocase I is well conserved between Gram positive and Gram negative bacteria (Bouhss et al., 2004). One of the more recent studies by Al-Dabbagh et al. (2008) describes the structure of this enzyme as consisting of ten transmembrane helices, five cytoplasmic loops and four periplasmic loops, with the N- and C- termini residing in the periplasm of the bacterial cell (Bouhss et al., 1999, Bouhss et al., 2004, Al-Dabbagh et al., 2008). From topological studies, it has been found that the five cytoplasmic loops are highly conserved, containing a total of 34 conserved amino acids, 19 of which are polar residues and 14 found to be essential for the activity of the enzyme (Bouhss et al., 2004, Al-Dabbagh et al., 2008).

1.2.8 Antibacterial cell wall targets

A number of antibiotics target the cell wall biomachinery (Ward, 1984). The most famous example is that of penicillin, which has activity against the transpeptidase by binding into its active site. Other examples include fosfomycin which inhibits the first enzyme in the pathway; MurA, and tunicamycin that shows activity against the cell wall enzyme MurG (Woodyer et al., 2006, Price and Tsvetanova, 2007).

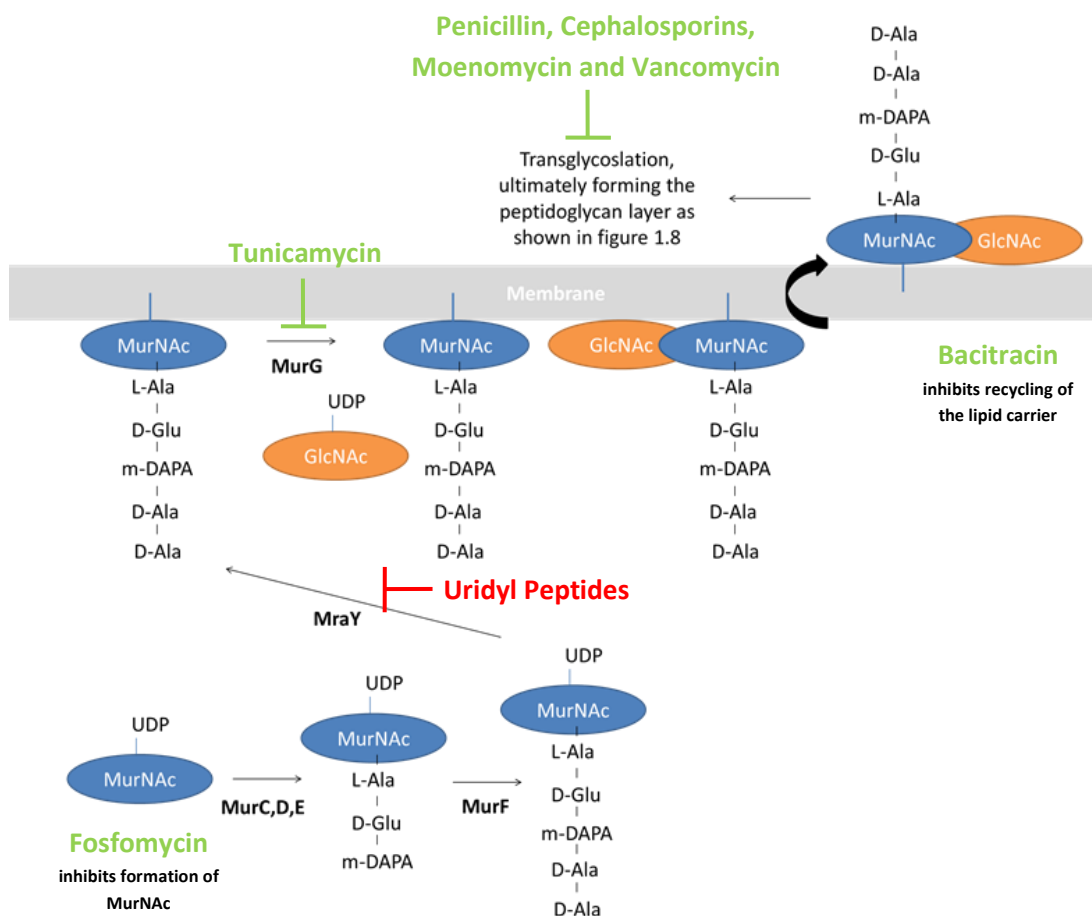


Figure 1.10 Diagrammatic representation of antibacterial targets in the cell wall biomachinery, also highlighting the target of the uridyl peptide class of metabolites (red).

1.3 Nucleoside antibiotics

1.3.1 General introduction

The nucleoside antibiotics are a group of antibiotics, all of which, as the name suggests, contain nucleoside residues (Kimura and Bugg, 2003). They inhibit the action of the cell wall enzyme translocase I, an enzyme encoded by *mraY* that catalyses an important stage in peptidoglycan biosynthesis and is as yet clinically unexploited. The antibacterial properties of the majority of the nucleoside antibiotics are non-specific, a property believed to be attributed to the fact that nucleoside metabolism is a primary process (Kimura and Bugg, 2003). A common and unfortunate characteristic of this family of antibiotics is that many are cytotoxic towards eukaryotic cells as well as prokaryotic cells, therefore making them

Chapter 1 - Introduction

redundant for clinical use (Nishimura et al., 1956). For example, the nucleoside antibiotic toyocamycin (Figure 1.11), one of the first in this class to be isolated, has good antibacterial activity against common pathogens such as *Mycobacterium tuberculosis*, but has been found to be toxic to eukaryotes during studies in mice. The toxicity is likely to be caused by its structural resemblance to adenosine, resulting in this anti-bacterial agent being inappropriate for clinical exploitation (Nishimura et al., 1956).

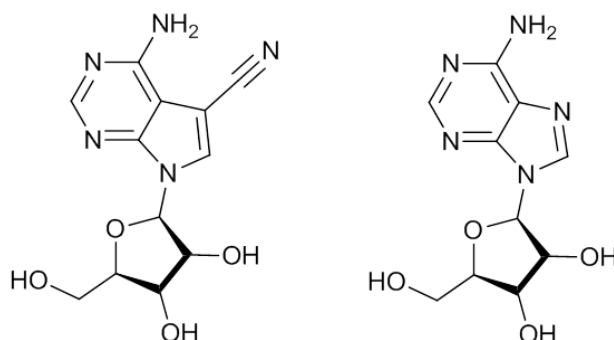


Figure 1.11 Structure of the nucleoside antibiotic toyocamycin (left), a nucleoside antibiotic found to be toxic in mice, its toxicity likely to be due to its striking resemblance to the ubiquitous nucleoside adenosine (right).

The nucleoside antibiotics are subdivided into three classes (Kimura and Bugg, 2003, Winn et al., 2010). The first class includes the natural products tunicamycins, streptoviridins and corynetoxins. These are characterised by the presence of a GlcNAc residue, O-linked sugar residues, a long fatty acid chain and the presence of the nucleoside uracil (Takatsuki et al., 1977, Eckardt, 1983, Winn et al., 2010). The second class are the structurally complex liposidomycins. They contain the aminosugar, aminoribose, the nucleoside, uridine, and fatty acyl units (Isono et al., 1985, Winn et al., 2010). The final class is the uridyl peptides and includes the napsamycins, sansamycins, mureidomycins and pacidamycins (Winn et al., 2010).

1.3.2 Discovery of the uridyl peptides

The first group of molecules discovered in the uridyl peptide class of metabolites were the mureidomycins and pacidamycins in 1989 when Chen et al. discovered pacidamycins 1-7

Chapter 1 - Introduction

and Isono et al. discovered mureidomycins A-D (Chen et al., 1989, Isono et al., 1989a). Another member of the class, the napsamycins were first isolated in 1994 and most recently the sansamycins were discovered (Chatterjee et al., 1994, Winn et al., 2010).

The uridyl peptides share a structurally common framework (shown in Figure 1.12) consisting of a central *N*-methyl (2*S*,3*S*)-diaminobutyric acid (DABA) residue, linked by a 4'-5' enamide bond to the nucleoside, 3'-deoxyuridine (Winn et al., 2010). The fourth amino acid in the peptide backbone is attached to the α -nitrogen of the DABA residue. An urea motif attaches to the fifth amino acid, which contains an aromatic side chain (Winn et al., 2010). Found linked to the $C\beta$ of DABA, depending on the natural product, is the *N*-terminal amino acid or dipeptide (Winn et al., 2010).

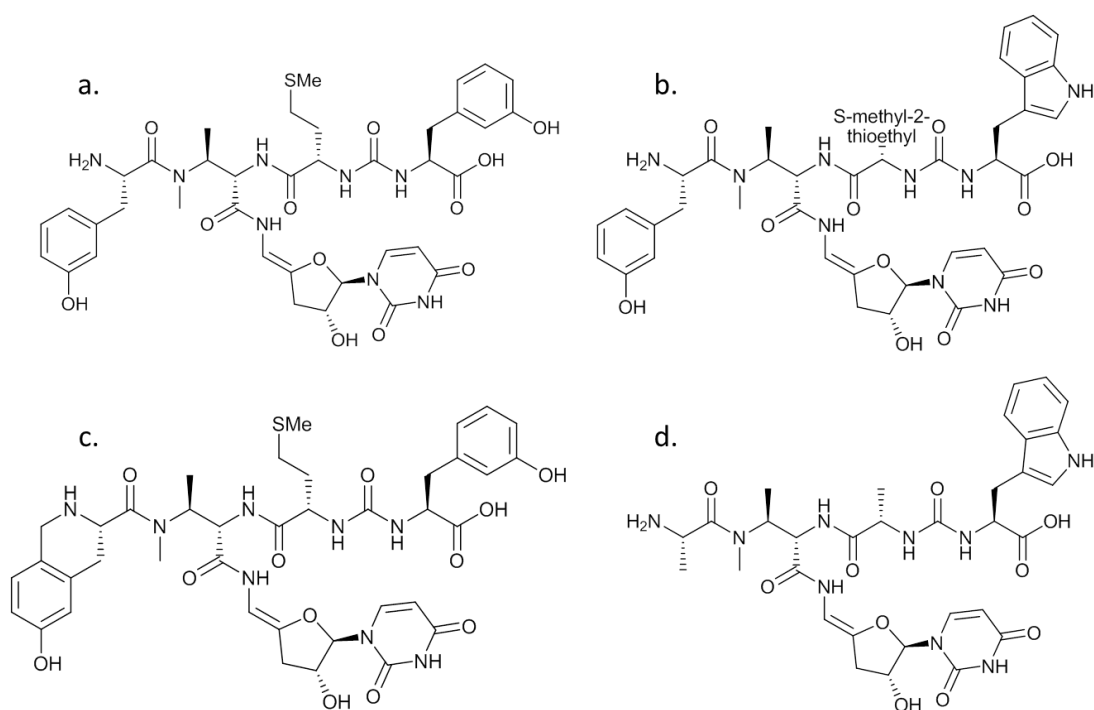


Figure 1.12 Diagrammatic representation of the structures of the different groups of uridyl peptides to re-iterate the significant similarities in their structures; (a) Mureidomycin A, (b) Sansamycin A, (c) Napsamycin A and (d) Pacidamycin D.

1.3.3 Uridyl peptide antibacterial activity

The uridyl peptides show a relatively narrow spectrum of activity. In general they show some activity against the *pseudomonads*, however, members of the class, such as the pacidamycins, only show activity against specific species within the pseudomonad family,

with the pacidamycins having activity *in vivo* against *Pseudomonas aeruginosa* only (Isono et al., 1989b, Isono et al., 1992, Fernandes et al., 1989). Studies have found that for the majority of *P. aeruginosa* strains, the pacidamycins have minimum inhibition concentrations (MIC) of between 0.1 and 100 $\mu\text{g mL}^{-1}$ depending on the functionality at the N- and C- termini of the peptide (Fronko et al., 2000). Studies into the use of the uridyl peptides as antibiotics have predominately been carried out using the mureidomycins. Mureidomycins A-D were found to protect mice against *P. aeruginosa* infection, however, parallel studies using the pacidamycins were not as successful (Fernandes et al., 1989, Isono et al., 1989a). Although it is stated that the pacidamycins show exclusive activity against *P. aeruginosa*, at high concentrations it has been found that pacidamycins can affect the growth of other bacteria, including members of the *Pseudomonas* and *Streptococci* genera and *Staphylococcus aureus* (Chen et al., 1989, Karwowski et al., 1989).

1.3.4 Uridyl peptides and their postulated inhibition mechanism

Studies into the inhibition mechanism of the uridyl peptides using mureidomycin A have suggested that they act as competitive inhibitors (Gentle and Bugg, 1999). It was postulated that the 4'5'-enamido and the nucleoside were important to the activity of these peptides (Gentle and Bugg, 1999). Further studies using a synthetic uridine analogue with the 4'5'-enamido was found not to inhibit translocase I (Gentle and Bugg, 1999). Later studies using pacidamycin D concluded the opposite, leaving a question mark over the necessity of this enamido (Winn et al., 2010). Other studies using mureidomycin A deduced that the N-terminal end of the peptide chain may have some significance to their ability to inhibit translocase I (Howard and Bugg, 2003). Studies showed that the N-terminus competes for the Mg^{2+} cofactor binding site of translocase I (Howard and Bugg, 2003). The core DABA residue has also been suggested to be an important component of these natural products, potentially improving their ability to prevent lysis of the peptide (Winn et al., 2010).

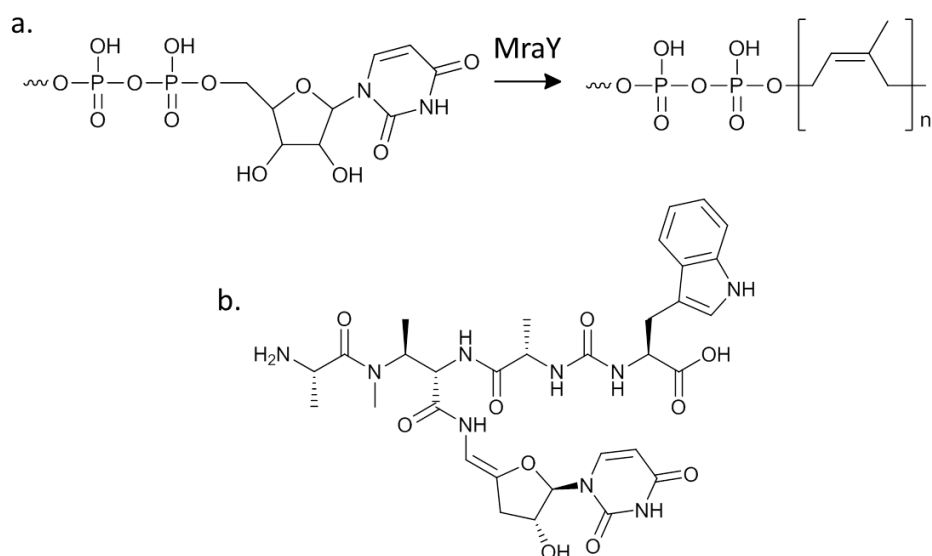


Figure 1.13 Showing the similarity of the reactive component of MrayY substrate (a) and the nucleoside moiety of the uridyl peptide (b) suggesting why the nucleoside was originally believed to be important in the antimicrobial activity of the uridyl peptides.

1.3.5 The pacidamycins

1.3.5.1 Introduction

The pacidamycins are uridyl peptide natural products produced by the bacterium *Streptomyces coeruleorubidus*. Pacidamycins 1 through to 7 were first isolated from *S. coeruleorubidus* strain AB 1183F-64 in 1989 by Chen et al. (Karwowski et al., 1989, Chen et al., 1989, Fernandes et al., 1989). Further pacidamycins were later isolated from *S. coeruleorubidus* strain NRRL 18730 and included pacidamycins D, 5T and 4N (Fronko et al., 2000). The structure of the pacidamycins has already been described. Figure 1.14 shows a more exhaustive list of pacidamycins. The total synthesis of pacidamycin D and a number of pacidamycin analogues has recently been reported by Okamoto et al. (Okamoto et al., 2011, Okamoto et al., 2012a, Okamoto et al., 2012b).

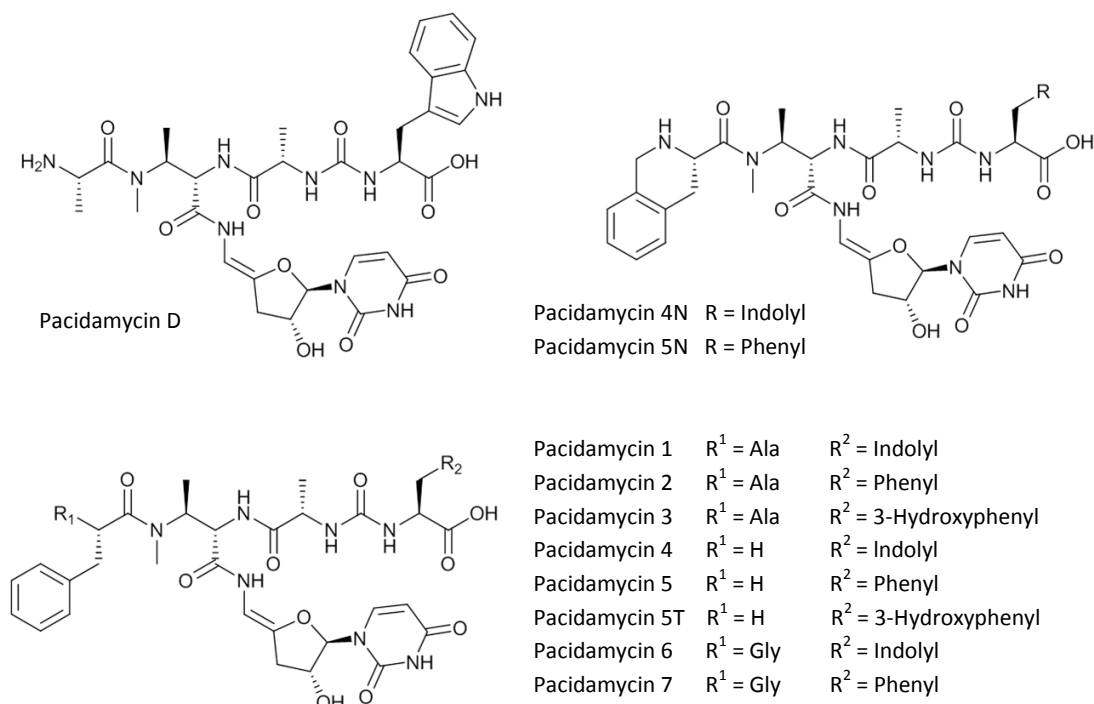


Figure 1.14 Diagram showing the structures of the naturally occurring pacidamycins.

1.3.5.2 The pacidamycin gene cluster

The pacidamycin gene cluster was the first uridyl peptide gene cluster to be characterised. The discovery was first reported by Rackham et al. (2010) and supported by the findings of Zhang et al. (2010) shortly after. Since 2010, gene clusters for napsamycins and sansamycins have also been identified (Kaysser et al., 2011, Wang et al., 2012). The pacidamycin gene cluster was identified using 454 sequencing which provided adequate genome coverage and 'systematic' gene disruption (using degenerate primers). By disrupting putative DABA synthases (six identified in total), Rackham et al. (2010) were able to determine which disruption eradicated pacidamycin production from *S. coeruleorubidus* (Rackham et al., 2010). They identified a 22 ORF gene cluster of the size 30.3 Kb that was believed to be responsible for the biosynthesis of pacidamycin. This hypothesis was confirmed by the heterologous expression of the gene cluster in a heterologous host (*S. lividans* TK24). Further analysis of the products of the heterologously expressed cluster showed that pacidamycin D and pacidamycin S were produced, but the most commonly produced pacidamycins that were usually generated by the native producer (pacidamycins 4, 4N, 5 and 5T) were not observed to be produced by the heterologous host. These

Chapter 1 - Introduction

pacidamycins contain the non-proteinogenic amino acid meta-tyrosine. It was postulated that the cluster is only the minimal pacidamycin gene cluster and that there were additional genes in the genome of *S. coeruleorubidus* that complement the minimal gene set to produce the suite of pacidamycins that are seen in the native producer (Rackham et al., 2010). Further to this, the minimal pacidamycin gene cluster showed significant similarity to a 25 ORF cluster found in *S. roseosporus* strain NRRL 15998/11379 (Rackham et al., 2010). Initial annotation of the 22 ORFs suggested there were three modular NRPSs at *pac12*, *pac14* (lacking an adenylation domain) and *pac16* (Rackham et al., 2010). Further NRPS domains were annotated at *pac15* and *21*, *pac4* and *9* and *pac8* being adenylation domains, condensation domains and acyl carrier proteins respectively (Rackham et al., 2010). Further to this *pac17* through *pac20* were believed to be responsible for the biosynthesis of the core DABA residue with *pac22*, the final ORF of the cluster catalysing the *N*-methylation of this residue (Rackham et al., 2010). Four hypothetical proteins were also annotated in the cluster as *pac1*, *2*, *7* and *13* (Rackham et al., 2010). Rackham et al. (2010) postulate that an ORF in the similar *S. roseosporus* homologous cluster encoding a phenyl alanine hydroxylase may have a homologue in *S. coeruleorubidus*, which was not present in the pacidamycin gene cluster, explaining the absence of production of any pacidamycin containing meta-tyrosine during heterologous expression (Rackham et al., 2010).

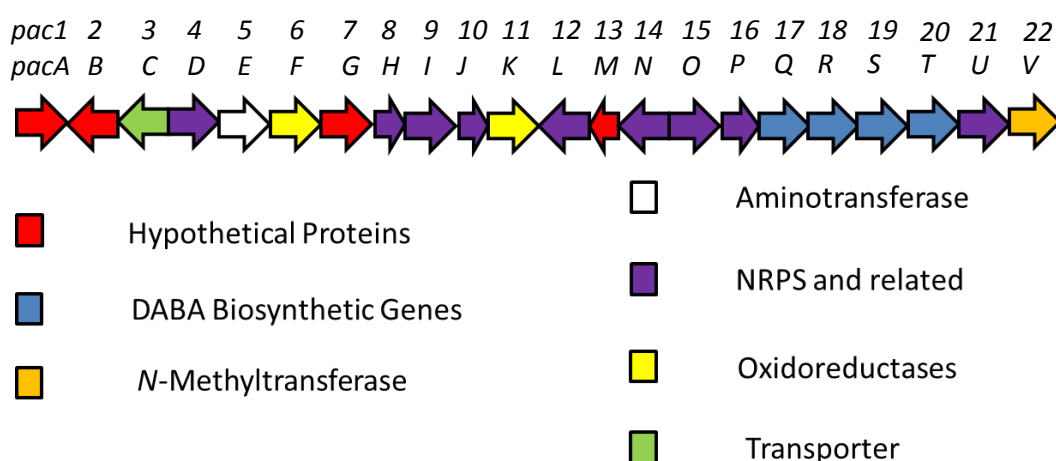


Figure 1.15 Illustration of the minimal pacidamycin gene cluster from *S. coeruleorubidus*. The genes have been coloured depending on their original functional designation, as established through sequence comparison to genes of known function in the database. Also included are both the Rackham et al. (2010) and Zhang et al. (2010) gene nomenclature.

1.3.5.3 Additional genes identified to be involved in pacidamycin biosynthesis

Further work reported by Grüşchow et al. (2011) identified the presence of a phenyl alanine hydroxylase (*phhA*) elsewhere in the *S. coeruleorubidus* genome that was responsible for the formation of the meta-tyrosine found at the N and C termini of a number of the pacidamycins. The work also identified a homologue of *pac21*, a gene in the minimal gene cluster of pacidamycin which was subsequently designated as *pac21h*. The paper reported that *pac21* was responsible for the activation and incorporation of alanine at the N terminus of the pacidamycin structure and *pac21h* was responsible for the activation and condensation of the meta-tyrosine produced by *phhA* at this position.

1.3.5.4 Further evidence in the assembly of pacidamycin

Work published by Zhang et al. (2011) showed that, *in vitro*, nine enzymes produced from the expression of genes in the pacidamycin gene cluster were essential for the assembly of pacidamycin. The work showed that Pac8 was responsible for the tethering of the core methylated non-proteinogenic amino acid DABA, the tethering being assisted by Pac16. DABA was methylated by Pac22 (DABA being tethered to Pac8 was not essential for its methylation). Pac10, 12 and 15 were found to be responsible for the assembly of the C-terminus of the peptide (ureido-dipeptide) and its assembly onto the DABA residue with the aid of Pac4. The N- terminus was incorporated into the structure by Pac21, and Pac9 was responsible for the addition of the nucleoside derivative.

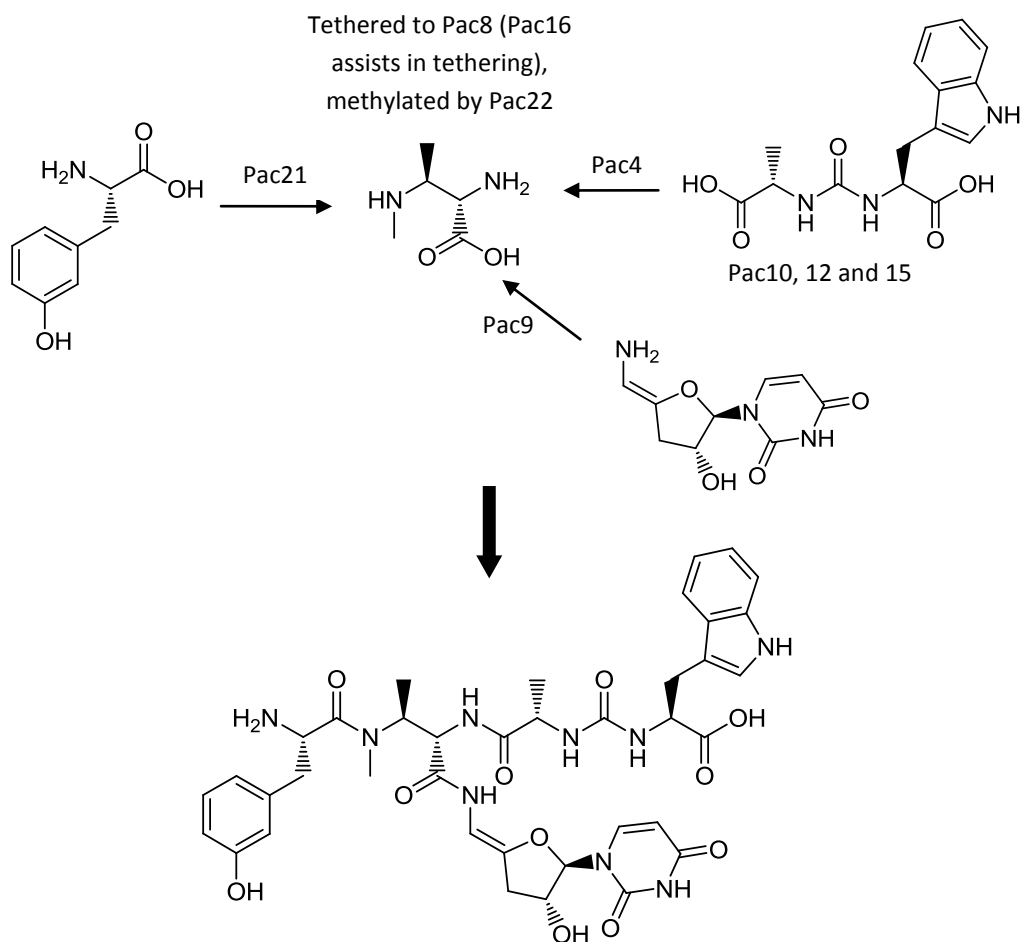


Figure 1.16 Diagrammatic representation of the enzymes involved in the assembly of pacidamycin.

1.4 Diamino acids and studies into their biosynthesis

Diamino acids have become of interest to biochemists in recent years due to their presence in many bioactive compounds. The simplest of these amino acids found in natural products is diaminopropionic acid (DAP) which has two enantiomeric forms and can be found in the tuberactinomycins and zwittermicin A. Diaminobutyric acid (DABA) has a total of four enantiomeric forms and is found in natural products such as the fruulimicins and the uridyl peptides. Even so, the full biosynthesis of these precursors of secondary metabolites are not fully understood. Some work has taken place into the biosynthesis of DAP (Carter et al., 1974). Studies by Carter et al. (1974) on the incorporation of diaminopropionic acid (DAP) into viomycin showed, by feeding experiments, that a variety of compounds could be used as a precursor for DAP, but the efficiency of incorporation of serine and glucose was ten to

Chapter 1 - Introduction

twenty fold higher (Carter et al., 1974). Suffice it to say, a likely candidate for a precursor of DABA would be threonine, an amino acid very similar to serine but with an additional methyl group, DABA having an additional methyl group compared to DAP. Bioassays conducted by Lam et al. (2008) showed that production of the antibiotic mureidomycin A in *Streptomyces flavidovirens* was significantly enhanced by the addition of 10 mM L-threonine or 100 mM L-phenylalanine to the growth media, however addition of uracil, L-methionine, meta-tyrosine, sodium acetate, or the D-enantiomers of phenylalanine or threonine, caused no increase in production (Lam et al., 2008). It was postulated that three possible pathways existed for the biosynthesis of DABA (Lam et al., 2008). The first is the condensation of glycine and acetyl coenzyme A by 2-amino-3-ketobutyrate ligase (KBL) and then transamination of the product (2-amino-3-ketobutyrate) to give DABA. A second plausible pathway is for L-threonine to be oxidised by threonine dehydrogenase, forming 2-amino-3-ketobutyrate followed by the transamination step. The final pathway would be via a β -replacement reaction, converting L-threonine to DABA in a PLP-dependent fashion (McGilvray et al., 1969, Lam et al., 2008). Lam et al. (2008) concluded that DABA is most likely to be biosynthesised via a PLP dependent β -replacement reaction of L-threonine. Their studies further showed that DABA is synthesised prior to the assembly of the antibiotic using UV/Vis assays of the phenylglyoxal derivative DABA (Lam et al. 2008). Absorbance (indicating presence of DABA) was highest two to three days after the beginning of growth of the native producer, where production of mureidomycin A was found to be highest at day six (Lam et al., 2008).

1.5 Proposed biosynthesis of DABA

Initial interrogation of the pacidamycin gene cluster by Rackham et al. (2010) suggests that four genes are responsible for the biosynthesis of the non-proteinogenic diamino butyric acid residue found at the core of the pacidamycin scaffold. These genes are *pac17* through *pac20*. Analysis of these genes suggest that *pac17*, *18* and *19* are translationally coupled and are annotated as an argininosuccinate lyase, a kinase and a DABA synthase, respectively, with the kinase requiring ATP as a substrate and the synthase requiring PLP as a cofactor. The final gene; *pac20* is annotated as being a PLP-dependent threonine aldolase (Rackham et al., 2010). From this evidence and work carried out by Lam et al. (2008) a mechanism has been postulated.

Chapter 1 - Introduction

It is believed that Pac20 is the link between primary and secondary metabolism. The enzyme is postulated to produce an excess of L-threonine, or produce threonine during the secondary metabolite production stage of the *Streptomyces* life cycle. It is believed that the threonine is phosphorylated at the hydroxyl position to produce O-phospho-L-threonine. This product is then utilized by the DABA synthase (Pac19) in a PLP-dependent β -replacement reaction, replacing the phosphate group for that of an amino group to produce (2S, 3S)-diaminobutyric acid (DABA), the amino group being supplied by the actions of Pac17 by breaking down a substrate such as aspartate to fumarate and ammonia. A diagrammatic representation of this postulated mechanism is shown in Figure 1.17.

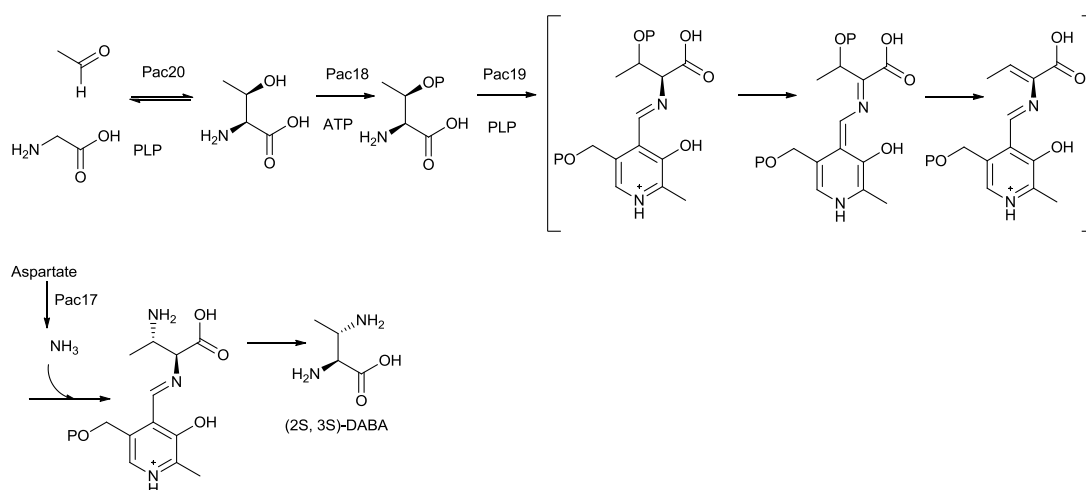


Figure 1.17 The postulated biosynthesis of DABA by the pacidamycin enzymes Pac17, 18, 19 and 20.

1.6 Project aims

This PhD project had two main aims:

1. To determine the mechanism by which the DABA moiety of pacidamycin is biosynthesised in *S. coeruleorubidus*.
2. To determine the functions of the four hypothetical proteins (*pac1*, *pac2*, *pac7* and *pac13*) that are present in the pacidamycin gene cluster.

Progress towards these aims will be reported in the subsequent chapters.

Materials and Methods

2.1 General

2.1.1 General comments

All components used for media and buffers, along with any biological or chemical reagents were purchased and supplied by commercial retailers and stored as advised by the manufacturer. All microorganism stocks were stored at -80°C . *Streptomyces sp.* were stored as spore suspensions in 25% (v/v) glycerol and *E.coli* strains were stored as cell suspensions in 25% (v/v) glycerol.

2.1.2 General reagents

Unless otherwise stated, all reagents were purchased from Sigma Aldrich.

1 Kb DNA gel marker – Promega

12 % Amersham precast protein gels – GE Healthcare

37 % Acrylamide – Severn Biotech Limited

Agarose – Melford

BigDye v 3.1 – Applied Biosystems

Carbenicillin disodium - Formedium

Chloramphenicol – Melford

dNTP's – Novagen

DAPase – Qiagen

DNA purification kit – Qiagen

Hi-trap chelating column – GE Healthcare

IPTG – Melford

Kanamycin sulfate – Melford

Pfu DNA polymerase – Invitrogen

Protease inhibitor cocktail tablets (EDTA free) – Roche

Chapter 2 – Materials and Methods

Instant Blue protein stain – Expedeon

Restriction enzymes – Roche / New England Biolabs

Taq polymerase - Roche

Thrombin – ICN

Tris base – Melford

2.1.3 General equipment

Äkta FPLC and Äkta FPLC express – GE Healthcare

Bench top centrifuge – Sorvall Legend RT (Heraeus rotor 75006445)

Cell disruptor – Constant Cell Disruption Solutions

Cell electroporator – Biorad Gene Pulser

Centrifuge – Sorvall Evolution RC

Crystallisation supplies – Hampton Research, Molecular Dimensions and Qiagen

Incubators (shaking) – New Brunswick Scientific and Infors HT

Incubators (static) – LTE

LC-MS – Agilent 1100 series for general LC-MS analysis and Thermo Scientific Finnigan Surveyor for LC-MS₂ analysis

Microcentrifuge – Heraeus Instumentation (rotor PP 1/97 #3324)

Microscope – Nikon SMZ800

pH meter – Hanna Instruments

Pipettes – Anachem

Protein casting apparatus and tank – Biorad mini protean II or GE Healthcare Amersham ECL tank

Sonicator – Sonics Vibra Cell

Chapter 2 – Materials and Methods

Spectrophotometer – Perkin Elmer Lambda 25

Thermal Cyclers – Eppendorf Mastercycler and MJ Research PTC 200 (gradient PCR)

Transilluminator – Uvitec

Vortex – Stuart Autovortex Mixer

Water bath – Grant

2.1.4 Medium

Components to produce medium were purchased from retailers including Fisher Scientific, Sigma-Aldrich, BD Biosciences, Melford, Formedium, and Alfa-Aesar unless stated otherwise.

Table 2.1 Recipes for all medium (solid and liquid) used during this study

Medium	Components (L ⁻¹)
Luria-Bertani (LB)	10 g tryptone, 5 g yeast extract, 10 g NaCl
Lennox (Len)	10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 g glucose
Auto-Induction Media LB broth based (AIM)	10 g tryptone, 5 g yeast extract, 3.3 g (NH ₄) ₂ SO ₄ , 6.8 g KH ₂ PO ₄ , 7.1 g Na ₂ HPO ₄ , 0.5 g Glucose, 2 g α-lactose, 0.15 g MgSO ₄ , 0.03 g trace elements
ISP-2	4 g yeast extract, 10 g malt extract, 4 g glucose, pH 7.2 using NaOH
Soy Flour Mannitol (SFM)	20 g mannitol, 20 g soy flour, tap water
Tryptone Soy Broth (TSB)	30 g tryptone soya broth powder
SOB	20 g tryptone, 5 g yeast extract, 0.5 g NaCl
Difco Nutrient Agar (DNA)	23 g Difco Nutrient Agar
INA5 (modified)	15 g soya flour, 5 g CaCO ₃ , 2 g NaCl, 30 mL glycerol
Minimal Media (MM)	2.14 g NH ₄ Cl, 0.6 g MgSO ₄ , 4.41 g K ₂ HPO ₄ , 20.99 g MOPS, 0.2 mL 50 mg mL ⁻¹ FeSO ₄ ·2H ₂ O, 0.2 mL 50 mg mL ⁻¹ MnCl ₂ ·4H ₂ O, 0.2 mL 50 mg mL ⁻¹ ZnSO ₄ ·7H ₂ O, 0.2 mL 50 mg mL ⁻¹ CaCl ₂ , made upto 950 mL with distilled water at pH 7.0

Chapter 2 – Materials and Methods

Maltose Solution	20% (m/v) maltose, dissolved at 40°C
R5	103 g sucrose, 0.25 g K ₂ SO ₄ , 10.1 g MgCl ₂ ·6H ₂ O, 10 g glucose, 0.1 g Difco aminoacids, 2 mL trace element solution(40 mg ZnCl ₂ , 200 mg FeCl ₃ ·6H ₂ O, 10 mg CuCl ₂ ·2H ₂ O, 10 mg MnCl ₂ , 10 mg Na ₂ B ₄ O ₇ ·10H ₂ O, 10 mg (NH ₄) ₆ Mo ₇ O ₂₄ ·4H ₂ O in 100 mL), 5 g Difco yeast extract, 5.73 g TES buffer, 10 mL KH ₂ PO ₄ (0.5 %), 4 mL CaCl ₂ ·2H ₂ O (5 M) 15 mL L-proline (20 %) 7 mL NaOH(1 N), made up to 1000 mL with distilled water

All medium were made up to the required volume with distilled water unless otherwise stated, followed by autoclaving at 121 °C, 1.3 bar for 20 min to ensure the medium were sterile. Heat-sensitive components were sterilised by filtration through a 0.2 µm membrane. If the medium were required in solid form, agar was introduced at 1.5 % (w/v) for standard media, 0.5 % (w/v) for soft medium and 2 % (w/v) for SFM agar.

2.1.5 Antibiotics

Table 2.2 Recipes for antibiotics used in this study

Antibiotic	Stock solution (mg mL ⁻¹)	Solvent	Concentration in media (µg mL ⁻¹)
Kanamycin sulfate (kan)	50	H ₂ O	50
Carbenicillin sodium salt (carb)	100	H ₂ O	100
Tetracyclin hydrochloride (tet)	5	70 % ethanol	5
Apramycin sulphate (apr)	50	H ₂ O	50
Chloramphenicol (chl)	25	Ethanol	25
Hygromycin B (hyg)	100	H ₂ O	25 - 100
Naladixic acid sodium salt (nal)	25	H ₂ O	25

All antibiotic solutions other than chl were sterilised by filtration through a 0.2 µm membrane. As chl was in ethanol no sterilisation was required. Stocks were stored at – 20 °C in 1.5 mL aliquots.

Chapter 2 – Materials and Methods

2.1.6 Buffers and stock solutions

Components to produce buffers and stock solutions were purchased from retailers including Fisher Scientific, Sigma-Aldrich and Alfa-Aesar unless stated otherwise.

Table 2.3 Recipes of general reagents used in this study

Buffer/stock solution	Components
DNA preparation	
Sodium dodecyl sulphate (SDS)	2 (w/v) SDS, dissolved at room temperature in distilled water
Alkaline lysis solution I	50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 $\mu\text{g mL}^{-1}$ RNase A
Alkaline lysis solution II	0.2 N NaOH, 1 % (w/v) SDS
Alkaline lysis solution III	3 M potassium acetate, 11.5 % (v/v) acetic acid
Tris-EDTA (TE)	10 mM Tris-HCl pH 8.0, 1 mM EDTA
DNA gel electrophoresis	
TAE 50x	2 M Tris-base pH 8.0, 1 M glacial acetic acid, 0.05 M EDTA, autoclaved.
Preparation of competent cells	
KMES buffer	20 mM KMES pH 5.8, 60 mM CaCl_2 , 5 mM MgCl_2 , 5 mM MnCl_2 , 10 % (v/v) glycerol, autoclaved.
Cell lysis	
Cell lysis buffer	50 mM Tris-HCl pH 8.0, 25 mM NaCl, 2.5 mM EDTA, 5 % (v/v) glycerol
Protein purification	
FPLC loading buffer (A)	50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 40 mM imidazole
FPLC elution buffer (B)	50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 0.5 M imidazole

Chapter 2 – Materials and Methods

Gel filtration buffer (GF)	20 mM Hepes pH 7.5, 150 mM NaCl
SDS polyacrylamide gel electrophoresis	
SDS PAGE solution A	30 % (w/v) acrylamide, 0.8% (w/v) bisacrylamide
SDS PAGE solution B	0.4 % (w/v) SDS, 1.125 M Tris-HCl pH 8.8
SDS PAGE loading dye (2x)	100 mM Tris-HCl pH 6.8, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 % (v/v) glycerol, 200 mM dithiothreitol
SDS PAGE running buffer	250 mM glycine, 25 mM Tris-base, 3 % (w/v) SDS
Liquid chromatography - mass spectrometry	
LC-MS buffer A	0.1 % (v/v) formic acid in H ₂ O
LC-MS buffer B	0.1 % (v/v) formic acid in acetonitrile

2.1.7 Plasmids

Table 2.4 List of plasmid names and genotypes used in this study

Plasmid	Relevant genotype	Supplier
Cloning and protein expression		
pLysS	<i>LysS cat</i>	Invitrogen
pLysE	<i>LysE cat</i>	Invitrogen
pGroESL	<i>GroESL tet</i>	Takara
pET28a(+)	<i>oriT-bla carb</i>	Invitrogen
pET21a(+)	<i>oriT-carb</i>	Invitrogen
Gene disruption and heterologous expression		
pIJ773	<i>oriT-acc(3)IV</i>	JIC
pIJ790	λ -RED (<i>gam bet exo</i>) <i>cat araC repA101ts</i>	JIC
pUZ8002	<i>neo tra</i>	JIC
BT340	<i>F-</i> , Δ (<i>argF-lac</i>)169, ϕ 80 <i>dlacZ58(M15)</i> , <i>glnV44(AS)</i> , λ , <i>rbcC1</i> , <i>gyrA96(NalR)</i> , <i>recA1</i> , <i>endA1</i> , <i>spoT1</i> , <i>thi-1</i> , <i>hsdR17</i> , <i>pCP20</i>	JIC

Chapter 2 – Materials and Methods

Expression constructs		
pDT001	pET28a(+) <i>pac1</i>	This project
pDT002	pET28a(+) <i>pac2</i>	This project
pDT003	pET28a(+) <i>pac7</i>	This project
pDT004	pET28a(+) <i>pac13</i>	This project
pDT005	pET28a(+) <i>pac17</i>	This project
pDT006	pET28a(+) <i>pac18</i>	This project
pDT007	pET28a(+) <i>pac19</i>	This project
pDT008	pET28a(+) <i>pac20</i>	This project
pDT009	pET21a(+) <i>pac17</i>	This project
pDT010	pET21a(+) <i>pac19</i>	This project
pJC001	pET28a(+) <i>pac22</i> (start 31098)	Mr J. Clouston
pJC002	pET28a(+) <i>pac22</i> (start 31125)	Mr J. Clouston
pDT011	pET28a(+) <i>pac17</i> (Arg108 – Ala108)	This project
pDT012	pET28a(+) <i>pac17</i> (Asn109 – Ala109)	This project
pDT013	pET28a(+) <i>pac17</i> (His155 – Ala155)	This project
pDT014	pET28a(+) <i>pac17</i> (Met279 – Ala279)	This project
pDT015	pET28a(+) <i>pac17</i> (Lys282 – Ala282)	This project
pDT016	pET28a(+) <i>pac17</i> (Asn284 – Ala284)	This project

pLysS, pLysE and pGroESL were used to attempt to improve the expression of recombinant protein in the heterologous *E. coli* host. pLysS and pLysE both encode for T7 lysozyme which is a natural inhibitor of T7 RNA polymerase allowing for tighter control of protein expression (reduces 'promotor leakiness'). pLysE produces larger amounts of T7 lysozyme in comparison to pLysS and therefore offers tighter control of recombinant protein expression. pGroESL encodes for chaperone proteins that aid in the correct folding of the recombinant protein in the heterologous host. pIJ773 was used as a template for amplification of the gene replacement cassette, pIJ790 in the homologous recombination of DNA and pUZ8002 in the conjugal transfer of cosmids from *E. coli* to *Streptomyces sp.*

Chapter 2 – Materials and Methods

2.1.8 Cosmids

Table 2.5 List of cosmid names and genotypes used in this study

Cosmid	Relevant genotype	Supplier
pIJ10702	Supercos I $\Delta_{neo}::aac(3)IV-oriTint\Phi C31-attP\Phi C31$	JIC
2H-5	<i>cos neo bla_{carb}</i>	UEA
8G-5	<i>cos neo bla_{carb}</i>	UEA
Gene disruption and heterologous expression		
2H-5- $\Delta pac7$	$\Delta pac7::aac(3)IV-oriT$	This project
2H-5- $\Delta pac13$	$\Delta pac13::aac(3)IV-oriT$	This project
2H-5- $\Delta pac19$	$\Delta pac19::aac(3)IV-oriT$	This project
2H-5- $\Delta pac20$	$\Delta pac20::aac(3)IV-oriT$	This project
2H-5- $\Delta pac17-pac20$	$\Delta pac17-pac20::aac(3)IV-oriT$	This project
2H-5-integration	2H-5 $\Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project
2H-5- $\Delta pac17$ -integration	2H-5 $\Delta pac17 \Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project
2H-5- $\Delta pac18$ -integration	2H-5 $\Delta pac18 \Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project
2H-5- $\Delta pac19$ -integration	2H-5 $\Delta pac19 \Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project
2H-5- $\Delta pac20$ -integration	2H-5 $\Delta pac20 \Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project
2H-5- $\Delta pac17-20$ -integration	2H-5 $\Delta pac17-pac20 \Delta_{neo}::aac(3)IV-oriT-int\Phi C31-attP\Phi C31$	This project

The pIJ10702 cosmid contains a streptomycete integration backbone which was used for homologous recombination into cosmid 2H-5 allowing for the heterologous expression of the genes on this cosmid. Cosmids 8G-5 and 2H-5 were created as part of a cosmid library produced by E. Rackham (UEA) and contain the minimal pacidamycin gene cluster (*pac1-pac22*). These cosmids were used as template DNA for the amplification of the *pac* cluster genes and also for heterologous expression of the *pac* gene cluster.

Chapter 2 – Materials and Methods

2.1.9 Microorganisms

Table 2.6 List of microorganism strains and their genotypes used in this study

Strain	Genotype	Supplier
Cloning and protein expression		
<i>E. coli</i> XL1-Blue-MR	$\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173$ <i>endA1supE44 recA1 lac</i>	UEA
<i>E. coli</i> DH5 α	F ⁻ <i>glnV44 thi-1 endA1 recA1 relA1 deoR gyrA96</i> <i>deoR λ80dlaclacZ $\Delta(lacIZYA-argF) \Delta M15$</i> <i>U169hsR17(rk⁻rm⁺)</i>	Invitrogen
<i>E. coli</i> BL21 (DE3)	F ⁻ <i>ompT gal dcm lon hsdS_B (r_B⁻m_B⁻) lacI lacUV5</i> with DE3 λ prophage	UEA
<i>E. coli imp⁻</i>	<i>imp⁻</i>	JIC
Gene disruption and heterologous expression		
<i>E. coli</i> ET12567	<i>dam-13::Tn9 dcm-6 hsdM hsdR ara14I</i>	UEA
<i>E. coli</i> BW25113	<i>lacIq rrnBT14 lacZWJ16 hsdR514</i> <i>araBADAH33 rhaBADLD78</i>	UEA
<i>S. coeruleorubidus</i>	Wild-type	UEA
<i>S. lividans</i> TK-24	<i>str-6 SLP2⁻ SLP3⁻</i>	UEA
<i>S. coeruleorubidus</i> DT-001	<i>2H-5Δpac7::aac(3)IV-oriT</i>	This project
<i>S. coeruleorubidus</i> DT-002	<i>2H-5Δpac13::aac(3)IV-oriT</i>	This project
<i>S. coeruleorubidus</i> DT-003	<i>2H-5Δpac19::aac(3)IV-oriT</i>	This project
<i>S. coeruleorubidus</i> DT-004	<i>2H-5Δpac20::aac(3)IV-oriT</i>	This project
<i>S. coeruleorubidus</i> DT-005	<i>2H-5Δpac17-pac20::aac(3)IV-oriT</i>	This project
<i>S. lividans</i> TK-24 DT-006	Cosmid 2H-5- integrated	This project
<i>S. lividans</i> TK-24 DT-007	Cosmid 2H-5 Δ pac17 Δ neo::aac(3)IV-oriT- <i>intΦC31-attPΦC31-</i> integrated	This project
<i>S. lividans</i> TK-24 DT-008	Cosmid 2H-5 Δ pac18 Δ neo::aac(3)IV-oriT- <i>intΦC31-attPΦC31-</i> integrated	This project
<i>S. lividans</i> TK-24 DT-009	Cosmid 2H-5 Δ pac19 Δ neo::aac(3)IV-oriT- <i>intΦC31-attPΦC31-</i> integrated	This project
<i>S. lividans</i> TK-24 DT-010	Cosmid 2H-5 Δ pac20 Δ neo::aac(3)IV-oriT- <i>intΦC31-attPΦC31-</i> integrated	This project
<i>S. lividans</i> TK-24 DT-011	Cosmid 2H-5 Δ pac17-pac20 Δ neo::aac(3)IV- <i>oriT-intΦC31-attPΦC31-</i> integrated	This project

Chapter 2 – Materials and Methods

2.2 Basic molecular biology methodology

2.2.1 Primer design

Specific primers are needed for the amplification of a desired gene/DNA of interest. Primers were designed by using the first 20 nucleotides (forward primer) and last 20 complementary nucleotides (reverse primer) of the sequence of interest. Suitable restriction sites were added to the 5' ends of the sequences that were compatible both with the insert (same restriction site not present in gene) and with the vector (restriction site present in the multiple cloning site of the plasmid). The melting temperature (T_m) and base stacking values for complementary primers were compared using internet server PROMEGA BIOMATH (www.promega.com/techserv/tools/biomath/calc11.html) and the number of nucleotides in the primer increased or decreased to allow the forward and reverse primer sequences to have similar T_m values. Primers were ordered from Sigma-Aldrich, they arrived desalted and were made up to a concentration of 100 μ M.

2.2.2 Polymerase chain reaction

2.2.2.1 Primer sequences

Table 2.7 List of all primers and their sequences used in this study

Primer Name	Sequence (5'-3')
1) Gene cloning and protein expression	
Pac1F	GGAGATCATATGCACACCAAGATCGAAGCCGAT
Pac1R	CGACTCGAGTCACGAAGGCTGAGCCCTCGA
Pac2F	ACCGGATCCATGGCTATTGGTTTTACCTCGGCTATC
Pac2R	GCAAAGCTTATGATTCGCTACAGGCAAGTC
Pac7F	GGAGATCATATGGCGCAGGTTCTAGCAGAAGC
Pac7R	CTCGAGTCAGTCGAGCCAGGTCAGGAAGT
Pac13F	AGAGCTCATATGACCAAGTACAAATACACACAG
Pac13R	GCTCTCGAGTAGTAAGGGCTCTCGCTTTC
Pac17F	GGACGACATATGGTGAGACTGACCGGTCGACTT
Pac17R	TCCGGATCCTCGTCAGACTCCCCCGG
Pac18F	GGACATATGACGAGCCGGCAGCTC
Pac18R	TCCGGATCCTCATTGAAGATCATTACTTGGTCCG

Chapter 2 – Materials and Methods

Pac19F	CGAGGACATATGATCTTCAATGACCTTGTCGA
Pac19R	TCCGGATCCCCTTGCTCAACCGCCAACC
Pac20F	AGGATCCATATGGCTGACCTGATCGAAATGAG
Pac20R	TCCCTCGAGTCAGCTCGCCTCGCGTAC
2) Site directed mutagenesis	
Pac17_108Arg/Ala_F	GACCGGAGCGCCAACGACCTC
Pac17_108Arg/Ala_R	GAGGTCGTTGGCGCTCCGGTC
Pac17_109Asn/Ala_F	CCGGAGCCGCGCCGACCTCCAGGC
Pac17_109Asn/Ala_R	GCCTGGAGGTCGGCGCGGCTCCGG
Pac17_155 His/Ala_R	CCGGGTACACGGCCCTCCAGTCGGC
Pac17_155His/Ala_R	GCCGACTGGAGGGCCGTGTACCCGG
Pac17_279Met/Ala_F	CTCCGCGGCCGCGCCCCAGAAGAAG
Pac17_279Met/Ala_R	CTTCTTCTGGGGCGCGGCCGCGGAG
Pac17_282Lys/Ala_F	CCATGCCCCAGGCCAAGAACTACC
Pac17_282Lys/Ala_R	GGTAGTTCTTCGCTGGGGCATGG
Pac17_284Asn/Ala_F	CCCAGAAGAAGGCCTACCCGCTGC
Pac17_284Asn/Ala_R	GCAGCGGGTAGGCCTTCTTCTGGG
Pac17_276Ser/Ala_F	GGC CGGCATCGCC GCGGCCATGC
Pac17_276Ser/Ala_R	GCATGGCCGCGGCGATGCCGGCC
3) Gene disruption	
Redirect_Pac1_F	CCACAAGTGGTCGAGGAAGACAGGAAGACGCTCTTCATGATTCCGG GGATCCGTCGACC
Redirect_Pac1_R	CACCGGAAGCGGCCGCACGGCGGGAGGCCGTCACTTCATGTAGGC TGGAGCTGCTTC
Redirect_Pac2_F	CTCGCTGGTTGGACACTAGATCATGAGGGCTGATACATGATTCCGGG GATCCGTCGACC
Redirect_Pac2_R	CCCGGTGGTGCCGGCGGGTACCAGCCGCCTCCCTTATGTAGGCT GGAGCTGCTTC
Redirect_Pac7_F	CTGGATCGCCAGCATCCAGAAAGGGAGATACAGGCCATGATTCCGG GGATCCGTCGACC
Redirect_Pac7_R	GGGCACCCGTCCTTGGACGCGAGGCCGCGGCCGGGGTCATGTAGGC TGGAGCTGCTTC

Chapter 2 – Materials and Methods

Redirect_Pac17_F	CTCATGGCGTGGTCGTCGAGCGCCTACGGGTTGGTCGAGATTCCGG GGATCCGTCGACC
Redirect_Pac17_R	GGTGCGAAGTCGACGTAGAAGGCAGTCAGGTGGCCGGTTGTAGG CTGGAGCTGCTTC
Redirect_Pac18_F	GGCGGCGGCGGATCGCTGAGGCTGAGCGGCTCGCTCCCGATTCCGG GGATCCGTCGACC
Redirect_Pac18_R	CTCGATCCGACGCAGCAGGTTCTCGATGCCCTGCGGCGGTGTAGGCT GGAGCTGCTTC
Redirect_Pac19_F	GCCAGGTCATTCGCGAGGCTCGCGAGAACGGGGTGCTGGATTCCGG GGATCCGTCGACC
Redirect_Pac19_R	TCGCCCCGCCTACGCGCGATCTGGACTTCGGCGACACCGTGTAGGCT GGAGCTGCTTC
Redirect_Pac20_F	ATGGCTGACCTGATCGAAATGAGAAGCGACACCTTCACGATTCCGG GGATCCGTCGACC
Redirect_Pac20_R	CCTCCTGTATTTGAAGAACCATCGGGATTGCGCTTCATGTAGGCT GGAGCTGCTTC
Redirect_Pac17-20F	GGTGCCGGCGCCGCTCTCTCCGACATCGCGCTGGCGATGATTCCGG GGATCCGTCGACC
Redirect_Pac22_F	ACCAGGCATTTTTTCTGATCGAGAAAGGTGACGGGATGATTCCGGG GATCCGTCGACC
Redirect_Pac22_R	CTACCGAAGGGCCGGACGACGAAAGCCGGTACCGATCTATGTAGGC TGGAGCTGCTTC
Pac13+100F	TGCTTCGAGTTCTCGTCGC
Pac13+100R	CGACATGCCATCCTCCATCCA
Pac17-20+100F	GGTTCGAGGACGCGGGTCGCT
Pac17-20+100R	CCCGTCAGGTGCATCGGCCTC
φC31_F	GAAGCGTTTTTCGGGAGTAGT
φC31_R	CGTTCATCCACATGGACCAGA
Pac17R_GC	TCCTTAATTAATCGTCGACTCCCCCGG

Chapter 2 – Materials and Methods

2.2.2.2 Optimisation of PCR annealing temperature

A 50 μL PCR master mixture was prepared consisting of:

H ₂ O	36.5 μL
Pfu Buffer (10 x)	5 μL
25 mM dNTPs	4 μL
Template DNA	0.5 μL
Forward primer	0.5 μL
Reverse primer	0.5 μL
DMSO	2.5 μL
Pfu polymerase	0.5 μL

11 μL of the master mixture was aliquoted into PCR tubes, giving 4 identical reaction mixtures. The reactions were placed in a thermal cycler and exposed to the following cycler conditions:

Temperature	Time (min)
94 °C	2
94 °C	1
55 – 69 °C gradient	1
72 °C	2 per 1 kb DNA
Repeated 29 times	
72 °C	10
4 °C	Hold

Once the primer annealing temperatures were optimized, the desired gene was amplified by PCR using the protocol above, but by making a single 25 μL reaction and using the optimum annealing temperature.

2.2.2.3 Colony PCR using *E.coli*

As proof reading activity is not necessary for verifying correct insert size, a Biomix ready reaction (Biolabs Ltd) was used.

Chapter 2 – Materials and Methods

The master mixture contained:

H ₂ O	19 µL
Biomix	25 µL
Plasmid preparation	2.5 µL
Forward primer	0.5 µL
Reverse primer	0.5 µL
DMSO	2.5 µL

The reaction mixture was placed in a thermal cycler and exposed to the following cycler conditions:

Temperature	Time (min)
94 °C	2
94 °C	1
55 °C	1
72 °C	1 per 1 kb DNA
Repeated 29 times	
72 °C	10
4 °C	Hold

2.2.2.4 Colony PCR using *Streptomyces sp.*

A single colony was added to 50 µL of DMSO and vortexed for 30 s. The sample was boiled for 10 min, cooled and vortexed for a further 30 s. A colony PCR reaction was set up using:

H ₂ O	7.6 µL
Biomix	10 µL
Colony/DMSO mixture	1 µL
Forward primer	0.5 µL
Reverse primer	0.5 µL
DMSO	0.4 µL

Chapter 2 – Materials and Methods

The reactions were placed in a thermal cycler and exposed to the following cycler conditions:

Temperature	Time (min)
94 °C	5
94 °C	1
55 °C	1
72 °C	1 per 1 kb DNA
Repeated 29 times	
72 °C	10
4 °C	Hold

2.2.3 Restriction, ligation and cloning into the pET vector system

pET28a(+) (obtained from a plasmid preparation of *E. coli* XL1-Blue containing pET28a(+)) and the desired PCR amplified gene were digested using restriction enzymes at 37 °C overnight to obtain compatible ends. The restriction mixtures consisted of:

H ₂ O	76 µL
Plasmid preparation or PCR product	10 µL
Suitable restriction enzyme buffer (10x)	10 µL
Restriction enzyme 1	2 µL
Restriction enzyme 2	2 µL

The restriction mixtures of the plasmid and gene were purified using a PCR purification kit by the respective spin protocol and analysed by DNA gel electrophoresis.

Chapter 2 – Materials and Methods

If the restriction digests appeared to be successful, a ligation reaction was set-up consisting of:

H ₂ O	8.5 µL
Ligase buffer (10 x)	2 µL
ATP (50 mM)	0.5 µL
DNA T4 ligase	1 µL
Purified restricted plasmid	0.5 µL
Purified restricted insert/gene	0.4 µL

The reaction was incubated overnight at 6 °C and the ligation mixture used in a transformation reaction.

2.2.4 Site directed mutagenesis (SDM) - adapted from the Stratagene protocol

Primers were designed that were homologous with the region of the gene containing the area to be mutated other than for one or two bases which would alter the amino acid to be an alanine residue.

The primers were used in a normal PCR reaction as previously described with thermal cycler conditions of:

Temperature	Time (min)
95 °C	1
95 °C	0.5
55 °C	0.5
68 °C	1 per 1 kb DNA
Repeated 29 times	
68 °C	10
4 °C	Hold

To remove non-mutated original template DNA the PCR reaction was incubated with 1 µL of DpnI for 1 h, the DpnI enzyme digesting the methylated DNA template. After DpnI

digestion 6 μL of the reaction mixture was used in a chemical transformation reaction into a heterologous host as described.

2.2.5 Preparation of competent cells

2.2.5.1 Preparation of chemically competent cells

LB agar plates were inoculated with glycerol stocks of the desired cell and incubated overnight at 37 °C. Single colonies were used to inoculate 10 mL LB broth and incubated overnight at 37 °C, 200 rpm. The overnight cultures were 50 fold diluted into 10 mL LB broth and grown for 2-3 h at 37 °C, 200 rpm until OD_{600} reached approximately 0.4. The cultures were chilled on ice for 10 min and centrifuged at 4000 rpm using a bench top centrifuge (Heraeus rotor 75006445) for 10 min at 4 °C. The supernatant was removed and the pellet resuspended in 10 mL of KMES buffer and incubated on ice for 1 h. The mixture was centrifuged at 4000 rpm (Heraeus rotor 75006445) for 10 min and resuspended in 1 mL KMES buffer 1 containing 10 % (v/v) glycerol. The cells were then aliquoted (in 100 μL volumes) into sterile 1.5 mL tubes and stored at – 80 °C until needed.

2.2.5.2 Transformation of chemically competent cells

Approximately 2 - 6 μL of a desired plasmid construct was added to 100 μL of *E. coli* competent cells and incubated on ice for 30 min. The mixture was then heat shocked at 42 °C for 30 s and returned to ice for approximately 2 min. 250 μL of room temperature LB was added to the mixture and the mixture incubated at 37 °C, 200 rpm for 1 h. Then 250 μL and 50 μL (5 x dilution) of the mixture was used to inoculate LB agar plates containing an antibiotic for selection and incubated overnight at 37 °C. Colonies from the incubated plates were selected and grown overnight in 10 mL LB broth containing the desired antibiotic selection. The culture was then used for plasmid purification and analysed by agarose gel electrophoresis.

2.2.5.3 Preparation of electrocompetent cells

A LB agar plate was inoculated with glycerol stocks of the desired cell type and incubated overnight at 37 °C. Single colonies were used to inoculate 10 mL LB broth and incubated overnight at 37 °C, 200 rpm. The overnight cultures were 50 fold diluted into 10 mL LB broth and grown for 2-3 h at 37 °C, 200 rpm until OD_{600} reached approximately 0.4. The

Chapter 2 – Materials and Methods

cultures were chilled on ice for 10 min and pelleted at 4000 rpm using a bench top centrifuge (Heraeus rotor 75006445) at 4 °C. The supernatant was removed and the pellet resuspended in 10 mL 10 % (v/v) glycerol and centrifuged at 4000 rpm for a further 10 min. The supernatant was removed and the pellet again resuspended in 10 mL 10 % (v/v) glycerol. The suspension was pelleted a final time, the pellet being resuspended in 100 µL of 10 % (v/v) glycerol and immediately being used in an electrotransformation reaction.

2.2.5.4 Transformation of electrocompetent cells

Approximately 5 µL of the desired plasmid/cosmid construct was added to 50 µL of *E. coli* electrocompetent cells. The cells were briefly chilled on ice and then electroporated under the condition; 200 Ω, 25 µF and 2.5 kV. The mixtures were returned to ice and 250 µL of ice cold LB medium added to the mixture and the mixture incubated at 37 °C, 200 rpm for 1 h. The entire mixture was used to inoculate 20 mL LB agar with correct antibiotic selection and incubated overnight at either 30 or 37 °C. Colonies from the incubated plates were selected and grown overnight in 10 mL LB broth containing the desired antibiotic selection and the plasmids purified by alkaline lysis or Qiagen plasmid purification and analysed by agarose gel electrophoresis.

2.2.6 DNA purification and analysis

2.2.6.1 Alkaline lysis – adapted from Sambrook et al. (2001)

An overnight culture (10 mL LB broth kan) was microcentrifuged for 1 min at 13000 rpm (rotor PP 1/97 #3324) and the supernatant removed. The pellet was re-suspended in 100 µL of ice cold alkaline lysis buffer I by vortexing. 200 µL of alkaline lysis buffer II was added to the mixture, mixed by inversion and incubated on ice for 5 min. 150 µL of alkaline lysis buffer III was added to the mixture, mixed by inversion and stored on ice for a further 5 min. The mixture was centrifuged at 13000 rpm (rotor PP 1/97 #3324) for 5 min. The supernatant was collected and transferred to a sterile 1.5 mL tube, treated with 2 vol of ethanol, vortexed and left to stand for 2 min. The mixture was centrifuged for a further 30 min at 13000 rpm (rotor PP 1/97 #3324) and the ethanol removed by aspiration. The pellet was washed with 1 mL 70 % ethanol and centrifuged for a final 5 min at 13000 rpm. After centrifugation the supernatant was removed and residual ethanol removed by room temperature evaporation. The pellet was re-suspended in 50 µL of milli-Q water.

2.2.6.2 DNA purification (Qiagen protocol)

Five volumes of Qiagen buffer PB was added to the restriction mixture and placed into a QIAquick spin column which was inserted into a 2 mL collection tube. The mixture was centrifuged at 13000 rpm (rotor PP 1/97 #3324) for 1 min, the flow through removed and 750 µL of Qiagen buffer PE added to the spin column. The column and collection tube were centrifuged at 13000 rpm (rotor PP 1/97 #3324) for 1 min, the flow through discarded and the column centrifuged for a second time at 13000 rpm for 1 min. The remaining ethanol was evaporated from the column by allowing the column to stand at room temperature for 5 min. 40 µL of EB buffer was added to the spin column and the column placed in a sterilised 1.5 mL eppendorf and centrifuged for 1 min at 13000 rpm (rotor PP 1/97 #3324) and the eluant collected.

2.2.6.3 Gel filtration (Qiagen protocol)

Gel purification is used to purify linear DNA. The mixture of DNA was loaded on a freshly prepared 1 % agarose gel, run at 120 V and the band representing the desired DNA excised using a scalpel. The gel fragment was weighed in a sterile 1.5 mL tube and 100 µL of Qiagen buffer QG added for every 100 mg of gel. The mixture was incubated at 50 °C for 10 min to allow the gel fragment to dissolve. One gel vol of isopropanol was added, the sample inserted into a spin column and the spin column placed into a collection tube and centrifuged for 1 min at 13000 rpm (rotor PP 1/97 #3324). The flow through was removed and an additional 500 µL of Qiagen buffer QG added to the spin column. The column was centrifuged for a further 1 min at 13000 rpm (rotor PP 1/97 #3324). 750 µL of Qiagen buffer PE was added to the spin column and the column centrifuged at 13000 rpm (rotor PP 1/97 #3324) for 1 min, the through flow discarded and the column centrifuged again at 13000 rpm (rotor PP 1/97 #3324) for 1 min. The spin column was dried for 5 min at room temperature to remove residual ethanol and placed in a sterile 1.5 mL tube. 40 µL EB buffer was added to the column and the column centrifuged for a final 1 min at 13000 rpm (rotor PP 1/97 #3324) and the eluant collected.

2.2.6.4 Restriction digest for plasmid analysis

A restriction reaction mixture was prepared:

H ₂ O	16 μ L
Plasmid with desired gene	1 μ L
Suitable restriction enzyme buffer (10x)	2 μ L
Restriction enzyme 1	0.5 μ L
Restriction enzyme 2	0.5 μ L

The mixture was incubated at 37 °C for ≥ 1 h and analysed using DNA agarose gel (1 %) electrophoresis.

2.2.6.5 Agarose gel electrophoresis

The agarose gel electrophoresis apparatus was assembled as per manufacturer's instructions and a 0.8 % (general analysis) or 1 % (gel extraction) agarose gel prepared with 0.8 g and 1 g agarose, respectively, 2.5 μ L ethidium bromide and 100 mL 1 x TAE buffer. Gels were loaded with 6 μ L 1 kb DNA ladder and 5-20 μ L of sample containing 6 x DNA loading dye. Samples were run at 120 V in 1 x TAE buffer for 20 – 40 min and visualized using a transilluminator.

2.2.6.6 DNA sequencing

Sequencing of plasmid constructs was undertaken either by the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge or by The Genome Analysis Centre (TGAC), Norwich Research Park, Norfolk. 100 ng μ L⁻¹ of plasmid construct was submitted for sequencing to the Cambridge facility. When using TGAC, the sequencing reaction was carried out in the laboratory using the Big Dye method.

Chapter 2 – Materials and Methods

The 10 µL Big Dye 2.0 sequencing reaction mixture consisted of:

H ₂ O	1 µL
BigDye sequencing buffer	1 µL
BigDye v.1	2 µL
Plasmid-insert template	5.5 µL
T7 Forward Primer (20 mM)	0.5 µL

Using thermal cycling conditions:

Temperature	Time (min)
95 °C	1
95 °C	0.5
50 °C	0.5
60 °C	4
Repeated 29 times	
72 °C	10
4 °C	Hold

The sequencing results were visualised using the SEQUENCE SCANNER software (Life Technologies).

2.2.7 Bioassay of pacidamycin activity

A solid medium bioassay approach was taken. The indicator organism was *E. coli imp⁻* and *E. coli* DH5α used as the control strain. The *E. coli* strains were grown overnight at 37 °C at 200 rpm. The cultures were diluted 100 fold into 10 mL broth and grown for approximately 4 h. They were 10 fold diluted into DNA agar (0.5 % agar) and 10 mL used to overlay DNA agar plates. Bioassay discs were placed onto the solid media and 10 µL of pacidamycin extract placed onto each disc along with a control of 50:50 H₂O:CH₃OH. The bioassays were incubated overnight at 37 °C and the ring of inhibition around the discs analysed.

2.3 Protein methodologies

2.3.1 Cell lysis

2.3.1.1 Cell lysis using cell lysis buffer (for SDS-PAGE analysis)

The pelleted cell culture was resuspended in 500 μL cell lysis buffer containing lysozyme (1 mg mL^{-1}), agitated until viscous and stored at $4 \text{ }^\circ\text{C}$ for 10 min followed by room temperature for 30 min. The viscous mixture was passed through a G23 needle (Terumo) until it became less viscous and centrifuged at 13000 rpm (rotor PP 1/97 #3324) for 15 min at $4 \text{ }^\circ\text{C}$. The supernatant (soluble protein) was separated from the pellet (insoluble protein) by centrifugation ready for SDS-PAGE analysis.

2.3.1.2 Cell lysis by sonication

The induced cell culture was centrifuged at 4000 rpm (Heraeus rotor 75006445) for 10 min and the pellet resuspended in 0.05 vol (of original culture vol) FPLC loading buffer A. An EDTA-free protease inhibitor tablet was added to the resuspension and the suspension was lysed by sonication (40 % intensity, 1 s on, 3 s off, 8 min). The pellet and supernatant were separated by centrifugation for 30 min at 16 000 rpm using a SS-34 rotor at $4 \text{ }^\circ\text{C}$.

2.3.1.3 Cell lysis by cell disruption

The induced cell culture was centrifuged at 6000 rpm (Rotor SLC 4X1000y) for 10 min and the pellet resuspended in 0.05 vol FPLC buffer A. An EDTA free protease tablet was added to the resuspension and the suspension was loaded into the cell disruptor and the cells lysed at a pressure of 25 Kpsi (*E. coli*) or 30 Kpsi twice (*S. coeruleorubidus*). The lysed suspension was returned to ice and centrifuged at 18000 rpm (rotor SS-34) at 4°C . The supernatant was then removed and used for protein purification.

2.3.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

For 'home-made' SDS-PAGE gels, the SDS-PAGE apparatus was prepared as per the manufacturer's instructions and a 12 % acrylamide running gel prepared by mixing 4 mL of SDS-PAGE solution A with 2.5 mL SDS-PAGE solution B, 3.5 mL of water, 50 μL of ammonium persulfate (APS) and 10 μL of tetramethylethylenediamine (TEMED). The gel was loaded into the apparatus and allowed to set. Once set, the iso-butanol layer was removed and a 5 % acrylamide loading gel prepared with 660 μL of solution A, 1 mL

Chapter 2 – Materials and Methods

solution C, 2.7 mL water, 50 μ L APS and 5 μ L TEMED. The loading gel was loaded on top of the running gel and combs inserted to form wells for loading the samples. Once set, the prepared protein samples were boiled for 10 min at >90 °C with 2 x SDS loading dye. The samples (10 μ L) and a marker (6 μ L) were loaded into the wells and SDS-PAGE running buffer was added to the running apparatus. The samples were run at 50 V – 200 V for approximately 1.5 h. Precast gels were loaded into a suitable running tank, loaded as per 'home-made' gels and run at 50 V – 200 V for \sim 1 h. The gels were stained for 2 h using Instant Blue.

2.3.3 Protein production trials

Colonies from overnight LB agar plates were selected and grown overnight in 10 mL LB broth containing antibiotic selection at 37 °C, 200 rpm. Plasmids were purified from the cultures by alkaline lysis and analysed on a DNA agarose gel. If the desired plasmid construct was present, the colony was reselected and grown overnight in LB broth as a starter culture. Overnight starter cultures were 1000 fold diluted into fresh LB broth and incubated for 4-5 h until OD_{600} was approximately 0.6-0.8. A 1 mL pre-induced sample was taken and the culture induced using 0.1 mM IPTG (final concentration) and incubated at either 37 °C for 4 h or 16 °C for 16 h. After the incubation time the culture was pelleted by centrifugation for 10 min at 13000 rpm (rotor PP 1/97 #3324). The pellet was lysed using cell lysis buffer and the supernatant obtained. The pre-induced, induced and supernatant samples were analysed by SDS-PAGE to determine the successfulness of the trial.

2.3.4 Protein production

Protein production was carried out differently for each protein produced dependent on the outcome of the initial protein production trials. The initial protein production trials allowed for the analysis of the best conditions to produce the maximum amount of viable, soluble protein. These conditions were then repeated on a larger scale to produce enough protein for subsequent studies.

2.3.5 Protein purification by affinity and size exclusion chromatography

A protein that has been produced to contain a hexa-histidine tag can be purified using metal chelate affinity chromatography. When a column containing a chromatographic matrix has been charged with a suitable metal, such as Ni^{2+} , the hexa-histidine tag and

other histidines found on the surface of the protein structure are able to bind. Proteins that do not contain the hexa-histidine tag will not bind to the matrix or will bind to a lesser affinity and be washed away from the matrix during a wash step in the purification process. As the protein of interest has reversibly bound to the matrix, a competitor (such as imidazole) can be used to elute the purified protein of interest away from the column for collection. Size exclusion chromatography allows for the separation of proteins and protein species (oligomeric state) on the basis of their size (and less importantly shape). Size exclusion chromatography can be used to estimate the oligomeric state of the protein when the column has been calibrated with molecular weight standards.

2.3.5.1 Protein purification using the ÄKTA express FPLC

The supernatant from cell lysis was loaded onto the ÄKTA Express automated FPLC system which contained a 5 mL His-Trap column (GE Healthcare) loaded with nickel and a Superdex 75 26/60 or Superdex 200 26/60 (GE Healthcare) gel filtration column. Buffers used were a loading/binding/wash buffer (buffer A), a nickel column elution buffer (buffer B), which was used to elute the protein of interest from the nickel column in a step wise fashion and a gel filtration buffer (GF buffer). Fractions of protein were collected and run on an SDS-PAGE gel to determine the purity of the sample. Fractions of high purity were pooled and concentrated. The concentrated samples were stored in gel filtration buffer for protein crystallography trials or gel filtration buffer with 10 % (v/v) glycerol for biochemical use. All samples were stored at -80 °C.

2.3.5.2 Protein purification using the ÄKTA FPLC

The supernatant from cell lysis was loaded onto a 1 mL His-Trap column (GE Healthcare) that had been charged with NiCl₂ prior to use. Buffer A was used to load the crude protein samples onto the affinity column and wash the column prior to the elution step. Any protein retained by the column was eluted using buffer B, the elution of the protein from the affinity column being caused by the increase in imidazole present due to the increase in volume of buffer B used. The elutant was collected in 1 mL fractions and the absorbance at 280 nm was recorded. Fractions containing protein were pooled and analysed by SDS-PAGE analysis. Purified protein was concentrated after purification and stored in gel filtration buffer for protein crystallography trials or gel filtration buffer with 10 % (v/v) glycerol for biochemical use.

2.3.6 Bradford assay

The Bradford assay (Bradford, 1986) was performed to determine the protein concentration of a given sample. A series of protein concentration standard solutions (using BSA) were prepared in a 0.15 M NaCl solution. A 2 mL cuvette was filled with 1 μ L of the standard solution, 999 μ L of water and 1 mL of Bradford reagent and the mixture inverted and incubated for 5 min. After 5 min the absorbance of the solution was measured at 595 nm, using a cuvette containing 1 mL water and 1 mL Bradford reagent as a blank control. This was repeated for each of the series of protein standards to produce a standard curve measuring absorbance against protein concentration. This standard curve was then used to determine the protein concentration of an unknown protein sample using the same preparation method as described.

2.3.7 Dynamic light scattering (DLS) analysis

The protein sample in solution was analysed by dynamic light scattering to deduce if any protein aggregation had occurred which could become problematic for crystallisation and/or enzymology studies. 25 μ L of the purified protein solution was passed through a centrifugal 0.1 μ m filter (Millipore) to remove any particulate material and 12 μ L of the filtered solution was added to the microsampling cell of a DLS. A total of twenty scattering measurements were taken at 20 °C and the results analysed using DYNAMICS V6 software computer package (Wyatt Technology).

2.3.8 Dialysis

Dialysis was used to change the constituents and pH of the buffer of a protein sample. The protein solution was placed into 3.5 kDa cut-off dialysis tubing (Thermo Scientific). The tubing was sealed using clips and placed into a 2 L beaker containing a magnetic flea and approximately 1.5 L of desired new buffer. The setup was stirred at a low RPM overnight at 6 °C.

2.3.9 His-tag removal

On occasion it is necessary to remove the affinity tag (in this case a hexa-histidine) tag from the N- or C- termini of the protein of interest in order for the protein to successfully crystallise or for it to be active *in vitro*. In the case of this work the pET vector system was used for gene cloning. This vector system as well as containing the hexa-histidine tag also

contains a thrombin cleavage site to enable removal of the tag. Other vector systems use other proteases to cleave affinity tags such as 3C protease (pOPIN vector system) and Tobacco Etch Virus (TEV) protease.

2.3.9.1 Removal using thrombin

Purified protein was dialysed overnight into 2.5 mM CaCl₂, 150 mM NaCl, 20 mM Tris HCl, pH 8.4 at 4 °C. Approximately 5 units of thrombin was added to the protein sample for every mg of protein and incubated for 4 h at room temperature, mixing at 100 rpm. Following cleavage, the mixture was purified by affinity chromatography (removing the lysed hexa-histidine tag) and the flow through applied to a Superdex 75 26/60 filtration column to separate the thrombin from the cleaved protein. The column was washed through with gel filtration buffer and the eluted peaks collected separately allowing for the separation of the desired His-tag free protein and thrombin.

2.3.9.2 His-tag removal using DAPase (Qiagen)

Thrombin is often not successful in cleaving the histidine tags from proteins and can very often cause degradation of the recombinant protein. DAPase is an enzyme belonging to the TAGzyme series of enzymes from Qiagen. The TAGzyme system will efficiently remove N-terminal His-tags from purified proteins. As the TAGzyme has a C-terminal histidine tag, once the histidine tag has been removed from the desired protein, it can be purified again by passing it through a nickel column (all histidine tagged proteins and his tags will be retained by the column) leaving the flow through containing just the His-tag free protein of interest. The DAPase enzyme will remove dipeptides from the N-terminus of the desired protein until it reaches a termination sequence.

Chapter 2 – Materials and Methods

The termination sequences it recognizes are:

Xaa-Xaa...Xaa-Xaa	Lys-Xaa Xaa-Xaa
Xaa-Xaa...Xaa-Xaa	Arg-Xaa Xaa-Xaa
Xaa-Xaa...Xaa-Xaa	Xaa-Xaa Pro-Xaa
Xaa-Xaa...Xaa-Xaa	Xaa-Pro Xaa-Xaa
Xaa-Xaa...Xaa-Xaa	Gln-Xaa Xaa-Xaa

One or more of these recognition sites are generally found within the thrombin cleavage site and therefore DAPase is a perfect substitute for thrombin.

The purified protein of interest was dialysed into 20 mM MOPS pH 6.9, 100 mM NaCl at 4 °C overnight. 10 mg of the dialysed protein was then added to activated DAPase (60 µL DAPase, 1050 µL water, 120 µL cysteamine HCl and 540 µL DAPase buffer) and incubated at room temperature for 3 to 4 h. The mixture was then purified using a 1 mL His-Trap column. The column was equilibrated using the dialysis buffer and the protein solution loaded and washed with 8 vol dialysis buffer and the flow through was collected (protein without His tag). The His tagged protein and the DAPase enzyme were removed from the His-Trap column using 10 vol of 50 mM Tris HCl, 0.5 M NaCl, 0.5 M imidazole at pH 8.0.

2.3.10 *In vitro* protein activity methods

2.3.10.1 Determination of Pac17 activity (reported by Weiner et al., 2010)

The activity of Pac17 was assessed by either the formation of or disappearance of fumarate. A standard curve of 1 mM fumarate and 1 mM L-aspartate (in 50 mM Na₂PO₄ pH 8.5) between wavelengths 200 and 500 nm at 20 °C was produced using a UV/Vis spectrometer (Perkin Elmer). This was then used to determine suitable wavelengths to measure. A continuous assay was carried out using a variety of co-factors, protein concentrations, buffers and metals. The activity of Pac17 was either determined by the formation of fumarate (at 240 nm), where L-aspartate was the substrate used, or the depletion of fumarate (at 270 nm), where fumarate and either ammonia chloride or ammonia bicarbonate were used as the substrate.

2.3.10.2 Determination of Pac19 activity (reported by Lam et al., 2008, adapted from Tripier et al., 2003)

The activity of Pac19 was assessed using a stopped assay, where derivatisation of the diamino acid product, DABA, with phenylglyoxal produced a derivative with an absorbance at 345 nm. Using DAP as a supplement for DABA, a standard curve was produced using a UV/Vis spectrometer (Perkin Elmer), varying the concentration of DAP from 0.1 mM to 3.0 mM and derivatising it with 0.2 vol phenylglyoxal solution (250 mM phenylglyoxal, 1:1 acetic acid:H₂O pH 6.0). The assay was carried out in gel filtration buffer with a pH varying between 6.0 and 8.0. A variety of conditions were used to determine the activity of Pac19 and included changing the concentration of the protein used, the concentration of substrates, the addition of co-factor (PLP) and incubation time. Derivatisation of the product was carried out by incubating the assay solution with 0.2 vol phenylglyoxal solution and the absorbance measured at 345 nm.

2.3.11 Protein crystallographic methods

2.3.11.1 Protein crystallisation

A number of different approaches can be used to crystallise proteins, however, all methods work on the similar principle which is to bring the protein in solution to a state of supersaturation. At this state nucleation and crystal growth can occur. In this study, crystallisations were set up using the vapour diffusion method. This method consists of mixing the protein of interest (in solution) with a precipitant solution in the form of a drop. This drop is then equilibrated against the precipitant solution present in a reservoir, the concentration of the precipitant solution being greater in the reservoir than in the drop, therefore the concentration of the precipitant increasing in the drop over time. As the concentration of the precipitant solution in the drop increases, hopefully a point at which the protein will begin to nucleate will be achieved, this nucleation being the first step to the formation of a protein crystal. Protein nucleation is dependent on the concentration of both the protein and the precipitant.

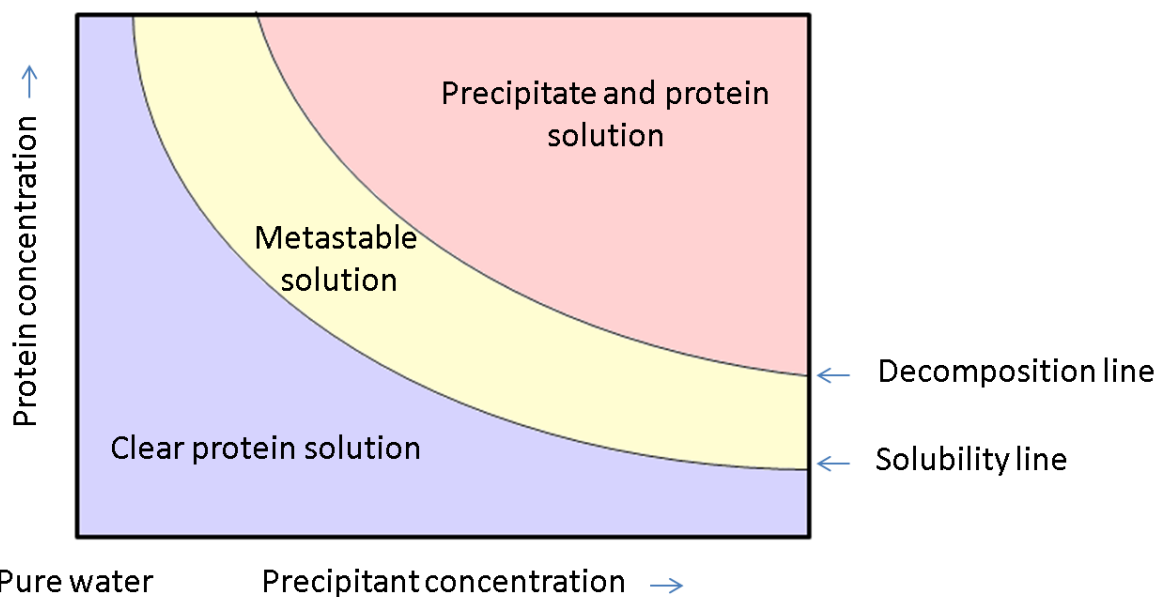


Figure 2.1 A basic solubility phase diagram providing a visual representation of the general observation that to produce protein crystals (metastable phase) a balance between the precipitant concentration and the protein concentration is necessary.

Figure 2.1 depicts a general solubility phase diagram. The objective of protein crystallisation is to produce a metastable solution of the protein and precipitant. If this can be achieved there is the possibility that the solution will separate into protein (in the form of crystals) and a saturated precipitant solution. If when crystal trials are set up, the protein or precipitant concentration are too high or too low, a two-phase region or single phase will be produced, respectively. A single phase will produce a clear protein solution, whereas a two-phase region will cause spontaneous decomposition (precipitation of the protein out of solution). Often in protein crystallisation trials, the protein and precipitant form a clear protein solution, but as diffusion takes place, both the protein and the precipitant become more concentrated in the solution and the solution moves closer towards a metastable phase.

Chapter 2 – Materials and Methods

Initial protein crystal screens were made using commercially available screens. The following screens and manufacturers were used:

JCSG – plus screen (Molecular Dimensions)

PACT premier screen (Molecular Dimensions)

Structure screen 1 (Molecular Dimensions)

Ammonium sulphate screen (Qiagen)

PEG suite screen (Qiagen)

Morpheus screen (Molecular Dimensions)

The screens were aliquoted into 96 well MRC plates (Molecular Dimensions) using the Freedom Evo liquid handling robot (Tecan), filling each reservoir with 50 μ L of the desired condition. The screen solutions and protein of interest were then mixed in sitting drops at a 1:1 ratio (total volume 0.6 μ l) using an OryxNano robot (Douglas Instruments Ltd). The screens were then sealed and stored at 20 °C.

Screens were checked regularly to observe any changes. Any conditions where crystals were observed were optimised using 24 well VDX optimisation plates (Molecular Dimensions) with a reservoir volume of 1 mL and a hanging drop volume of 2 – 3 μ L on a plastic cover slip which was used to seal the well. Optimisation of the conditions included changing the concentration of precipitate, the concentration of the protein, the ratio of protein to screen condition, changing the pH, co-crystallisation with other proteins or likely ligands and the addition of cryoprotectant.

2.3.11.2 Protein crystal cryo-protection and data collection

Before data collection, the crystal was protected against ice formation. Cryoprotectant can be added to the screen during optimisation or prior to crystal cooling by making up a buffer that contains the condition in which the protein crystallised plus 10 – 20 % (v/v) cryoprotectant (unless enough cryoprotectant is already present in the crystallisation condition). The most common cryoprotectants used are glycerol and ethylene glycol. If necessary, crystals were placed into the cryoprotectant for approximately 10 s and then mounted onto litholoops (Molecular Dimensions). Mounted crystals were either used in-house for initial analysis using a Rigaku RU-H3RHB rotating anode X-ray generator (50 kV

and 100 mA) fitted with osmic confocal optics and a copper target (Cu K_{α} ; $\lambda = 1.542 \text{ \AA}$) and Mar-345 image plate detector (X-ray Research) or plunged into liquid nitrogen and stored and taken to Diamond Light Source Synchrotron, Harwell, Oxfordshire.

2.3.11.3 Data collection

X-ray data were collected at the Diamond Light Source synchrotron, Oxfordshire either on the macromolecular crystallography beamlines (i02, i03, i04 and i04-1) or the microfocus beamline (i24). The diffracted X-rays were recorded on either a charge couple device (CCD) or a Pilatus detector at the beamline. Initially three test images were collected at $\phi = 0, 45$ and 90° to determine the quality of the diffraction and also to estimate the resolution. These images were indexed to deduce the crystal symmetry and processed using MOSFLM (Leslie and Powell, 2007) and EDNA (Diamond Light Source) to calculate the most suitable collection conditions (such as starting orientation and the number of degrees needed to be collected) to produce a full and highly resolved dataset. The recommendations from these analyses were then used to collect complete datasets at the synchrotron.

2.3.11.4 Data processing

Data collected either in-house or at Diamond Light Source was integrated using MOSFLM (Leslie and Powell, 2007), XIA2 or XDS (Kabsch, 2010) and scaled and merged using SCALA (Evans, 2006). Model building and refinement was carried out using the CCP4 suite (Winn et al., 2011). When a dataset was collected, a subset of data consisting of 5 % of the total number of reflections was set aside to be used for the free R factor (Kleywegt and Brünger, 1996) calculation during model building and refinement.

2.3.11.5 Molecular replacement

Molecular replacement is a crystallographic technique that can use information from similar protein structures to calculate estimates for the phase and is one of the methods used to overcome the phase problem. The method can often be used when a already solved molecule is in a different crystallographic form (isomorphous replacement) or when a known protein structure has a high sequence identity to the protein of interest. In isomorphous replacement, generally the difference in phases are so small that the phases from the search model can be used directly to compute the electron density (ρ) for each

point (x,y,z) in the unit cell from the native intensities of the new protein to produce an electron density map of the new protein query. If the search protein is isomorphous to the query protein, molecular replacement can be done via REFMAC5 (Murshudov et al., 1997), entering the merged data as the .mtz file and the .pdb file of the search protein, differences between the two structures can then be identified by the use of a visualising software such as COOT (Emsley and Cowtan, 2004). When the query and the search proteins are not isomorphous it is necessary to identify the position and orientation of the query model in the new unit cell by rotating and translating the model to deduce the phases that are most likely to be that of the query protein. Two programs in the CCP4 suite are able to do this; PHASER (McCoy, 2007) which calculates both the rotation and translation functions and gives you a result that has the highest likelihood of being correct and MOLREP (Vagin and Teplyakov, 1997) which can be commanded to calculate the most likely rotation and translation functions together automatically or calculate one and then the other. In this study both methods explained above were used for initial structure determination.

2.3.11.6 Phase improvement, model building and refinement

Model building and refinement can be done both automatically and manually using a series of programs from the CCP4 suite. The use of these tools allows for the improvement of the electron density map and therefore the protein model that was produced during molecular replacement or by other phase solving techniques. A number of automated model building programs can be used and include BUCANEER (Cowtan, 2006), which once given the primary sequence of the query protein can begin to use the electron density map to build more of the protein structure and ARPwARP (Perrakis et al., 1999) which can add water molecules into the model. By using a structure visualisation program such as COOT (Emsley and Cowtan, 2004), manual building/rebuilding of the protein structure can be done which allows the user to build the protein structure to complement the electron density map produced by phase solving and/or refinement. Model refinement can be carried out using PARROT which can modify electron density and REFMAC5 (Murshudov et al., 1997) which will automatically refine a structure using rigid body or restrained refinement depending on which parameters the program has been commanded to execute. Cycles of manual and automated model building and refinement of these alterations is carried out until the user is happy that no more improvement of the model can be done. The improvement of the model is traditionally monitored using the R_{work} and R_{free} values, however, in recent years it

has become common to also use the Mean $I/\sigma(I)$ and FOM as it is believed these give a more accurate representation of the accuracy of the built model, model improvement was also monitored by the 'fit' of the structure into the electron density and the stereochemical properties of the model.

2.3.11.7 Structural validation

Validation of the built structure is important to confirm that no bias has been built into the structure and that the resulting electron density maps from initial data collection have been interpreted accurately. A number of programs and software can be used for validation. The building software COOT (Emsley and Cowtan, 2004) has a number of validation tools which including production of Ramachandran plots (Ramachandran et al., 1963) (assessment of the torsion angles of the amino acid chains), assessment of the electron density for the built in waters and the position of amino acids in the protein in correlation to the electron density map. Further validation can be done externally of COOT via a number servers and programs. One online server commonly used is MOLPROBITY (<http://molprobity.biochem.duke.edu>) (Chen et al., 2010) which analyses the sterics and the geometry of the completed protein model by analysing Ramachandran outliers, energetic clashes in bonding, bond angle distortions and rotamer outliers. MOLPROBITY also supplies the user with a score indicating, based on the results, how accurate the structure would be in comparison to the structures currently deposited in the Protein Data Bank (PDB).

2.4 Streptomyces genetic methodologies

2.4.1 Culturing *Streptomyces*

2.4.2 Culturing *Streptomyces sp.* on solid medium

Two species of *Streptomyces* were used in this study. *Streptomyces coeruleorubidus* was grown on ISP₂ solid media and *Streptomyces lividans* was grown on SFM solid media. Cultures were produced by the inoculation of the agar plate with *Streptomyces* spores. After initial growth, a single colony from the plate was used for a subsequent inoculation of a fresh agar plate to ensure all colonies on a single plate had originated from the same colony. If possible a selectable marker was used to reduce the risk of contamination.

2.4.2.1 Culturing *S. coeruleorubidus* in liquid medium

Liquid ISP₂ (10 mL) in a spring baffled glass vial was inoculated with 50 µL of the desired spore stock and incubated at 30 °C, 250 rpm for 2 d. To produce the main *Streptomyces* culture, 500 µL of the starter culture was used to inoculate a further 10 mL ISP₂ in a spring baffled vial and the culture incubated for 5 – 6 d until a ring of spores was observed forming on the glass vial at the liquid-gas interface.

2.4.2.2 *S. coeruleorubidus* feeding experiments in liquid medium

From a 2 d 10 mL ISP₂ starter culture, 10 mL ISP₂ in a baffled vial was inoculated with 500 µL of the starter culture to produce the main culture and grown for 3 d at 30 °C, 250 rpm. After 3 d, the cultures were inoculated with the desired compound with a final concentration of 10 mM and incubated for a further 3 d at 30 °C, 250 rpm.

2.4.2.3 Culturing *S. lividans* in liquid medium

Liquid R5 medium (10 mL) in a spring baffled glass vial was inoculated with 50 µL of the desired spore stock and incubated at 30 °C, 250 rpm for 2 d. To produce the main *Streptomyces* culture, 500 µL of the starter culture was used to inoculate a further 10 mL R5 in a spring baffled vial and the culture incubated for 5 d.

2.4.3 Generation of genetic mutants by homologous recombination in *S. coeruleorubidus*

A method to determine the function of a gene is to disrupt/delete that gene within the genome of the native host. This can be done by the gene disruption method (REDIRECT™ Technology, outlined in Figure 2.2) developed at the JIC (Gust et al., 2003). The method relies on two naturally occurring processes; recombination and conjugation. By creating a PCR product that contains an antibiotic resistant gene sequence and flanking regions which complement the DNA sequence either side of the gene to be disrupted, recombination can be used to replace the gene of interest with that of the antibiotic gene cassette on a cosmid, the cassette allowing for a means of selection. The mutated cosmid can then be introduced into a non-methylating *E. coli* host containing a plasmid which allows for the formation of a pilus by *E. coli* and therefore conjugation of the cosmid from itself into another bacteria, this being either the gene of interest's natural host or a heterologous host.

Chapter 2 – Materials and Methods

Long primers (58 and 59 nucleotides (nt) in length) were designed to flank the replacement region (the gene of interest). The 39 nt sequence at the 5' end (forward primer) of the gene to be replaced and 39 nt of the complementary sequence of the 3' end (reverse primer) were taken. A designated 20 nt sequence of the interruption cassette (5' ATT CCG GGG ATC CGT CGA CC 3') was added to the 3' end of the forward primer sequence and a designated 19 nt sequence of the interruption cassette (5' TGT AGG CTG GAG CTG CTT C 3') added to the 3' end of the reverse primer sequence.

The interruption cassette containing the complementary sequence for recombination (obtained from the primer sequences) was prepared using a normal PCR reaction recipe with the disruption cassette template provided by Dr S. Grüşchow (UEA). The mixture was exposed to thermal cycler conditions:

Temperature	Time (min)
94 °C	2
94 °C	0.75
55 °C	0.75
72 °C	6
Repeated 29 times	
72 °C	5
4 °C	Hold

The PCR product was analysed by agarose gel electrophoresis to ensure correct amplification and the remainder of the product purified by Qiagen PCR purification.

E. coli BW25113/pIJ790 containing specific *S. coeruleorubidus* cosmids (2H-5 or 8G-5) were cultured overnight at 30 °C in 10 mL LB broth containing chl and 100 µL of the culture used to inoculate 10 mL LB containing carb, kan, chl and 100 µL 1 M L-arabinose and incubated at 30 °C for 3 – 4 h until OD₆₀₀ was approximately 0.4. The L-arabinose induces the promoter, allowing for the λ red genes responsible for recombination to be expressed. The culture was pelleted by centrifugation at 4000 rpm (Heraeus rotor 75006445), 4 °C and the pellet resuspended in 10 mL 10 % (v/v) glycerol, centrifuged for a further 10 min at 4000

Chapter 2 – Materials and Methods

rpm (Heraeus rotor 75006445), 4 °C, the glycerol removed and the pellet resuspended in 5 mL 10 % (v/v) glycerol, which was again centrifuged at 4000 rpm (Heraeus rotor 75006445) for 10 min at 4 °C. The glycerol was removed and the pellet re-suspended in 100 µL 10 % (v/v) glycerol. 50 µL of the resuspended mixture was inoculated with approximately 100 ng of the amplified disruption cassette with primer ends and transformed by electroporation. After electroporation, 250 µL of liquid LB was added to the transformation mixture and incubated at 37 °C, 200 rpm for 1 h and the entire culture used to inoculate 20 mL LB agar containing kan and apra. Gene disruption was confirmed by inoculating 5 mL LB broth containing kan and apra with colonies from the plates, incubating overnight at 37 °C, 200 rpm and analysing with alkaline lysis, using the purified plasmid for colony PCR analysis.

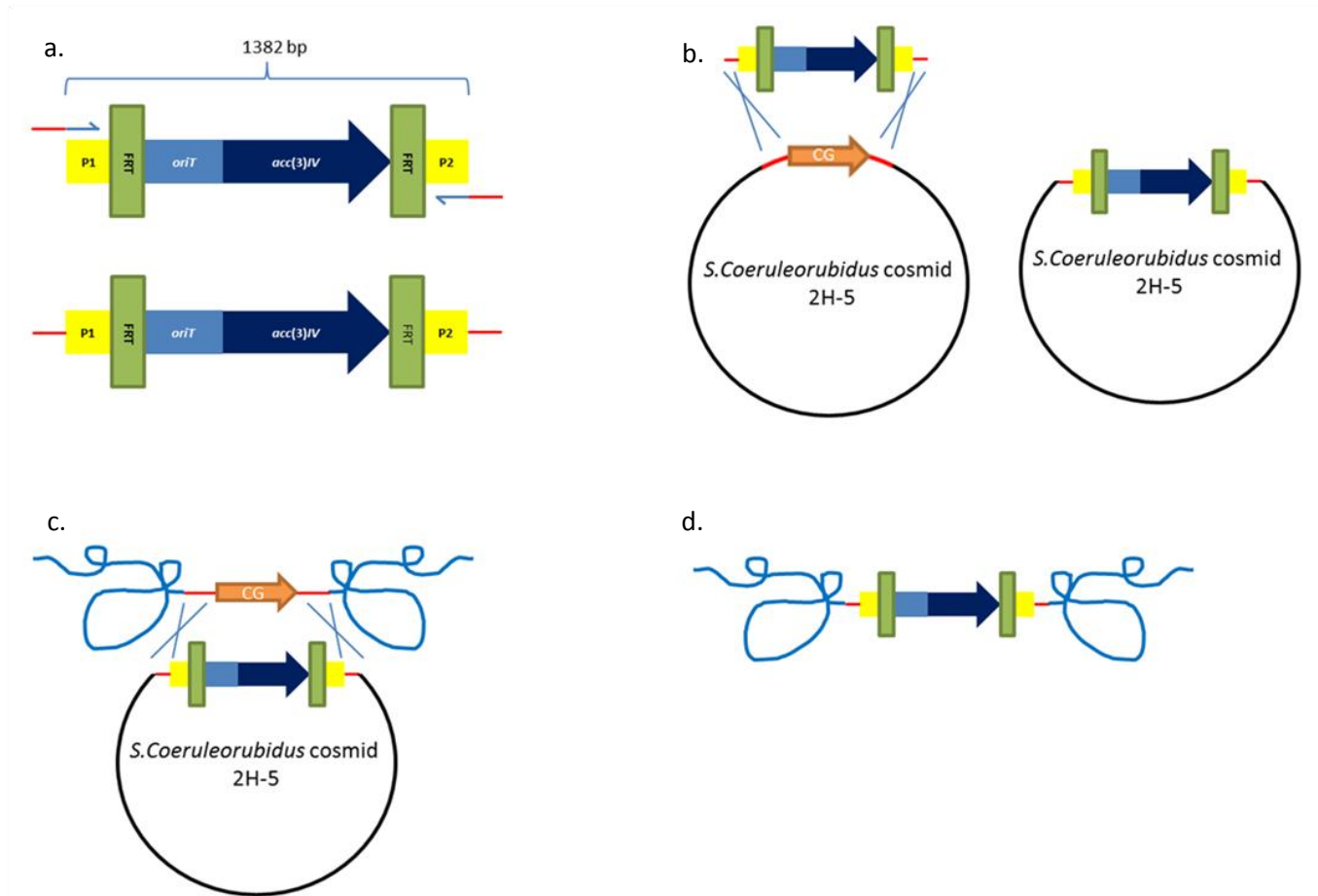


Figure 2.2 An illustration of the Re-direct protocol. The resistance cassette is amplified using specific primers (a) and by recombination replaces the gene of interest on a cosmid containing the pacidamycin gene cluster (b). The cosmid containing the resistance cassette in the place of the gene of interest is then conjugated in to the pacidamycin producer where another recombination event occurs (c) and the gene of interest in the native producers genome is replaced with the resistance cassette (d), eliminating the expression of the gene

2.4.4 Conjugation into *Streptomyces sp.*

This conjugation method was used for recombination into both *S. coeruleorubidus* and *S. lividans*.

On confirmation of successful interruption, 100 μL of electrocompetent *E. coli* ET12567/pUZ8002 was inoculated with 2 μL of plasmid preparation and incubated on ice for 30 min. The *E. coli* was transformed using electroporation. 250 μL of LB broth was added and the culture incubated at 37 $^{\circ}\text{C}$, 200 rpm for 1 h. The culture was used to inoculate 20 mL LB agar with apra and carb and incubated overnight at 37 $^{\circ}\text{C}$. 10 mL LB broth containing kan, apra and chl was inoculated with colonies from the overnight plates and incubated overnight at 37 $^{\circ}\text{C}$, 200 rpm. The overnight culture was 50 fold diluted into 10 mL LB broth with kan, apra and chl and incubated at 37 $^{\circ}\text{C}$ until OD_{600} was approximately 0.4. The culture was then pelleted by centrifugation at 4000 rpm (Heraeus rotor 75006445), 4 $^{\circ}\text{C}$ and the pellet washed twice with 10 mL Lennox broth, finally being resuspended in 1 mL Lennox broth. 200 μL of the re-suspended culture was used to inoculate 40 mL SFM agar with 10 mM MgCl_2 along with 200 μL of *Streptomyces sp.* culture (produced from 100 μL spore stock in 900 μL TSB, incubated for 5 h at 30 $^{\circ}\text{C}$, 250 rpm) and incubated at 30 $^{\circ}\text{C}$ overnight. Negative control plates containing 200 μL of *Streptomyces sp.* and positive control plates containing the *E. coli* cultures were also produced. After 20 h the plates were overlaid with 2 mL soft TSB agar with apra (500 $\mu\text{g mL}^{-1}$) and nal (500 $\mu\text{g mL}^{-1}$) and incubated at 30 $^{\circ}\text{C}$ for approximately 5 d. Once adequate growth of the *Streptomyces sp.* had appeared (and spores could be seen) colonies from the plates were used to inoculate agar plates of the correct growth medium (SFM for *S. lividans* and ISP_2 for *S. coeruleorubidus*) containing apra and nal and incubated at 30 $^{\circ}\text{C}$. Once colonies of sufficient size appeared, they were used to inoculate (in patches) new plates of growth medium containing either apra or kan. The replicates allowed for the observation of colonies in which recombination and double cross-over had occurred. Colonies that were resistant to apra and sensitive to kan indicated that kan resistance available on the recombinated cosmid backbone had been lost and therefore the gene of interest had been replaced with the apramycin resistance cassette.

2.4.5 Heterologous expression in *Streptomyces*

Sometimes it is necessary to study the function of a gene in a heterologous host. Examples of this is when a scar must be introduced in place of the gene of interest, meaning there would be no antibiotic selection if introduced into the natural producer, therefore an integrated backbone can be inserted in place of the original cosmid backbone so that an antibiotic selection is reintroduced. The PCR product that was produced during the original PCR step, as well as containing flanking complementary regions and the antibiotic resistance gene also contains flippase recognition target (FRT) regions that are recognised by a flippase that will remove anything in between the two FRT sites. The loss in antibiotic resistance can then be selected for. A second round of recombination is then necessary to replace the original backbone of the cosmid with that of an integrative cosmid backbone containing an antibiotic resistance gene. The cosmid, once conjugated into the heterologous host is then able to integrate into a phage integration site and become incorporated into the genome of the heterologous host.

2.4.5.1 FLP-mediated recombination

For heterologous expression studies, it was necessary to remove the apramycin resistance cassette and replace it with a FLP scar (81 bp). This is because, for heterologous expression the cosmid backbone would eventually need to be replaced to allow for integration of the cosmid into the Φ C31-*attB* site of the *Streptomyces* host, as the resistance gene on the backbone is the same as that of the cassette.

E. coli DH5 α /BT340 was grown in selective 10 mL L broth containing chl overnight at 30 °C , 200 rpm. A 1 in 100 dilution of the overnight culture was used to inoculate 10 mL selective L broth and grown at 30 °C for 4 h. Electrocompetent cells were produced and 100 ng of the cosmid was added to 50 μ L of the cells and electroporation transformation was carried out. A 500 μ L aliquot of the transformation mixture was plated onto DNA selective medium (chl and apra) and incubated for 2 d at 30 °C. Two colonies from each plate were selected and streaked onto L agar without any selection and incubated over night at 42 °C to promote flippase production and FLP-recombination. Single colonies from these plates (approximately 15) were streaked onto replica L agar plates containing either apra or kan and incubated overnight at 37 °C. The plates were compared to identify colonies that retained kan resistance but lost apra resistance which indicated that the FLP-recombination

event had been successful. Successful FLP-recombination was subsequently reconfirmed by PCR analysis.

2.4.5.2 Recombination of SspI fragment into cosmid backbone

E. coli DH5 α /pIJ10702 was incubated overnight at 30 °C, 200 rpm in 10 mL LB broth containing apra. The culture was pelleted by centrifugation at 4000 rpm (Heraeus rotor 75006445) for 10 min and the plasmid purified by Qiagen plasmid purification. The plasmid was digested using restriction enzyme SspI and a fragment of approximately 5247 bp purified by Qiagen gel extraction. Electrocompetent cells of *E. coli* BW25113/pIJ790 containing the 2H-5 cosmid with the FLP scar in the place of the gene of interest and expressing the λ red genes were produced and approximately 100 ng of the SspI fragment were mixed together and transformed by electroporation. Recombination of the new backbone was confirmed by the cosmid obtaining apra resistance.

2.4.5.3 Cosmid integration into a *Streptomyces* heterologous host

Integration of the cosmid containing both the FLP scar in place of the gene of interest and the integration backbone from pIJ10702 was performed using the same methodology as for the conjugation of a resistance cassette into a *Streptomyces sp.* The cosmid of interest was introduced into a non-methylating strain of *E. coli* (ET12567/pUZ8002) and conjugated with the *Streptomyces sp.* as previously described. Exconjugates were selected for by their resistance to apra and subsequently confirmed by colony PCR, checking for both the presence of the mutated gene and for the presence of the integration site by using primers Φ C31_F and Φ C31_R.

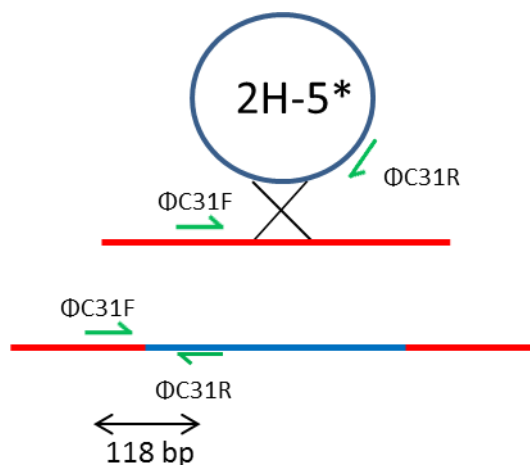


Figure 2.3 Illustration of the integration of the cosmid (2H-5*) into the genome of *S. lividans* TK24, and the position of the primers Φ C31F and Φ C31R which are used to confirm integration of the cosmid into the integration site.

2.4.6 Pacidamycin extraction and analysis

2.4.6.1 Pacidamycin extraction

After culturing the *Streptomyces sp.*, the mycelia and broth were separated by centrifugation at 3000 rpm (Heraeus rotor 75006445), 10 min, 4 °C. The medium was incubated with 50 % (v/v) XAD-16 overnight at 4 °C on a shaker. The resin was collected and washed with approximately 10 mL milliQ-H₂O and metabolites eluted from the resin using approximately 10 mL methanol. The solvent was removed *in vacuo* at 40 °C and the residue dissolved in 1 mL milliQ-H₂O and used for LC-MS analysis.

2.4.6.2 Detection of pacidamycin by LC-MS analysis

The extract components (5 μ L) were separated by reverse phase high pressure liquid chromatography with a C₁₈ X-Bridge column, flow rate 0.35 mL min⁻¹, 10-95 % LC-MS buffer B for 11 min and the masses of a number of the most commonly observed pacidamycins monitored (for pacidamycin in the natural producer) and the mass of pacidamycin D (m/z 712.0) monitored for all heterologous expression extracts.

2.4.6.3 Confirmation of pacidamycin production by LC-MS/MS analysis

On occasion, to ensure that what was observed during LC-MS analysis is in fact the expected product, LC-MS/MS analysis is necessary. MS₂ allows for the analysis of a fragmentation pattern from a parent ion, in this case the parent ion being that of a pacidamycin. A distinct fragmentation will occur to a molecule depending on its structure, therefore MS₂ analysis allows for a more confident assessment that the correct mass that was detected during LC-MS is in fact the molecule you were searching for. The extract components (10 µL) were separated using reverse phase high pressure liquid chromatography with a C₁₈ column, flow rate of 0.2 mL/min, 10-90 % LC-MS buffer B for 24 min and then 10 % LC-MS buffer B for a further 7 min. The parent ion of pacidamycin D (m/z 712) was monitored and MS₂ carried out on this ion to confirm the fragmentation pattern is the same as that expected for the fragmentation of pacidamycin D.

Cloning, Protein Production and Crystallisation

3.1 Introduction

As described in the aims of this thesis, the function of four hypothetical protein genes and five genes believed to be responsible for the biosynthesis of (2S,3S)-diaminobutyric acid present at the core of the pacidamycin structure are being investigated. This chapter discusses the initial bioinformatics analysis of pacidamycin gene cluster genes along with bioinformatics interrogation of each of the genes of interest. The over expression of these genes and production and purification of their respective proteins will also be discussed.

3.2 Results

3.2.1 Bioinformatics analysis of the proteins of interest

An initial bioinformatics analysis was conducted on each of the proteins of interest to deduce some parameters. The bioinformatics analysis of the Pac17 protein will be discussed in detail and a summary of the outcome of the analysis for all other query proteins will be presented in Table 3.1.

The *pac17* nucleotide sequence was translated into the amino acid sequence using the EXPASY TRANSLATE server (<http://web.expasy.org/translate/>). The amino acid sequence (sequence in Appendix 1) was used in a number of online servers such as NCBI BLAST (<http://blast.ncbi.nlm.nih.gov/>) (Altschul et al., 1990), PROTPARAM (<http://web.expasy.org/protparam/>) and XTALPRED (<http://ffas.burnham.org/XtalPred/cgi/xtal.pl>) (Slabinski et al., 2007a, Slabinski et al., 2007b). Pac17 was annotated by Rackham et al. (2010) as an argininosuccinate lyase like enzyme and was postulated to be involved in the biosynthesis of the core DABA residue found in pacidamycin. Analysis of the amino acid sequence by PROTPARAM estimated that the molecular weight of the 498 residue protein is 53564.6 Da and its isoelectric point (pI) is 5.51. Interrogation of the amino acid sequence by FOLDINDEX (<http://bip.weizmann.ac.il/fldbin/findex>) (Prilusky et al., 2005) suggested that the protein is predominately folded with a small region predicted to be unfolded near the centre of the protein (Figure 3.1)

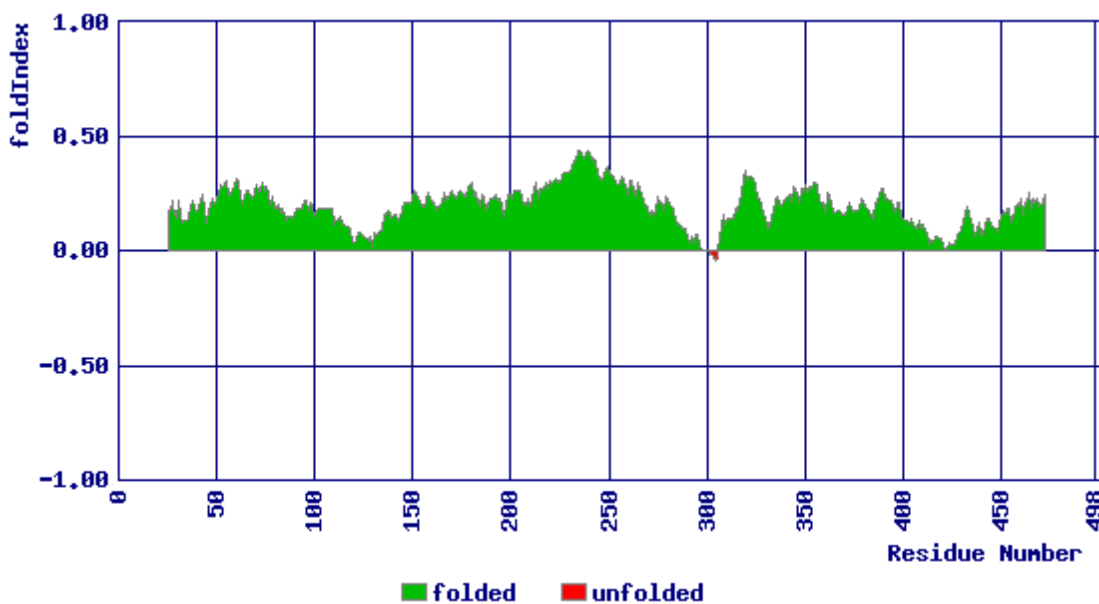


Figure 3.1 FOLDINDEX output when the Pac17 amino acid sequence was used as the query sequence. Pac17 is predicted to be predominately folded.

Further analysis of the Pac17 amino acid sequence was carried out using the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) (Jones, 1999). PSIPRED prediction of the secondary structure of Pac17 suggested that the protein would contain 18 α -helices and 1 β -sheet in each monomer, the β -sheet appearing near the centre of the protein at residues 267-268 (Figure 3.2).

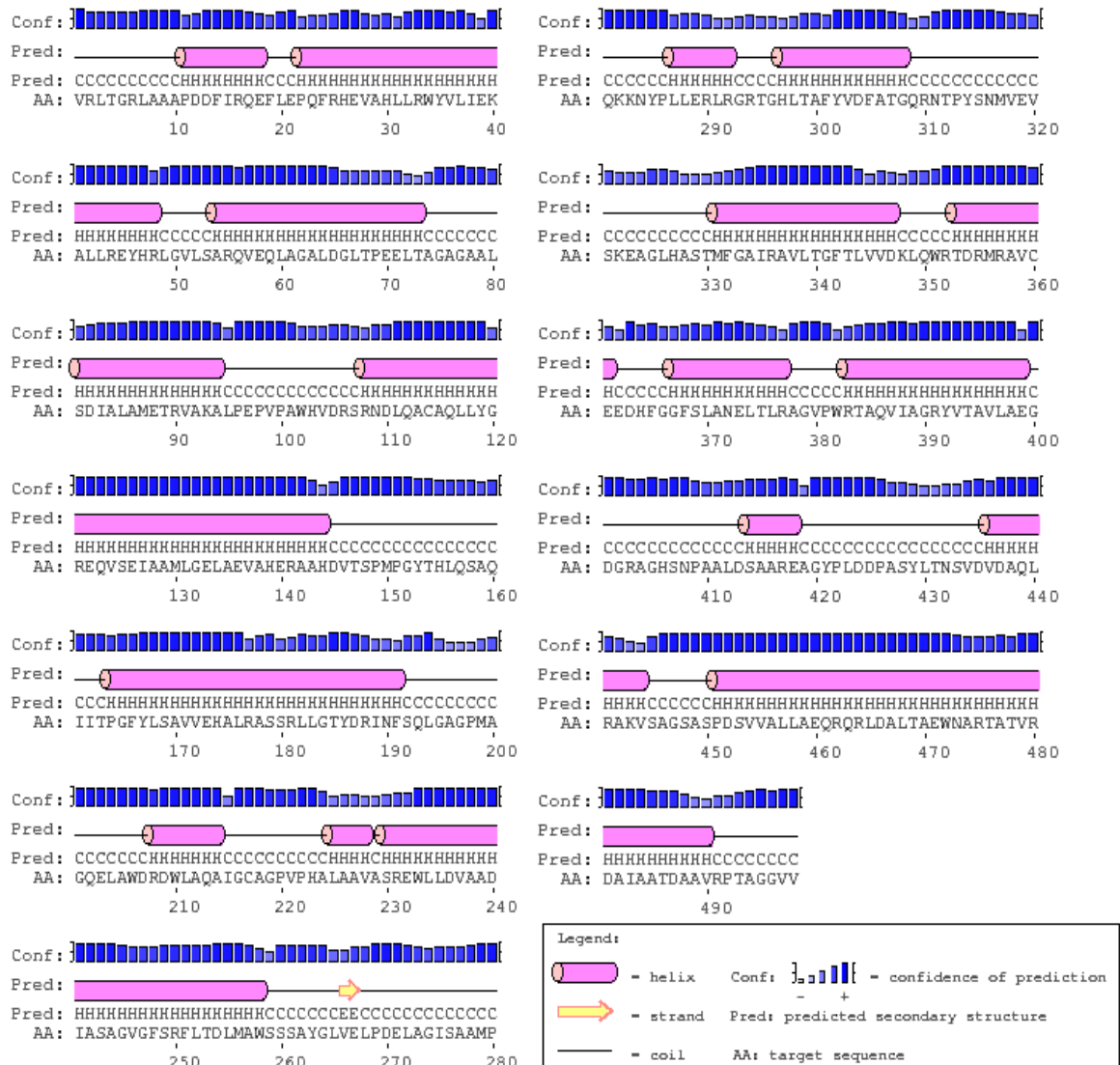

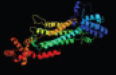

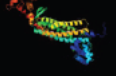

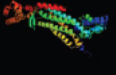



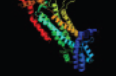




Figure 3.2 PSIPRED output of the Pac17 amino acid sequence. The protein is predicted to consist of 18 α-helices and 1 β-sheet near the centre of the structure.

To complement the secondary structure prediction of the protein, the Pac17 sequence was then analysed by the PHYRE2 server (<http://sbg.bio.ic.ac.uk/phyre2>) (Kelley and Sternberg, 2009). The PHYRE2 server predicts the 3D structure of the query protein using the information from previously solved protein structures that have homology to the query protein. The analysis identified that Pac17 is likely to be a member of the L-aspartase superfamily, which is consistent with the hypothesis for the protein’s function as an aspartase.

#	Template	Alignment Coverage	3D Model	Confidence	% I.d.	Template Information
1	d1lj7a_	 Alignment		100.0	27	Fold: L-aspartase-like Superfamily: L-aspartase-like Family: L-aspartase/fumarase
2	d1l0aa_	 Alignment		100.0	23	Fold: L-aspartase-like Superfamily: L-aspartase-like Family: L-aspartase/fumarase
3	d1k62a_	 Alignment		100.0	25	Fold: L-aspartase-like Superfamily: L-aspartase-like Family: L-aspartase/fumarase
4	d1ljva_	 Alignment		100.0	23	Fold: L-aspartase-like Superfamily: L-aspartase-like Family: L-aspartase/fumarase
5	d1hy0a_	 Alignment		100.0	23	Fold: L-aspartase-like Superfamily: L-aspartase-like Family: L-aspartase/fumarase
6	c2e9fC_	 Alignment		100.0	28	PDB header: lyase Chain: C; PDB Molecule: argini nosuccinate lyase; PDBTitle: crystal structure of t.th.hb8 argini nosuccinate lyase complexed with2 l-arginine

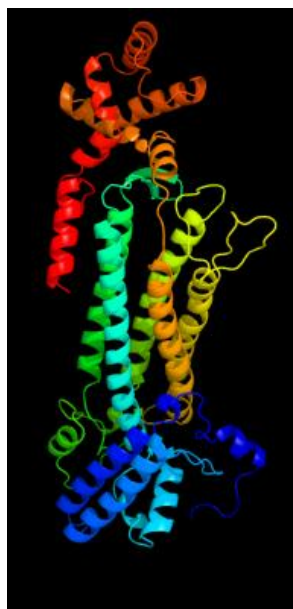


Figure 3.3 PHYRE2 output when the Pac17 sequence is used as the query sequence. The structures of the highest sequence homologs are first collected (top) and the structural information from these hits is then used to predict the 3D structure of the query sequence (Pac17), shown at the bottom.

Chapter 3 – Cloning, Protein Production and Crystallisation

A final bioinformatics analysis was carried out using XTALPRED (<http://ffas.burnham.org/XtalPred-cgi/xtal.pl>). XTALPRED analyses the probability that a protein would crystallise based on eight characteristics of the protein including the protein's length, isoelectric point and regions of predicted order.

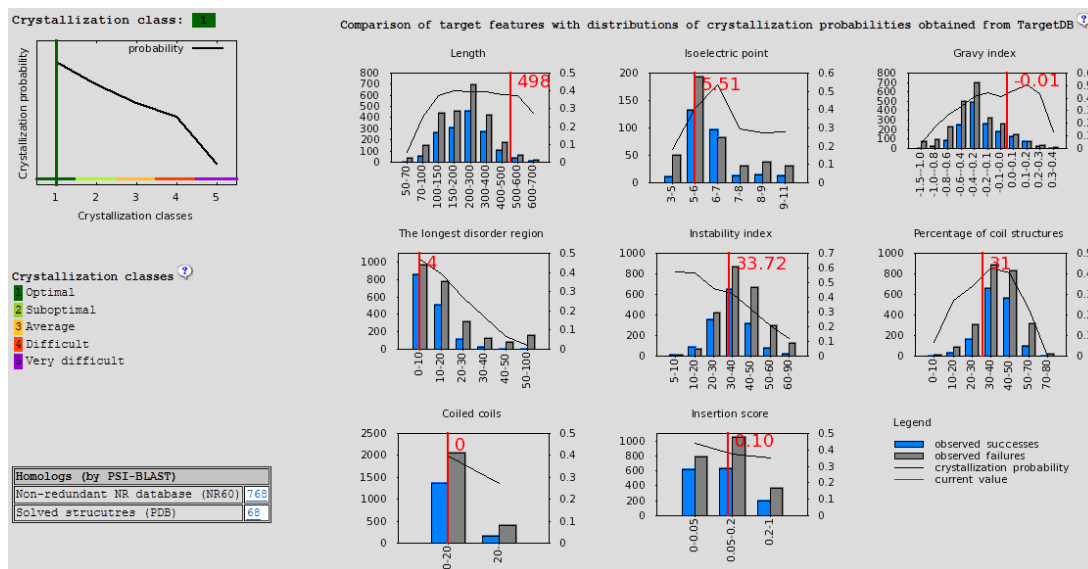


Figure 3.4 XTALPRED output for Pac17. From analysis of 8 characteristics of the protein, the server predicts that the protein is likely crystallise, suggesting that the chances of obtaining X-ray diffraction quality crystals are possible

Table 3.1 shows a summary of the bioinformatics analysis carried out on the other proteins of interest in this study; the four hypothetical proteins (Pac1, Pac2, Pac7 and Pac13) and the other proteins hypothesised to be involved in the biosynthesis of DABA (Pac18, Pac19, Pac20 and Pac22)

Table 3.1 Summary of the bioinformatics analysis carried out on the proteins investigated in this study

Protein	Uniprot Entry	Annotated by Rackham et al. (2010)	No of $\alpha\alpha$	Molecular weight (Da)	Calculated pI	BLAST 'hits' – accessed Jan 2010	Secondary structure prediction	PHYRE2 'hits' result	XTALPRED result
Pac1	E2EKN3	Hypothetical protein	258	28655.8	8.97	Putative uncharacterised protein	Folded. 11 β -sheets towards N-terminal and 5 α -helices at C-terminal	No significant hit	Difficult
Pac2	E2EKN4	Hypothetical protein	350	39445.1	5.23	Putative uncharacterised protein	Folded. 14 α -helices and 4 β -sheets	Acyl-CoA N-acyltransferase	Average
Pac7	E2EKN9	Hypothetical protein	244	27332.8	5.62	Putative uncharacterised protein	Folded. 8 α -helices and 10 β -sheets in centre of protein	Rmlc-like cupins	Suboptimal
Pac13	E2EKP5	Hypothetical protein	121	13815.3	5.27	Cupin-2 domain containing protein	Folded. 10 β -sheets and a C-terminal α -helix	No significant hit	Difficult
Pac17	E2EKP9	Argininosuccinate lyase	498	53564.6	5.51	Argininosuccinate lyase/aspartase/fumara se	Folded. 18 α -helices and 1 β -sheet in centre of protein	L-aspartase/fumarase family	Optimal
Pac18	E2EKQ0	Kinase	318	33635.2	8.42	Putative kinase/homoserine kinase	Folded. 13 β -sheets (mainly at termini) and 7 α -helices	Kinase	Suboptimal
Pac19	E2EKQ1	Synthase	755	79256.1	5.61	PLP-dependent synthase/lyase	Folded. 22 α -helices and 28 β -sheets	N-terminal synthase, C-terminal ligase	Difficult
Pac20	E2EKQ2	Threonine aldolase	350	36584.7	4.98	Threonine aldolase/allo-threonine aldolase	Folded. 16 α -helices and 10 β -sheets	PLP-dependent transferase	Suboptimal
Pac22	E2EKQ4	N-methyltransferase	250	27999.2	4.61	SAM-dependent methyltransferase	Folded. 9 α -helices and 6 β -sheets	S-adenosyl-L-methionine-dependent methyltransferases	Optimal

3.2.2 Pac17

3.2.2.1 Gene cloning of *pac17*

The *pac17* gene of *S. coeruleorubidus* was amplified by PCR using cosmid 2H-5, which contained the minimal pacidamycin gene cluster, with Pac17F primer containing a NdeI restriction site and Pac17R primer containing a BamHI site. The amplified DNA was NdeI/BamHI digested and ligated into the NdeI/BamHI digested expression vectors pET28a(+) and pET21a(+) to produce expression constructs pDT005 which encoded for the Pac17 protein (amino acid sequence Appendix 1) tethered to an N-terminus hexahistidine tag and thrombin cleavage site and pDT009 which encoded for the native Pac17 protein (pDT009 was used in expression studies of Pac18), respectively. The addition of this tag added an additional twenty amino acids onto the N-terminus of the protein with the sequence MSSHHHHHSSGLVPRGSH giving a total molecular weight of 55724.9 Da. The protein without the additional amino acid residues having a molecular weight of 53564.6 Da. The sequence of the constructs were confirmed by analysis by the DNA sequencing service at Cambridge University, UK. The pDT005 expression vector was introduced into *E. coli* BL21(DE3) cells by chemical transformation.

3.2.2.2 Expression studies of *pac17*

Protein production studies were carried out as described in chapter 2. The cells were lysed using the cell lysis buffer method and the supernatant analysed by SDS-PAGE to deduce the relative protein solubility and yield. The study suggested that the optimum expression conditions for the pDT005 construct was using AIM broth medium at 37 °C for 4 h followed by 16 °C for 16 h.

3.2.2.3 Large scale expression and purification of Pac17

An overnight starter culture (40 mL) was used to inoculate 4 L AIM broth and incubated using the same conditions as deduced from the *pac17* expression studies. After expression the cells were harvested and the pellet resuspended in FPLC loading buffer A and lysed by sonication. The supernatant was collected by centrifugation and loaded onto a ÄKTA express FPLC using a programme consisting of affinity chromatography purification and size exclusion chromatography, using a pre charged Ni²⁺ column and a Superdex 200 Hiload HP gel-filtration column which had been previously calibrated using molecular weight gel-filtration standards (as described in chapter 2), respectively. From gel filtration, the

Chapter 3 – Cloning, Protein Production and Crystallisation

molecular size of the homo-multimeric protein was estimated as approximately 150 kDa which would suggest the protein to be a trimer. The fractions containing the Pac17 protein which were confirmed by analysis using SDS-PAGE were pooled and concentrated to approximately 11 mg mL⁻¹ (determined by the Bradford assay). The purity of Pac17 was greater than 95% as determined by SDS-PAGE analysis (Figure 3.5). The protein yield was calculated to be approximately 24 mg L⁻¹ of culture. DLS analysis showed a monomodal distribution with 16 % polydispersity and an estimated molecular size of 108 kDa.

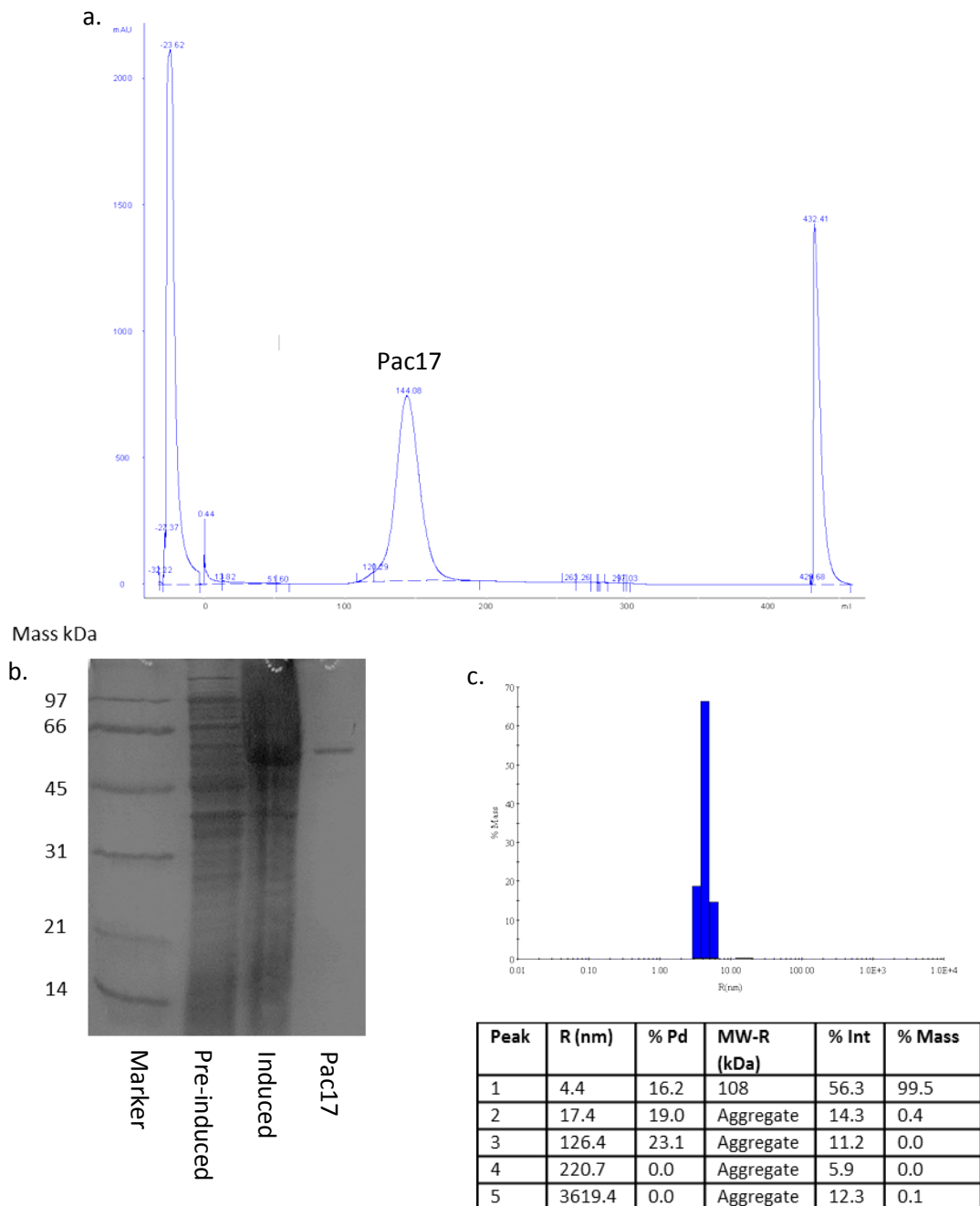


Figure 3.5 Showing (a) the gel filtration chromatogram from the purification of Pac17. Using a calibration curve for the gel filtration column it was estimated that the molecular weight of the protein was approximately 150 kDa (b) the SDS-PAGE result showing the soluble protein profile in the cell culture before and after induction and the purified Pac17 (size approx. 55.7 kDa) and (c) The DLS profile of the purified protein, which shows a predominant monomodal, monodispersed species (>99 %) within the solution with an estimated size of 108 kDa.

3.2.2.4 Crystallisation of Pac17

As the purity of the protein preparation was high (above 95 %), crystallisation trials were undertaken using the sitting drop vapour diffusion format in 96-well MRC plates, using five commercially available screens (as explained in chapter 2) at a temperature of 20 °C. Each drop consisted of 0.3 µL of well solution and 0.3 µL of protein solution and a well volume of 50 µL. The final protein concentration was approximately 6.5 mg mL⁻¹. Crystals grew within 24 h at 20 °C from a number of crystallisation conditions. The full crystallographic analysis of Pac17 is described in chapter 4.

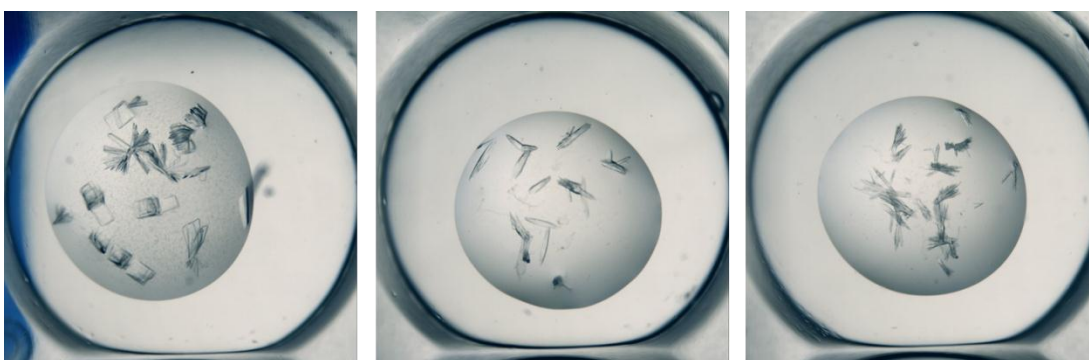


Figure 3.6 showing three examples of crystals observed from the initial crystallisation screens of Pac17. Optimisation of these conditions is discussed further in chapter 4.

3.2.3 Pac18

3.2.3.1 Gene cloning of *pac18*

The *pac18* gene of the pacidamycin producer was amplified by PCR using primers Pac18F and Pac18R which contained the restriction sites NdeI and BamHI, respectively, and cosmid 2H-5 as the template DNA. The amplified DNA was restricted with the aforementioned restriction enzymes and ligated into the NdeI/BamHI restricted plasmid pET28a(+) to produce plasmid construct pDT006 which encoded for the Pac18 protein (amino acid sequence Appendix 1) along with a twenty amino acid extension at its N-terminus of sequence MGSSHHHHHSSGLVPRGSH to allow for the expression of a hexa-histidine affinity tag for purification and a thrombin cleavage site. The addition of the affinity tag and cleavage site gave the expressed protein a total molecular weight of 35667.3 Da. The sequence of the construct was confirmed by analysis by the DNA sequencing service at Cambridge University, UK. The pDT006 expression vector was introduced into *E. coli* BL21(DE3) cells by chemical transformation.

3.2.3.2 Expression studies of *pac18*

Initial protein production studies were carried out as described in chapter 2. The cells were lysed using the cell lysis buffer method and the supernatant analysed by SDS-PAGE to deduce relative protein solubility and yield. From this original small scale production of Pac18, it was determined that Pac18 protein was being produced, however, none of this protein appeared to be soluble and therefore problematic for protein purification.

A new approach was taken to try and obtain soluble amounts of the protein. The plasmid construct pDT006 was introduced into a number of different *E. coli* expression strains (BL21(DE3) pLysS, Solu BL21, BL21 containing the GroESL plasmid and BL21 Arctic Express). Expression studies were carried out as explained previously, using both LB and AIM media in an attempt to produce soluble Pac18, however all of these conditions and *E. coli* strains failed to produce any soluble Pac18 protein.

Due to the difficulty of producing soluble Pac18, the more aggressive method of *in vitro* denaturing and refolding was attempted. As the protein was being produced by the expression host, the insoluble protein found in the pellet of the lysed cell could be denatured in an 8 M solution of urea (6 M guanidine-HCl can also be used). The denatured protein can then be purified using the same affinity chromatography method (however the buffers contain 8 M urea) and the urea can be removed from the solution either by step dialysis (gradually reducing the concentration of urea) or by rapidly reducing the urea concentration by injecting the protein solution into a large quantity of buffer that does not contain urea. Both of these methods were carried out, however neither produced soluble Pac18, with the protein precipitating out of the solution immediately with the 'quick urea change' method and at a concentration of about 2 M urea in the gradual dialysis method.

One final approach was attempted to try and produce soluble Pac18. It has been previously stated that the *pac17*, *pac18* and *pac19* genes are translationally coupled within the genome of *S. coeruleorubidus* which suggests that the proteins are likely to be produced in stoichiometric quantities. The fact that the genes are translationally coupled could also suggest that the resulting proteins produced from the expression of these genes may form a complex. This would be consistent with the hypothesis that they work together to produce DABA. As a result, it was thought that if *pac17* or *pac19* was co-expressed with *pac18*, it may allow for the two proteins to form a complex and stabilise the Pac18 protein

Chapter 3 – Cloning, Protein Production and Crystallisation

so that it will remain soluble. Both the *pac17* and *pac19* genes had been cloned into a pET21a(+) plasmid for native expression (i.e. without a tag), giving expression constructs pDT009 and pDT0010, respectively. Both of these plasmids could be selected for by their resistance to carb. Competent cells of *E. coli* BL21(DE3) containing pDT006 were prepared (as described in chapter 2) and pDT009 and pDT010 introduced into the cells by chemical transformation. Both plasmid constructs in the desired strains could be identified by using $50 \mu\text{g mL}^{-1}$ kan to select for the presence of pDT006 and $100 \mu\text{g mL}^{-1}$ carb to select for the presence of either pDT009 or pDT010. Expression trials were carried out in the same way as previously described and the presence of soluble Pac18 analysed. Unfortunately, when Pac18 is produced with either Pac17 or Pac19, no soluble Pac18 protein was detected.

A number of reasons could explain why the methods described have been unsuccessful in producing soluble Pac18. Firstly Pac18 is believed to be a kinase, a class of protein that is notoriously difficult to solubilise. Further explanations for the failure to produce soluble Pac18 could be that as three genes are believed to be translationally coupled, all three genes need to be produced together to obtain soluble Pac18, however as Pac17 and Pac19 can be produced on their own, with no issues of insolubility, this may not be the main cause. Another explanation could be that the heterologous host is lacking some factor that is essential for protein solubility.

Since the insolubility of Pac18 could be caused by the use of heterologous expression, a large scale culture (2 L) of *S. coeruleorubidus* was produced and the cells pelleted and lysed by cell disruption. The optimum conditions for lysis of *S. coeruleorubidus* cells had been deduced previously by conducting a series of cell disruption experiments and determining the concentration of soluble protein present by Bradford analysis. The supernatant produced from the lysis was then incubated with hexa-histidine tagged Pac17 and Pac19 in the hope that if an association occurred between Pac18 and one of these proteins, one would be able to purify Pac18 away from the supernatant and, possibly, with a co-factor that is essential for the solubility of Pac18. Unfortunately, this attempt was unsuccessful and so no further attempt was made to try and solubilise Pac18 for biochemical and biophysical analysis.

3.2.4 Pac19

3.2.4.1 Gene cloning of *pac19*

The *pac19* gene of *S. coeruleorubidus* was amplified by PCR using cosmid 2H-5 as the template DNA and primers Pac19F and Pac19R containing restriction sites NdeI and BamHI, respectively. The amplified DNA was restricted using the two specified restriction enzymes and ligated into a NdeI/BamHI restricted pET28a(+) and pET21a(+) plasmid to produce plasmid constructs pDT007 and pDT010, respectively. pDT007 encoded for the Pac19 protein (amino acid sequence Appendix 1) with a N-terminus hexa-histidine tag and a thrombin cleavage site adding an additional twenty amino acids onto the protein with sequence MGSSHHHHHSSGLVPRGSH, giving the protein a mass of 81288.3 Da. The pDT010 construct encoded for the native Pac19 protein with a mass of 79125.0 Da. The sequence of the constructs were confirmed by analysis by the DNA sequencing service at Cambridge University, UK. The pDT007 expression vector was introduced into *E. coli* BL21(DE3) cells by chemical transformation.

3.2.4.2 Expression studies of *pac19*

Protein production studies were carried out as previously described. The cells were lysed using the cell lysis buffer method and the supernatant analysed by SDS-PAGE to deduce relative protein solubility and yield. The study suggested that the optimum expression conditions for the pDT007 construct were using either LB or AIM broth medium at 37 °C for 4 h and 16 °C for 16 h, for ease, it was decided to grow the expression host (*E. coli* BL21(DE3)) in AIM medium.

3.2.4.3 Large scale expression of *pac19* and purification of Pac19

An overnight starter culture (10 mL) was used to inoculate 1 L AIM broth which was incubated using the same conditions as deduced from the *pac19* expression studies. After expression, the cells were harvested and the pellet resuspended in FPLC loading buffer A and lysed by sonication. The supernatant was collected and loaded onto a ÄKTA express FPLC using a programme consisting of an affinity chromatography purification step and a size exclusion chromatography step using a pre charged Ni²⁺ column and a Superdex 75 Hiload HP gel-filtration column which had been previously calibrated using molecular weight gel-filtration standards (as described in chapter 2). From the gel filtration, it was suggested that Pac19 exists in a monomeric state as it was estimated that the molecular

Chapter 3 – Cloning, Protein Production and Crystallisation

weight of protein being eluted from the column was 100 kDa. The fractions containing the Pac19 protein which were confirmed by analysis using SDS-PAGE were distinctly pale yellow in colour, which suggested that its predicted co-factor, PLP, was purified with the protein. The Pac19 fractions were pooled and concentrated to approximately 10 mg mL⁻¹ (determined by Bradford analysis). The purity of Pac19 was greater than 90% which was determined by SDS-PAGE analysis (Figure 3.7). DLS analysis suggested that 10 % of the sample was aggregated. Analysis using the Bradford assay determined that the protein yield was approximately 19 mg L⁻¹ of culture.

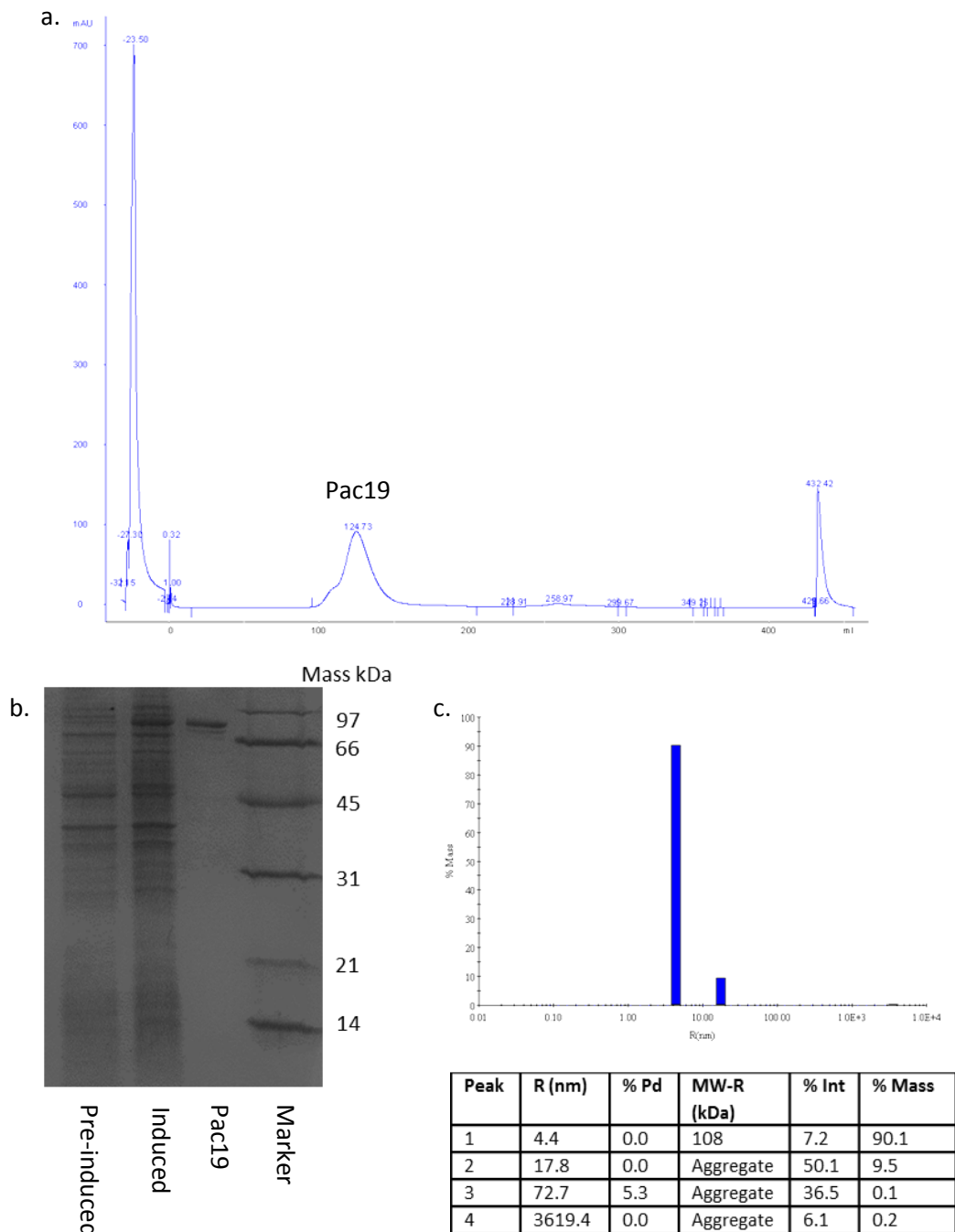


Figure 3.7 Showing (a) the gel filtration chromatogram from the purification of Pac19. Using the calibration curve for the gel filtration column it was estimated that the molecular weight of the protein was approximately 100 kDa (b) the SDS-PAGE result showing the soluble protein profile in the cell culture before and after induction and the purified Pac19 (size approx. 81.0 kDa) and (c) The DLS profile of the purified protein suggests approximately 10 % of the protein in the sample is aggregated as its hydrodynamic radius is more than 10.0 nm.

3.2.4.4 Crystallisation of Pac19

As the purity of the protein preparation was high (above 90 %), crystallisation trials were undertaken using the sitting drop vapour diffusion format in 96-well MRC plates using five commercially available screens (as explained in chapter 2) at a temperature of 20 °C. Each drop consisted of 0.3 µL of well solution and 0.3 µL of protein solution and a well volume of 50 µL. The final protein concentration was approximately 5 mg mL⁻¹. No crystals grew within six weeks. A number of the crystallisation conditions formed granular precipitate (example Figure 3.8) and so a number of other crystallisation trials were setup with conditions that included changing the precipitant concentration, increasing and decreasing the concentration of the Pac19 protein in the drop to 2.5 mg mL⁻¹ and 7 mg mL⁻¹, respectively, and also co-crystallising the protein with its postulated substrates; L-threonine and L-phospho-O-threonine. Unfortunately no protein crystals were observed in any of these conditions.

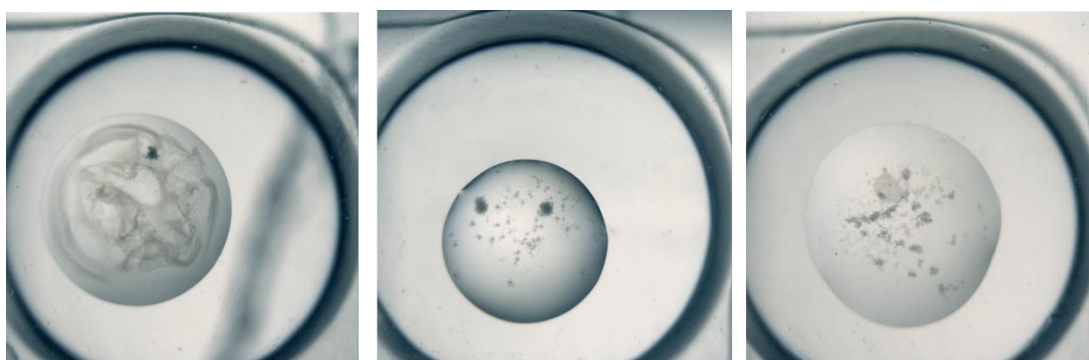


Figure 3.8 Examples of granular precipitation in the initial crystallisation screens of Pac19, all three conditions contain PEG as a precipitant and a pH of between 6.5 and 8.0

3.2.5 Pac20

3.2.5.1 Gene cloning of *pac20*

The *pac20* gene of *S. coeruleorubidus* was amplified by PCR using primers Pac20F and Pac20R which contain the recognition site for the NdeI and XhoI restriction enzymes, respectively. The amplified DNA was restricted using these restriction enzymes and ligated into a NdeI and BamHI restricted pET28a(+) plasmid to produce plasmid pDT008. pDT008 encoded for the Pac20 protein (amino acid sequence Appendix 1) with a N-terminal hexa-histidine tag and thrombin cleavage site adding an additional twenty one amino acids onto the protein with sequence MGSSHHHHHSSGLVPRGSH, giving the protein a mass of

38616.8 Da. The construct sequence was confirmed by analysis by the DNA sequencing service at Cambridge University, UK. The pDT008 expression vector was introduced into *E. coli* BL21(DE3) cells by chemical transformation.

3.2.5.2 Expression studies of *pac20*

Protein production studies were carried out as before. The cells were lysed using the cell lysis buffer method and the supernatant analysed by SDS-PAGE to deduce relative protein solubility and yield. The study suggested that the optimum expression conditions for the pDT008 construct was using AIM broth medium at 37 °C for 4 h and 16 °C for 16 h.

3.2.5.3 Large scale expression of *pac20* and purification of Pac20

An overnight starter culture (10 mL) was used to inoculate 1 L AIM broth which was incubated using the same conditions as deduced from the *pac20* expression studies. After expression, the cells were harvested and the pellet resuspended in FPLC loading buffer A and lysed by sonication. The supernatant was collected by centrifugation and loaded onto a ÄKTA express FPLC using a programme consisting of an affinity chromatography purification step and a size exclusion chromatography step using a pre charged Ni²⁺ column and a Superdex 75 Hiload HP gel-filtration column which had been previously calibrated using molecular weight gel-filtration standards (as described in chapter 2). Initially, no protein was eluted from the chromatographic matrix onto the gel filtration column, however, through SDS-PAGE analysis it had been confirmed that soluble Pac20 protein was present in the fraction. Two reasons could explain the absence of protein: the protein was being retained by the chromatographic matrix (the concentration of chelating agent was too low to decrease the interactions between the protein and the matrix) or the protein's hexa-histidine affinity tag was not binding to the column matrix and eluting in the flow through. To try to overcome this issue, it was first repeated, increasing the concentration of chelating agent (imidazole) in the ÄKTA buffer B from 0.5 M to 1 M. As a result, the protein purification was performed using the manual ÄKTA FPLC and so only the affinity chromatography step was carried out. The eluted fractions from this purification were collected, the protein sample concentrated to approximately 14 mg mL⁻¹ and the presence and purity of Pac20 confirmed by SDS-PAGE analysis. Pac20 eluted from the chromatographic matrix at an imidazole concentration of approximately 0.7 M, a relatively high concentration for a protein that only contains a hexa-histidine affinity tag. The purity

Chapter 3 – Cloning, Protein Production and Crystallisation

of Pac20 was greater than 95% which was determined by SDS-PAGE analysis (Figure 3.9). Analysis using the Bradford assay determined that the protein yield was approximately 30 mg L⁻¹ of culture. Pac20 may have associated to the chromatographic matrix more tightly than the majority of other proteins because of a number of reasons; the hexa-histidine affinity tag could be more exposed to the chromatographic matrix in comparison to other His-tagged proteins, making the histidine:nickel complex more stable, or the presence of other basic amino acids such as histidine, arginine or lysine on the surface of the protein could be coordinating with the nickel in the matrix. Analysis of the amino acid composition of Pac20 using PROTPARAM (<http://web.expasy.org/protparam/>) suggests the protein only contains 2.3 % histidine, 6.3 % arginine and 2 % lysine, which in comparison to other proteins, is not remarkably high.

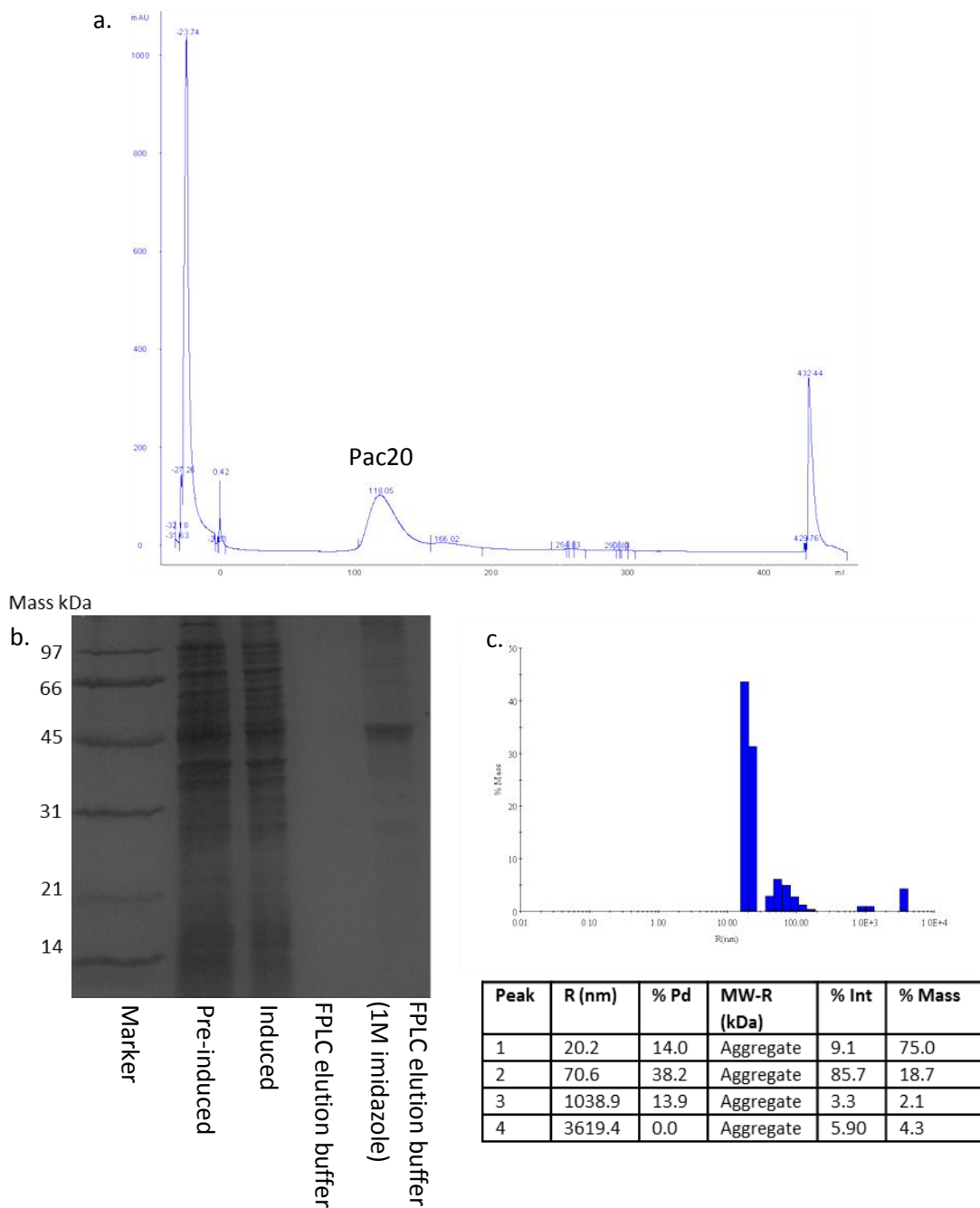


Figure 3.9 Showing (a) the gel filtration chromatogram from the purification of Pac20. Using the calibration curve for the gel filtration column it was estimated that the molecular weight of the protein was approximately 117 kDa (b) the SDS-PAGE result showing the soluble protein profile in the cell culture before and after induction and the purified Pac20 (protein size approx. 38.6 kDa) and (c) The DLS profile of the purified protein, which shows the protein sample is aggregated.

The SDS-PAGE analysis of the purified Pac20 protein was inconsistent with the expected size of the protein. To confirm the identity of the purified protein, a sample was submitted to the JIC Proteomics Facility for intact mass analysis. The analysis confirmed that the major species in the sample had a m/z of 38615.63 Da (Figure 3.10), corresponding closely to the expected mass of Pac20 (38616.8 Da). The SDS-PAGE may have suggested a mass higher than expected due to the low pI of the Pac20 protein.

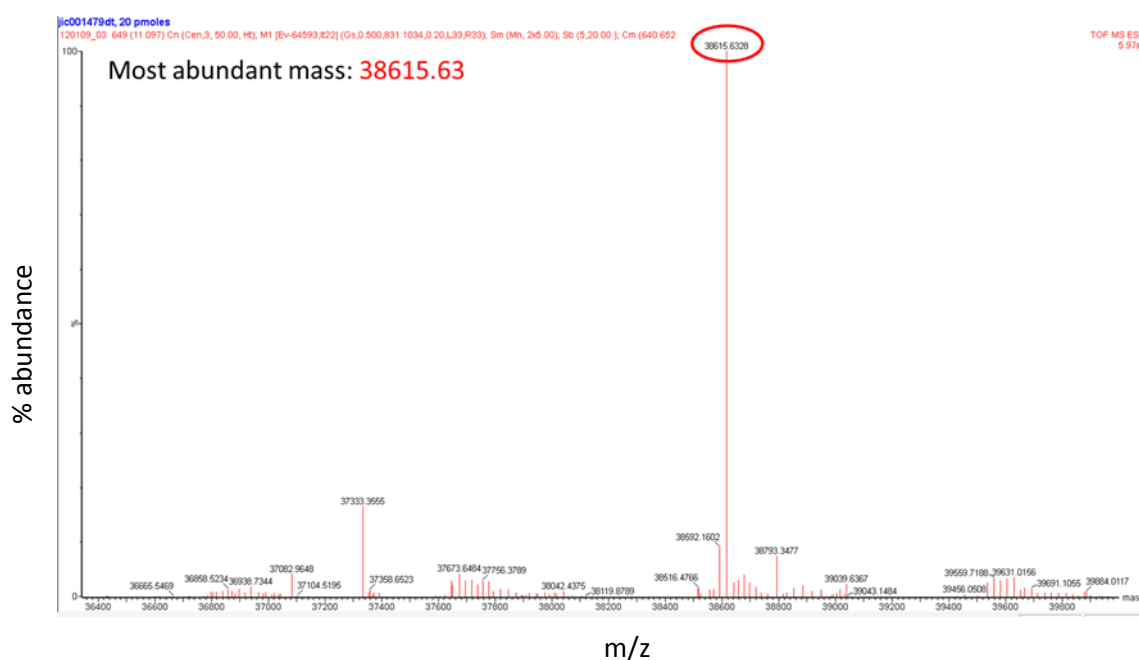


Figure 3.10 MALDI-mass spectrometry analysis of the purified Pac20 sample showing the most abundant mass as 38.615.63 Da, the calculated mass of Pac20 being 38616.8 Da.

3.2.5.4 Crystallisation of Pac20

As the purity of the protein preparation was high (above 95 %), crystallisation trials were undertaken using the sitting drop vapour diffusion format in 96-well MRC plates using five commercially available screens at a temperature of 20 °C. Each drop consisted of 0.3 μL of well solution and 0.3 μL of protein solution and a well volume of 50 μL . The final protein concentration was approximately 7 mg mL^{-1} . Precipitate was observed in the majority of the screens. The presence of a lot of precipitant is consistent with the DLS suggesting the protein is largely aggregated and not happy in solution.

3.2.6 Pac22

The cloning and initial protein production studies for Pac22 were conducted by J. Clouston, UEA. He produced two expression constructs of Pac22; pJC001 and pJC002. pJC001 encoded for 27 extra bases and encoded a protein of 249 amino acids in length, which also contained a N-terminal hexa-histidine tag and thrombin cleavage site adding an additional twenty amino acids onto the protein with sequence MGSSHHHHHSSGLVPRGSH, giving the protein a mass of 30031.4 Da. The pJC002 constructs also contained a N-terminal hexa-histidine tag, however the gene encoded for a slightly smaller protein of 240 amino acids which had a molecular weight of 29032.3 Da. J. Clouston had determined that large scale production of either of the Pac22 constructs could be carried out in LB broth with a selectable marker (in this case kan as the plasmid backbone was that of pET28a(+)).

3.2.6.1 Large scale expression of *pac22* and purification of Pac22

An overnight starter culture (10 mL) of the smaller Pac22 construct (pJC002) was used to inoculate 1 L LB broth, which was incubated using the same conditions as deduced from the Pac22 expression studies. After expression, the cells were harvested and the pellet resuspended in FPLC loading buffer A and lysed by sonication. The supernatant was collected and loaded onto a ÄKTA express FPLC using a programme consisting of an affinity chromatography purification step and a size exclusion chromatography step using a pre charged Ni²⁺ column and a Superdex 75 Hiload HP gel-filtration column, which had been previously calibrated using molecular weight gel-filtration standards (as described in chapter 2). From the gel filtration step, it was suggested that Pac22 is a monomer as it was estimated that the molecular weight of protein being eluted from the column was approximately 32.5 kDa. The Pac22 fractions were pooled and concentrated to approximately 10 mg mL⁻¹ (determined by Bradford analysis). The purity of Pac22 was greater than 95%, which was determined by SDS-PAGE analysis (Figure 3.11). Analysis using the Bradford assay determined that the protein yield was approximately 16 mg L⁻¹ of culture. DLS analysis suggested that the protein existed as a variety of species.

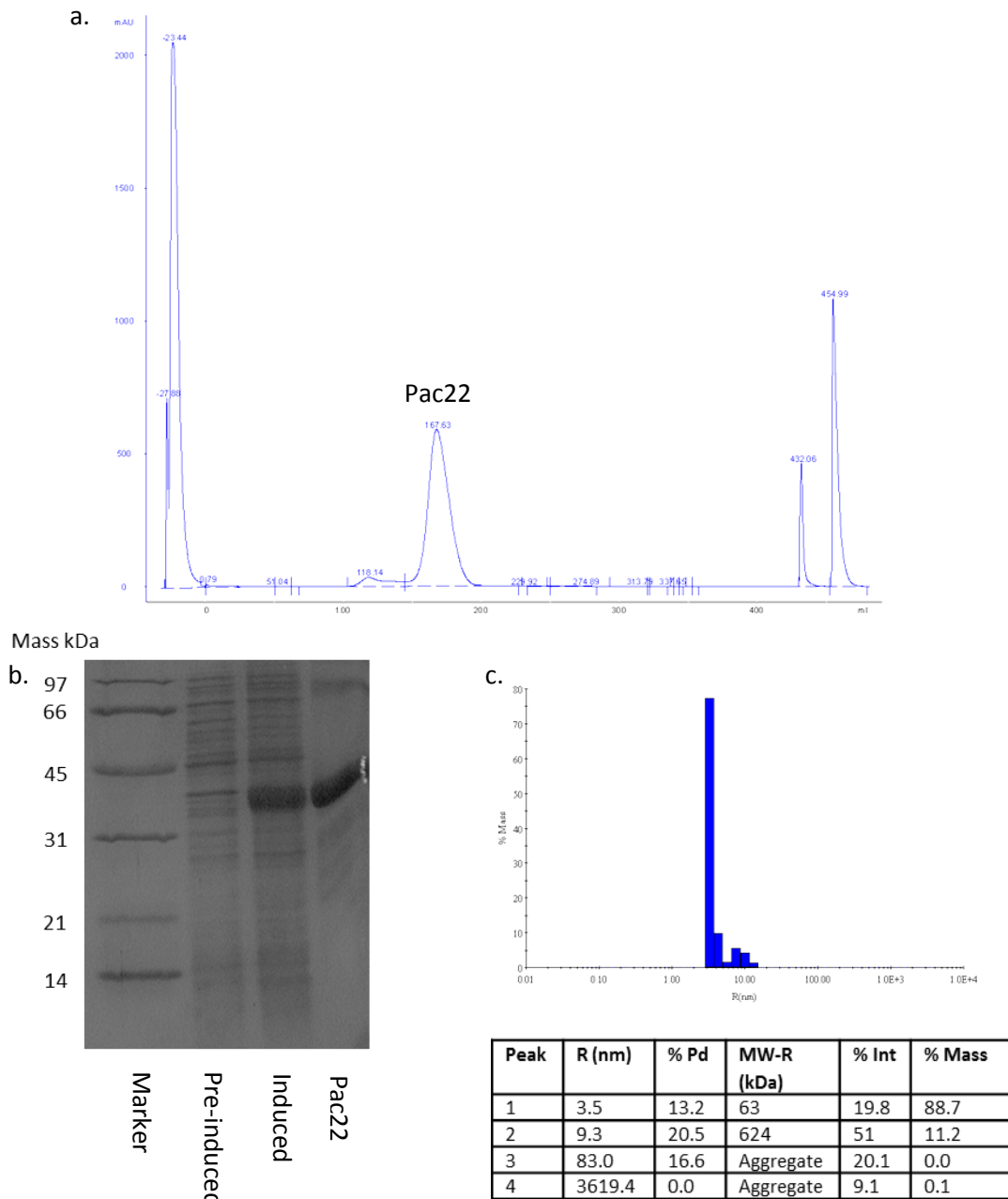


Figure 3.11 Showing (a) the gel filtration chromatogram from the purification of Pac22. Using the calibration curve for the gel filtration column it was estimated that the molecular weight of the protein was approximately 32.5 kDa (b) the SDS-PAGE result showing the soluble protein profile in the cell culture before and after induction and the purified Pac22 (size approx. 29.0 kDa) and (c) The DLS profile of the purified protein suggests that the protein exists as a number of species.

Although the gel filtration estimation of size is consistent with the size of the smaller Pac22 construct, SDS-PAGE analysis suggested that the protein species was of a larger size

(approximately 35 kDa). A sample of the purified protein was submitted to the JIC proteomics facility for mass analysis. This gave a mass of 29019.7 Da, which is considered within the value calculated from the amino acid sequence of Pac22 (29032.3 Da).

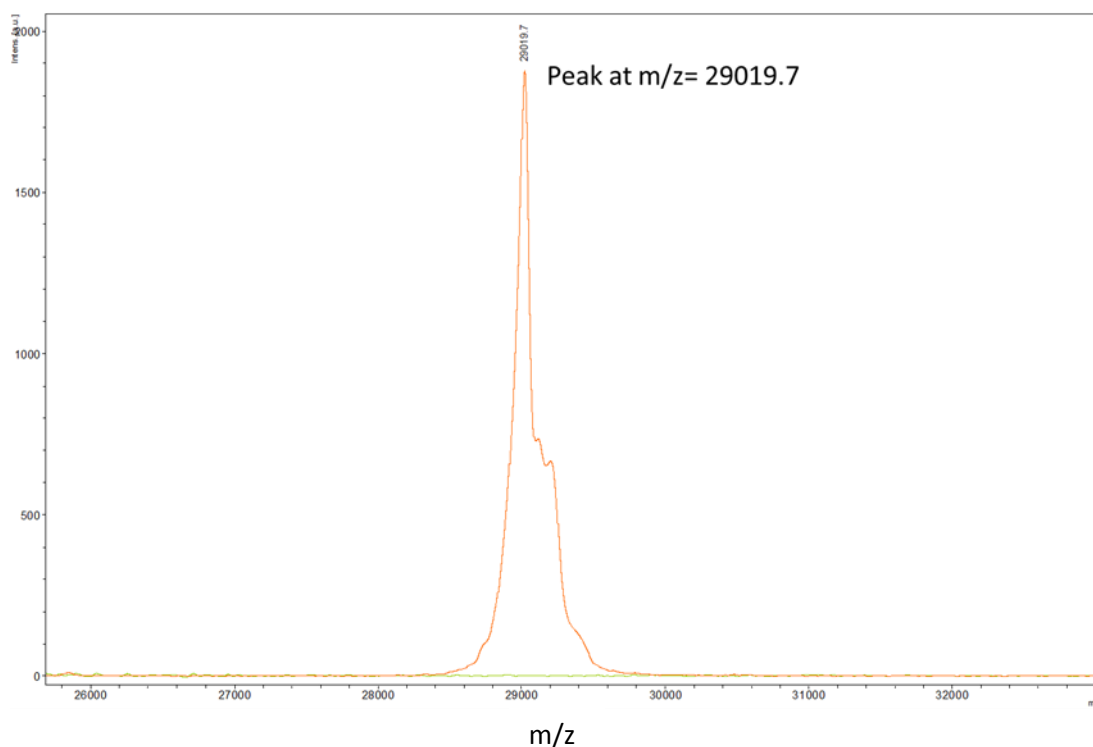


Figure 3.12 Mass spectrometry (electrospray spectrum) result for purified Pac22 submission. The smaller construct for Pac22 has an expected MW of 29032.3 Da. Additional peaks are present approximately +200 larger than the expected mass of Pac22 which may suggest that some of the protein is bound to another molecule, for example, PLP (a predicted co-factor for the protein).

3.2.6.2 Crystallisation of Pac22

As the purity of the protein preparation for both constructs was high (above 95 %), crystallisation trials were undertaken using the sitting drop vapour diffusion format in 96-well MRC plates using five commercially available screens at a temperature of 20 °C. Each drop consisted of 0.3 µl of well solution and 0.3 µl of protein solution and a well volume of 50 µl. The final protein concentration was approximately 8 mg mL⁻¹. After six weeks, no crystals were observed in any of the screens, however, some granular precipitate was observed.

Chapter 3 – Cloning, Protein Production and Crystallisation

From bioinformatics analysis using the DISOPRED server (<http://bioinf.cs.ucl.ac.uk/disopred/>) (Ward et al., 2004) it was determined that the affinity tag present on the N-terminus of the protein was rather disordered and could be affecting the ability of the Pac22 to crystallise. An unsuccessful attempt to remove the affinity tag was attempted using the thrombin cleavage method (outlined in chapter 2) but the tag was eventually successfully cleaved using the DAPase method (outlined in chapter 2).

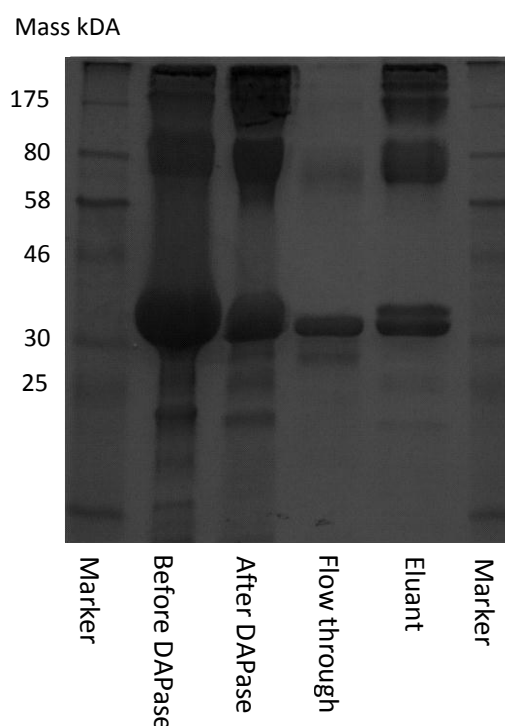


Figure 3.13 SDS-PAGE analysis of the DAPase cleavage of the Pac22 N terminus affinity tag. The pre treated (Before DAPase) and post treated (After DAPase) solutions are shown followed by the purified tag-less Pac22 (Flow through) and the remaining tagged protein (Eluant). It should be stated that the eluent contained both tagged and untagged Pac22 which may suggest the protein associates with itself and is not found in a monomeric state.

The tag-cleaved Pac22 protein was concentrated to approximately 16 mg mL^{-1} and crystallisation trials were undertaken using the sitting drop vapour diffusion format in 96-well MRC plates using five commercially available screens at a temperature of $20 \text{ }^{\circ}\text{C}$. Each drop consisted of $0.3 \text{ }\mu\text{L}$ of well solution and $0.3 \text{ }\mu\text{L}$ of protein solution and a well volume of $50 \text{ }\mu\text{L}$. The final protein concentration was approximately 8 mg mL^{-1} . After a

Chapter 3 – Cloning, Protein Production and Crystallisation

week a crystal was observed in one of the drops with well conditions; 0.2 M zinc acetate, 0.1 M sodium cacodylate pH 6.5, 10 % (v/v) 2-propanol.

To determine if the crystal was that of salt or protein, it was assessed for tryptophan fluorescence using a UV pen and the fluorescence observed in a dark room. The crystal showed fluorescence, suggesting that it was a protein crystal.

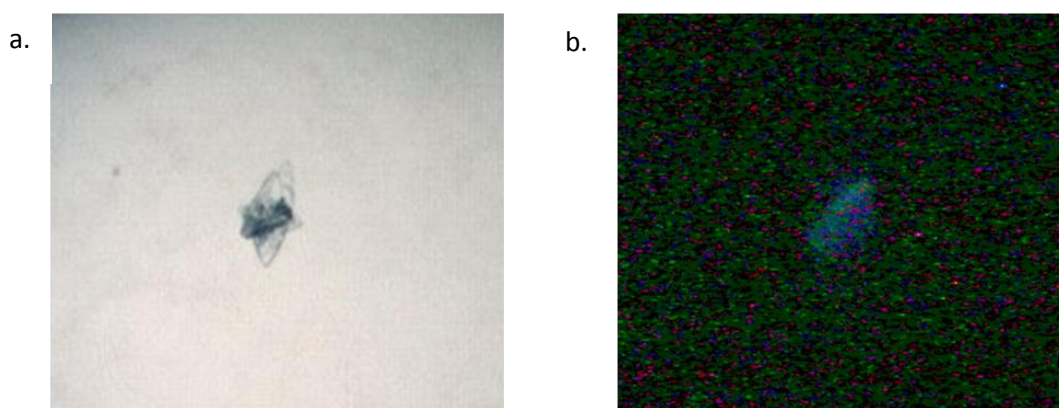


Figure 3.14 Showing (a) the protein crystal observed in the crystallisation drop condition and (b) the same drop under UV radiation, showing weak fluorescence

As it was confirmed that the crystal was likely to be a protein crystal, a series of optimisations were carried out, using similar conditions to those of the initial hit, but with variations in the concentration of precipitant (PEG), concentration of the protein and buffers, the pH and with the addition of ligands (in this case SAM and SAH, as these are likely cofactors, and glycine and threonine, which are likely substrates and products) and incubating the plates at 20 °C, RT and 4 °C. Although many optimisation attempts were carried out, none produced crystals for further analysis.

3.2.7 Gene cloning of hypothetical protein genes

The cloning of the hypothetical genes to produce expression constructs was carried out in the same way. The gene of interest from the pacidamycin gene cluster was amplified from a cosmid DNA template, using cosmid 2H-5 which contained the minimal pacidamycin gene cluster. Primers were used in a standard PCR amplification procedure to produce a DNA

Chapter 3 – Cloning, Protein Production and Crystallisation

fragment corresponding to the desired gene size. The amplified DNA was digested by two restriction enzymes to produce 'sticky ends' and ligated into a restricted pET28a(+) vector to produce an expression construct. The construct encoded for the protein of interest (amino acid sequence Appendix 1) tethered to an N-terminal hexa-histidine tag and thrombin cleavage site. The addition of the tag added an additional twenty amino acids onto the N-terminus of the protein with the sequence GSSHSHHHSSGLVPRGSHM . The sequences of the constructs were confirmed by analysis by the DNA sequencing service at Cambridge University, UK and introduced into *E. coli* BL21(DE3) cells by chemical transformation. Table 3.2 reports the specific primers used, construct names and the genes cloned.

Table 3.2 The parameters used for the cloning of the hypothetical protein genes

Gene	Protein	Primers used	Restriction enzymes used	Molecular weight of protein + tag (Da)	Construct name
<i>pac1</i>	Pac1	Pac1F and Pac1R	NdeI and XhoI	30688.0	pDT001
<i>pac2</i>	Pac2	Pac2F and Pac2R	NdeI and XhoI	41477.3	pDT002
<i>pac7</i>	Pac7	Pac7F and Pac7R	NdeI and XhoI	29346.9	pDT003
<i>pac13</i>	Pac13	Pac13F and Pac13R	NdeI and XhoI	15647.5	pDT004

3.2.7.1 Discovery of the function of *pac2* and a homolog of *pac1*

During this study, a paper was published by the Walsh Laboratory in Harvard, reporting the function of *pac2* in the biosynthesis of pacidamycin. It was found that Pac2 is a tRNA-dependent transferase which catalyses the addition of an alanyl residue to the N-terminus of the tetrapeptide of pacidamycin while it is tethered to Pac8. The alanyl group is transferred from alanyl-tRNA and completes the assembly of the pacidamycin peptide backbone (Zhang et al., 2011).

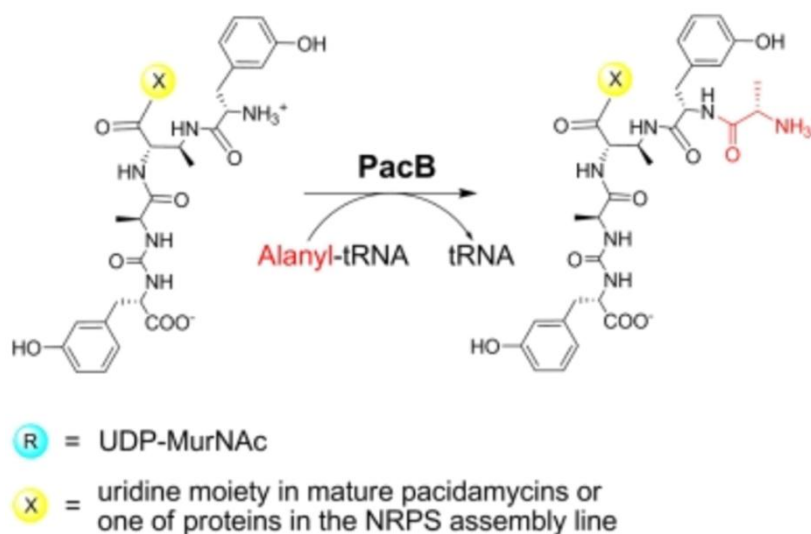


Figure 3.15 Image from Zhang et al. (2011) paper, illustrating the function of Pac2 (designated PacB by Zhang et al.) transferring an alanyl group from tRNA onto the pacidamycin structure as is it tethered to the NRPS assemble line.

Further to this, in 2013 a homolog of *pac1*, *ssaA*, was found to be a transcriptional regulator for the sansamycin gene cluster (sansamycins also being a member of the uridyl peptide class of natural products) (Li et al., 2013).

3.3 Conclusions

Gene cloning and protein production studies were carried out on the four 'hypothetical protein' genes and the postulated DABA biosynthetic genes of the pacidamycin gene cluster found in *S. coeruleorubidus*. Proteins postulated to be involved in the biosynthesis of DABA that were successfully heterologously produced in *E. coli* were used in protein crystallisation trials, and a number of optimisations of initial hits were attempted, particularly for Pac19 and Pac22, to attempt to produce protein crystals, however, these attempts failed to produce X-ray diffraction quality protein crystals. Protein crystals were produced from the Pac17 crystallisation trials. The crystal structure determination of Pac17 will be discussed fully in chapter 4.

Further studies into the structure and function of the hypothetical proteins were not continued in this study due to problems in producing gene disruptions (discussed in chapter 5) and the discovery of the function of Pac2 by Zhang et al. (2011).

Structural Investigation of the Pac17 Protein

Chapter 4 – Structural Investigation of the Pac17 Protein

4.1 Introduction

Pac17 is a protein encoded for by the gene *pac17* in the pacidamycin gene biosynthetic cluster of *Streptomyces coeruleorubidus*. At present the gene is believed to be involved in the biosynthesis of the core DABA residue of the pacidamycin structure as previously described in the introductory chapter of this thesis (chapter 1). As stated in chapter 3, the *pac17* gene is 1497 bp in length and has a high GC content of approximately 71 %. It is hypothesised that *pac17* is translationally coupled to *pac18* which in turn is coupled to *pac19* (illustrated in Figure 4.1).

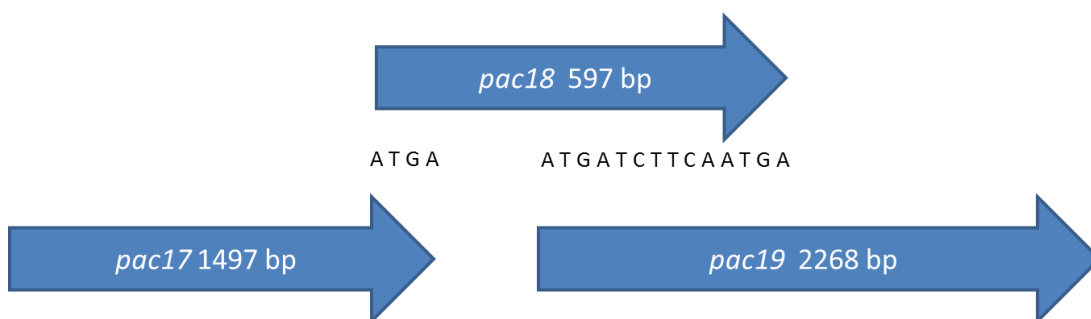


Figure 4.1 Diagrammatic representation of the translational coupling of *pac17* through *19*, showing the bases that are present in the gene overlaps.

From the bioinformatics analysis discussed in chapter 3, the amino acid sequence of Pac17 suggests that it is a member of the lyase like protein family with the best 'hits' in BLAST being that of argininosuccinate lyases, aspartases and fumarases. Apparent orthologs of Pac17 are observed in a number of streptomycetes, most notably in *S. roseosporus* which possesses a homologous gene cluster to *S. coeruleorubidus*.

The gene cloning, protein production and purification and initial crystallisation of Pac17 were carried out as described in chapter 3. Within this chapter, crystal optimisation and X-ray data collection will be discussed, along with the presentation of the apo and ligand bound structures of Pac17, the ligand being that of aspartate. Further evidence for the likely active site region and substrate is also investigated by the co-crystallisation of the protein with other ligands and also site directed mutagenesis studies.

4.2 Results

4.2.1 Optimisation of the crystallisation of Pac17

From the initial crystallisation screens (described in chapter 3), conditions where crystal 'hits' were observed were optimised in 24 well hanging-drop vapour diffusion plates. The best crystals appeared in a precipitant solution consisting of 15 % (w/v) PEG 3350, 0.2 M potassium sodium tartrate, 0.1 M bis tris propane pH 7.5. The crystals formed were approximately 100 X 50 X 300 μm . This optimisation condition was repeated, with the addition of 15 % (v/v) glycerol to the buffer in the well solution in preparation for cryogenic data collection at the synchrotron. The resulting crystals were mounted onto litho-loops, flash cooled by plunging into liquid nitrogen and stored ready for transport to the synchrotron.

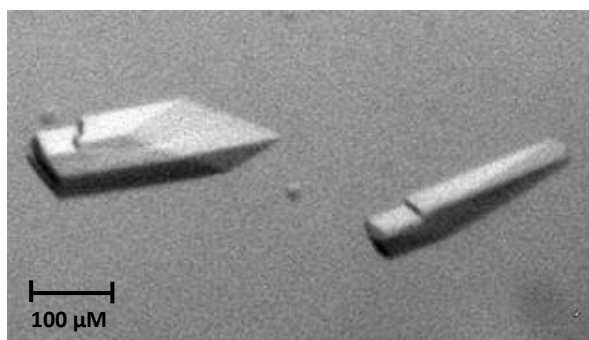


Figure 4.2 Picture of stereotypical crystals observed in the optimisation screens of Pac17.

4.2.2 Native data collection of a single Pac17 crystal

For data collection, a single crystal was transferred to the goniostat on station I02 at Diamond Light Source (Oxfordshire), maintaining the temperature at $-173\text{ }^{\circ}\text{C}$ and diffraction recorded. A total of 1000 images were collected in a single sweep at 0.2° oscillations to a maximum resolution of 1.9 \AA . The space group was C2 with unit cell parameters $a = 214.12$, $b = 70.88$, $c = 142.22\text{ \AA}$, $\beta = 92.96$. The cell parameters were used to predict the contents of the asymmetric unit (ASU). The values of 61.8% or 49.0% corresponded to three or four monomers present in the ASU, respectively. Given that the biological units of the closest structural homologues are tetrameric, it seemed likely that the ASU of Pac17 contained a homotetramer.

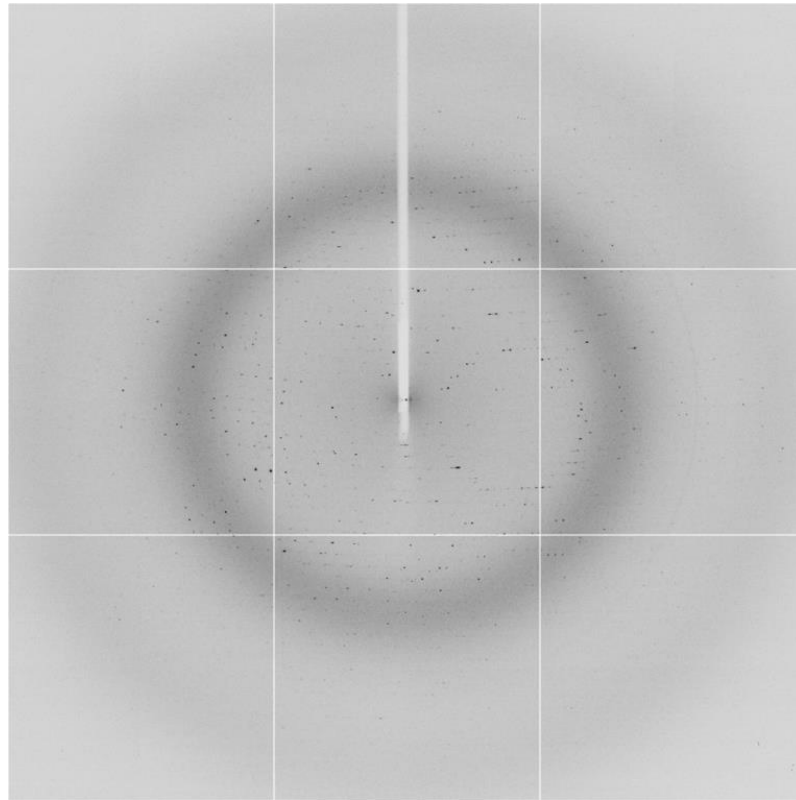


Figure 4.3 Diffraction image of Pac17 native crystal (data collected from station I02 at Diamond Light Source).

4.2.2.1 Pac17 data processing

The data collected were processed using XDS (Kabsch, 2010) and scaled using *SCALA* (Evans, 2006). The subsequent processing and analysis were carried out using the CCP4 suite of programs (Winn et al., 2011). To be able to calculate the free crystallographic R factor (R_{free}) value during model refinement, a random 5% of the reflections collected were placed in a subset that was not used in refinement (Brünger, 1993). The data collection statistics are reported in Table 4.1.

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.1 Statistics for X-ray data collection of Pac17.

Number of crystals	1
Beamline	I02, Diamond Light Source, UK
Wavelength (Å)	0.9795
Detector	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	290.7
Rotation range per image (°)	0.2
Exposure time per image (s)	0.25
Beam transmission (%)	27.2
Total rotation range (°)	200.0
Resolution range (Å)	67.28 - 1.90 (2.00 - 1.90)
Space Group	C2
Cell parameters (Å/°)	$a = 214.12, b = 70.88, c = 142.22, \beta = 92.96$
Total no. of measured intensities	672568 (74768)
Unique reflections	166584 (23088)
Multiplicity	4.0 (3.2)
Mean $I/\sigma(I)$	8.5 (2.0)
Completeness (%)	99.3 (95.1)
$R_{\text{merge}}^{\dagger}$	0.127 (0.583)
$R_{\text{meas}}^{\ddagger}$	0.147 (0.704)
Wilson B value (Å ²)	15.6

$^{\dagger} R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. $^{\ddagger} R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Inspection of the self-rotation function calculated from the experimental data using MOLREP (Vagin and Teplyakov, 1997) suggested that the ASU consisted of a homo-tetramer with 222 symmetry (Figure 4.4).

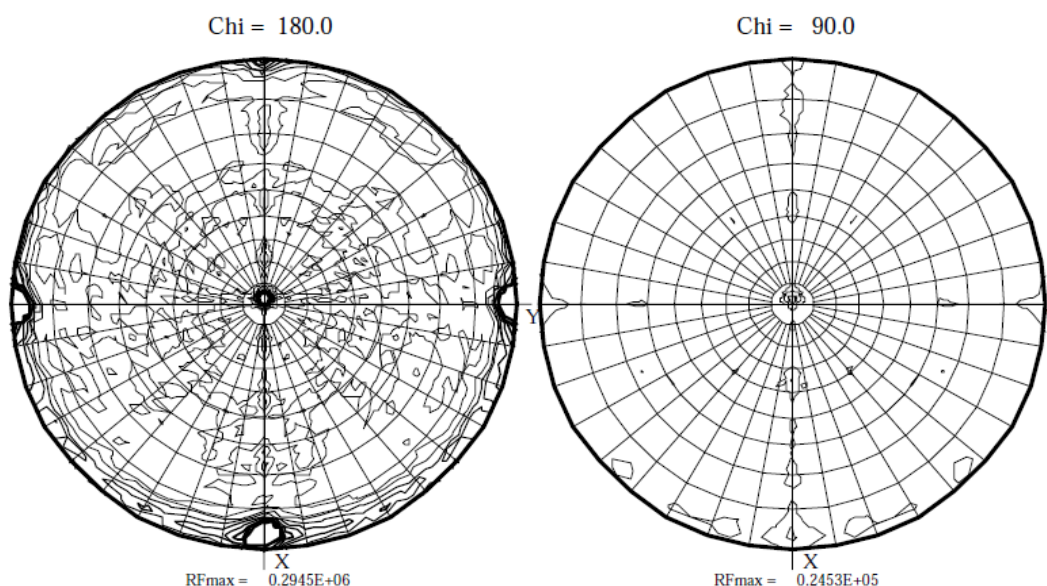


Figure 4.4 Self rotation function generated in MOLREP for the Pac17 structure showing sections $\text{Chi} = 180.0^\circ$ and $\text{Chi} = 90.0^\circ$. The peaks at $\text{Chi}=180^\circ$ indicate mutually perpendicular two-fold axes consistent with 222 symmetry and showing that the crystal symmetry is close to orthorhombic. The lack of peaks on $\text{Chi} = 90.0^\circ$ show that there is no four-fold symmetry.

4.2.2.2 Molecular replacement to solve the Pac17 apo structure

Through analysis of the primary sequence of Pac17 against entries in the Protein Data Bank (PDB) (www.rcsb.org) (Berman et al., 2000), a structure was identified as a potential candidate for use in molecular replacement. This candidate was PDB entry 2E9F, an argininosuccinate lyase from *Thermus thermophilus* which had a sequence coverage to Pac17 of 71% with an identity of 31% giving an overall identity of 22%. Figure 4.5 shows the alignment of Pac17 and 2E9F from CLUSTAL OMEGA (www.ebi.ac.uk/Tools/msa/clustalo) (Chenna et al., 2003).

Chapter 4 – Structural Investigation of the Pac17 Protein

```
Pac17      ---VRLTGRLLAAPDDFIRQEFLEPQFRHEVAHLLRWYVLI EKALLREYHRLGVLSARQV
2E9F      MAHRTWGGRFEGEPDALAARFNASLAFDRALWREDLWQ---NRVHARMLHAVGLLSAEEL
          **:. ** : : . * : : * : : . * * :*:***:

Pac17      EQLAGALDGLTPEELTAG---AGAALSDIALAMETRVAKALPEFVPAWHVDRSRNDLQAC
2E9F      EAILKGLDRIEE-IEAGTFPWRREELEDVHMNLEARLTELVGPPGGKLTARSRNDQVAT
          * : . ** : * : **          * . : : * : : * * * * * *

Pac17      AQLLYGREQVSEIAAMLGELAEVAHER--AAHDVTSMPGYTHLQSAQIITPGFYLSAVV
2E9F      DLRLYLRGAIDELLALLLALRRVLVREAEKHLDPYVLPGYTHLQRAQPVLLAHWFLLAYY
          ** * : . * : * * * * . .          * : * * * * * * * * : . . : *

Pac17      EHALRASSRLGTYDRINFSQLGAGPMAGQELAWDRDWLAQAIGCAGVPVHALAAVASRE
2E9F      EMLKRDAGRIEDAKERLINESPLGAAALAGTGFPIDRHFTARELGFKAPMRNSLDAVASRD
          * * : . * * : : * * * * . * * : * * : * : * * . * : . * * * * * :

Pac17      WLLDVAADIASAGVGFSLFLTDLMAWSSSAYGLVELPELAGISAAMPQKNYPLLERLR
2E9F      FALEVLSALNIGMLHLSRMAEELILYSTEEFGFVEVPDAFATGSSIMPQKNPDILELIR
          : * : * : : . : : * : : * : : * : * * * * * * * * : * * : * * * * * * * *

Pac17      GRTGHLTAFYVDFATGQRNTPYSNMVEV--SKEAGLHASTMFGAIRAVLTGFTLVVDKLQ
2E9F      AKAGRVLGAFVGLSAVVKGLPLAYNKDLQEDKEPLDA---LATYRDSLRLAALLPGLK
          . : * : : . * : : : * : : : . * * * * . : : * * : : : * :

Pac17      WRTDRMRVCEEDHFGGFSLANEL---TLRAGVPWRTAQVIAGRYVTAVLAEGDGRAGHS
2E9F      WRRERMWRAE---GGYTLATELADYLAEKGLPFREAHHVGRVLRRLVEEGRALKDLT
          ** : * * . *          * * * * * * * * . * * * * * : : * * . :

Pac17      NPAALDSAAREAGYPLDDPASLYLNSVDVDAQLRAKVSAGSASPDSVVALLAEQRLDA
2E9F      LEE-----LQAHHPLEAEDA--LPLLRLLETAIHRRRSYGGTAPFAVREERLEEAKKEVGL
          : * : * * : : : : : : * * * * * * * * * * * * * * : :

Pac17      LTAENNARTATVRDAIAATDAAVRPTAGGVV
2E9F      D-----
```

Figure 4.5 Alignment of Pac17 with 2E9F, generated by Clustal Omega.

Molecular replacement was carried out using the program PHASER (McCoy et al., 2007) from the CCP4 suite of programs and the polyaniline monomer and tetramer structures of 2E9F (generated by CHAINSAW (Stein, 2008)). A convincing solution was obtained with the tetramer model. This gave sensible packing in the unit cell with no clashes and no large gaps in the lattice. A self-rotation function calculated from the structure factors was comparable to that calculated from the experimental structure factors which suggests the orientation is correct. Since one of the model twofold axes was parallel to the crystallographic twofold axis, the application of twofold crystallographic symmetry resulted in similarly oriented tetramers within the same unit cell which were therefore also related by translational symmetry alone (Figure 4.6).

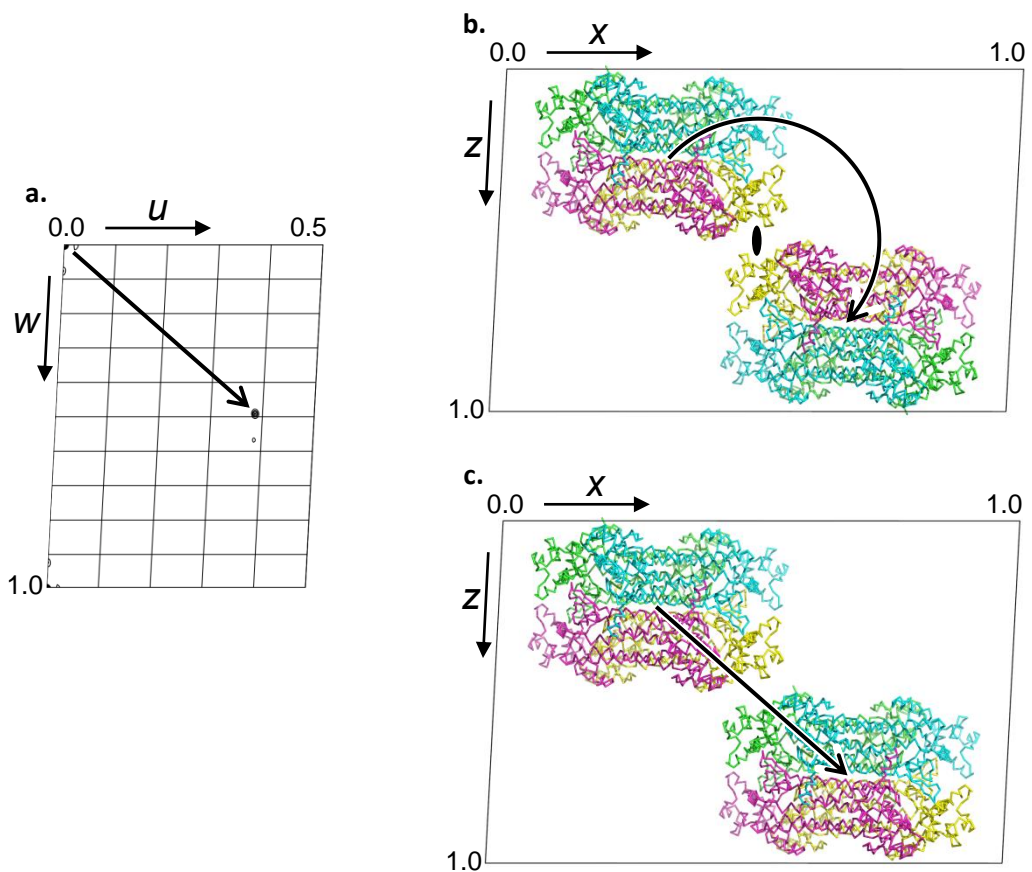


Figure 4.6 Showing (a) Self Patterson function (section $v = 0$) of the experimental data, showing the pseudotranslation vector of 0.386, 0.000, 0.491 and (b) and (c) showing that the application of twofold crystallographic symmetry and translation by the pseudotranslation vector, respectively, are essentially equivalent (these images were used in Tromans et al., 2012).

Analysis of the PHASER (McCoy et al., 2007) output when using the polyaniline monomer 2E9F model showed that the four subunits were orientated correctly and two dimers would form the biological unit, however, PHASER failed to combine these correctly in the translation function to form the tetramer. An explanation for PHASER's failure to do this could be due to how densely the tetramer is packed (the subunits form a tight core of helices, similar to that of membrane protein), another explanation being the influence of the pseudo-symmetry present due to the translation and rotation function giving a copy of the tetramer in the exact same orientation. It is worth noting that the estimated oligomeric state of Pac17 from gel-filtration was believed to be that of a trimer, with an estimated oligomeric mass of approximately 150 kDa. The densely populated core of the tetramer

could have caused the molecular mass estimated by size exclusion chromatography to be towards a lower oligomeric state.

4.2.2.3 Model building and refinement of the Pac17 Apo Structure

The structure produced by PHASER using the polyalanine 2E9F tetramer was rebuilt in COOT (Emsley and Cowtan, 2004). Density improvement of the resulting map was carried out in PARROT (Zhang et al., 1997) and the resulting phases used in BUCANEER (Cowtan, 2006) for automated model building. Using BUCANEER (Cowtan, 2006) with its default settings (5 cycles) was found to be insufficient to produce a workable Pac17 model with an acceptable model being produced after 15 cycles. This was shown by the significant improvement in the figure of merit (FOM), R_{free} and fractional completeness of the model between cycles 5 and 15 (Figure 4.7).

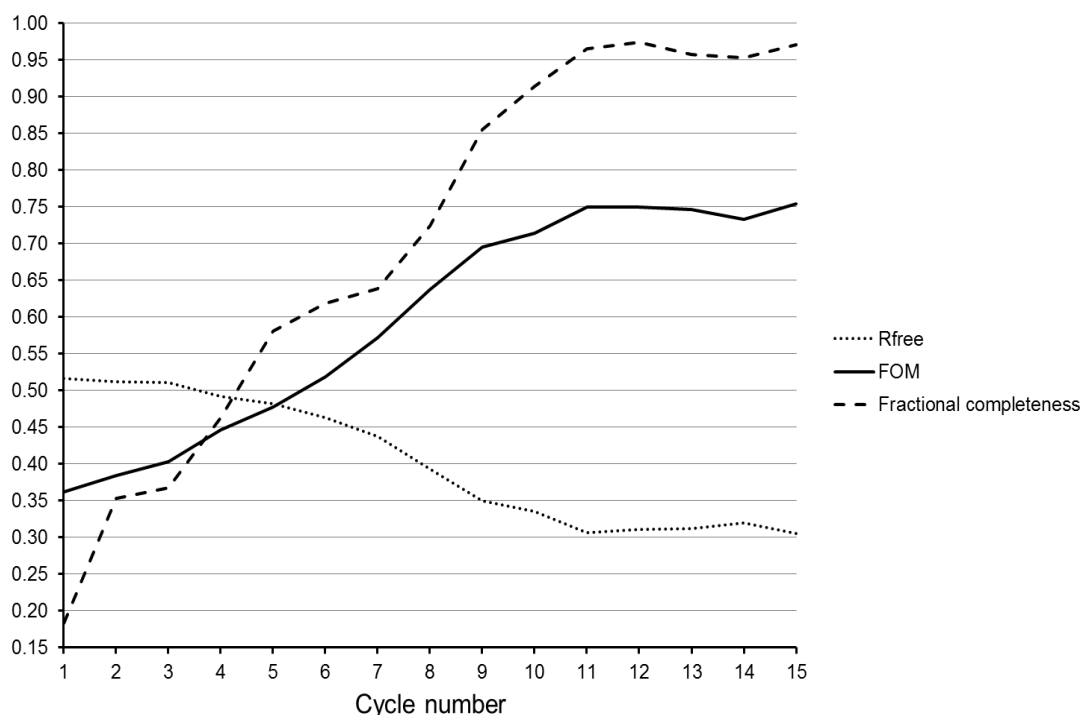


Figure 4.7 Showing the change in the R_{free} , FOM and fractional completeness of the Pac17 model after each cycle of BUCANEER (Cowtan, 2006).

Chapter 4 – Structural Investigation of the Pac17 Protein

The output model from BUCANEER (Cowtan, 2006) was completed by rounds of manual building using the program COOT (Emsley and Cowtan, 2004) and refinement carried out by the CCP4 suite of programs. Completion of model building was difficult due to the dense helical core of the protein with a number of flexible loops within the structure difficult to build. Interestingly, these loops lie in the conserved regions of the protein, deduced by CONSURF which will be described later in the chapter. The methods used for the structure solution of apo Pac17 is summarised in Figure 4.8.

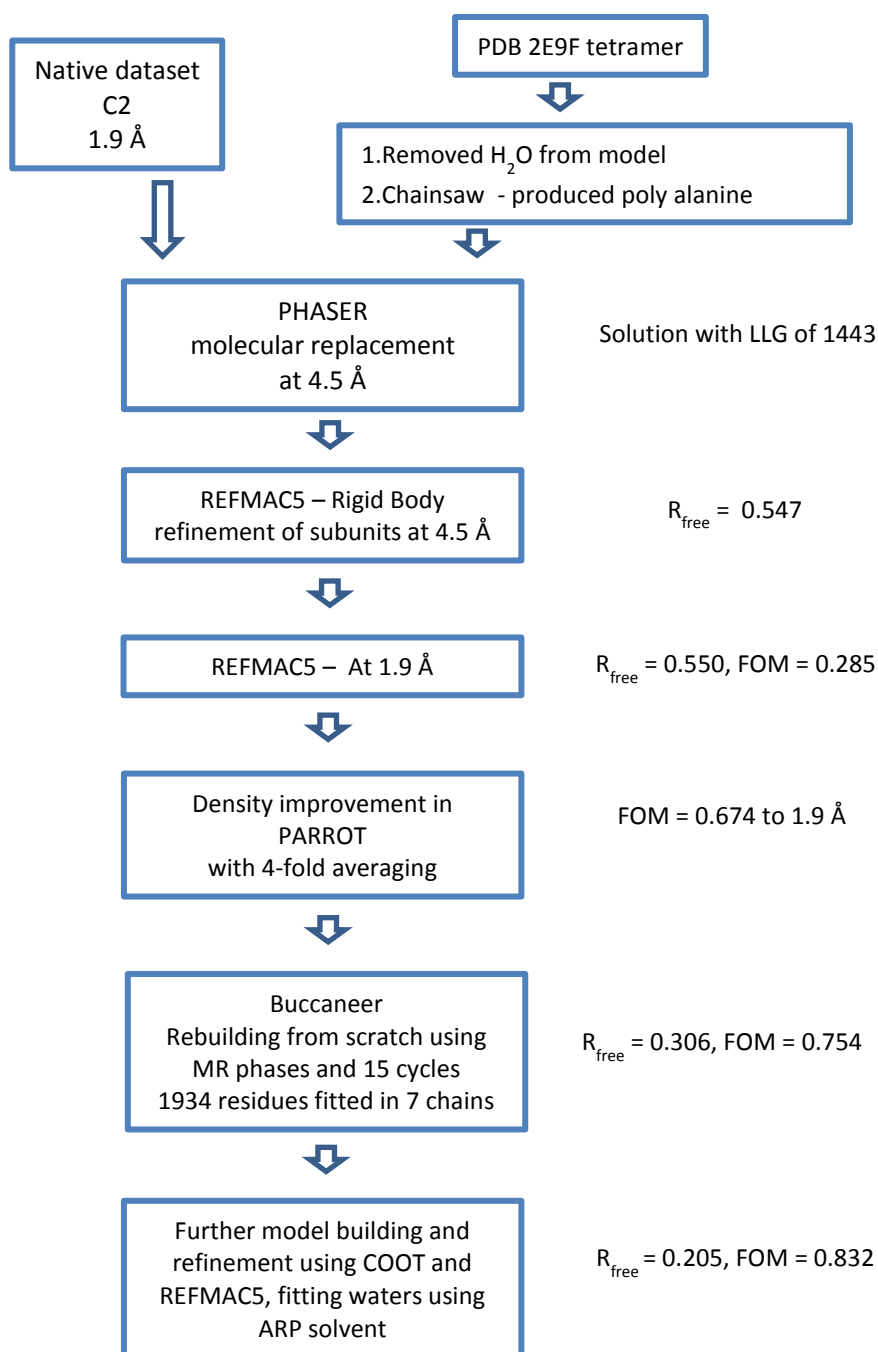


Figure 4.8 Summary of the solving of the apo structure of Pac17.

4.2.3 Higher resolution dataset of apo Pac17

Further on in this chapter, co-crystallisation of Pac17 with a number of ligands will be discussed. One of the datasets obtained from a crystal of Pac17 produced in a condition of 20 % PEG 3350 (w/v), 0.2 M potassium sodium tartrate, 0.1 M bis tris propane, pH 7.5 and 1mM asparagine, cryo-protected with a cryo-protectant of the condition containing 15 %

Chapter 4 – Structural Investigation of the Pac17 Protein

glycerol was found to be identical to that of the apo Pac17 model, however the resolution of this dataset was 1.8 Å.

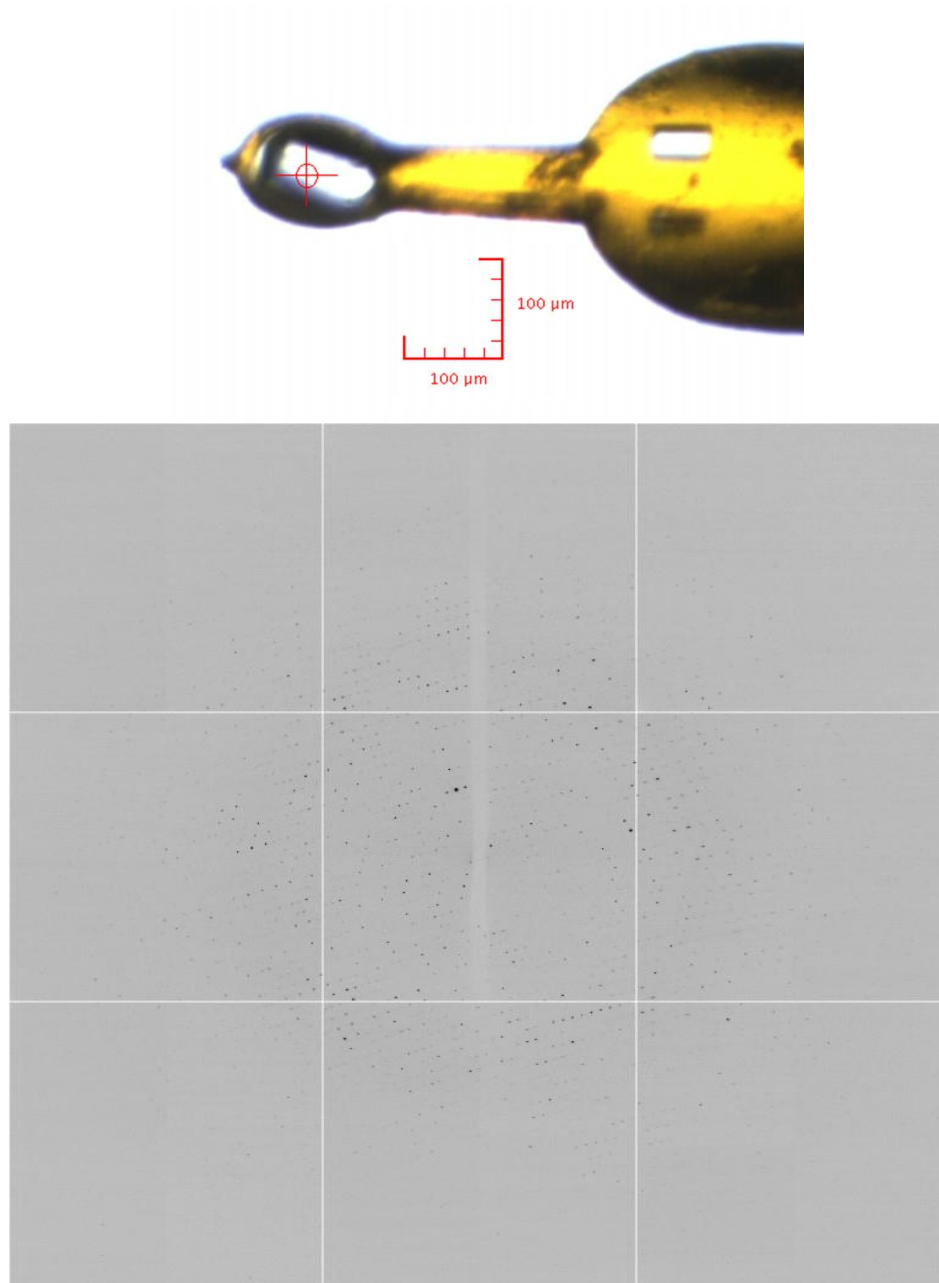


Figure 4.9 Image showing the crystal used for data collection loaded into a litholoop (top) and an example of the diffraction data collected.

The data were collected at the Diamond Light Source (Oxfordshire) on station I04 with the temperature maintained at -173 °C and diffraction data recorded. A total of 900 images

Chapter 4 – Structural Investigation of the Pac17 Protein

were collected in a single sweep of the crystal in 0.2° oscillations. The data was integrated and merged using XDS and SCALA (Evans, 2006), respectively, and the phases solved by rigid body refinement in REFMAC5 (Murshudov et al., 1997) using the slightly lower resolution Pac17 model that had been previously built and refined. This resulting model was improved by manual model building in the program COOT (Emsley and Cowtan, 2004) and refinement using the program REFMAC5 (Murshudov et al., 1997). The statistics for this data collection were:

Table 4.2 Statistics for X-ray higher resolution data collection of the Pac17.

Number of crystals	1
Beamline	I04, Diamond Light Source, UK
Wavelength (Å)	0.9795 Å
Detector	ADSC Q315r
Crystal-to-detector distance (mm)	272.0
Rotation range per image (°)	0.2
Exposure time per image (s)	0.25
Beam transmission (%)	35
Total rotation range (°)	180.0
Resolution range (Å)	44.16 - 1.81 (1.84 – 1.81)
Space Group	C2
Cell parameters (Å/°)	a = 213.54, b = 70.56, c = 142.28, β = 93.05
Total no. of measured intensities	692761 (34465)
Unique reflections	191246 (12502)
Multiplicity	3.6 (2.7)

Chapter 4 – Structural Investigation of the Pac17 Protein

Mean $I/\sigma(I)$	13.0 (1.6)
Completeness (%)	98.3 (83.8)
$R_{\text{merge}}^{\dagger}$	0.068 (0.515)
$R_{\text{meas}}^{\ddagger}$	0.098 (0.706)
Wilson B value (\AA^2)	15.8

$^{\dagger} R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. $^{\ddagger} R_{\text{meas}} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Completion of model building of the 1.8 \AA resolution apo structure for Pac17 was done by the building in of waters into the remaining positive electron density by using the CCP4 program ARP/wARP (Perrakis et al., 2001) and checking manually in COOT (Emsley and Cowtan, 2004). The method of structural solution is summarised in Figure 4.10. The final structural parameters are summarised in Table 4.3.

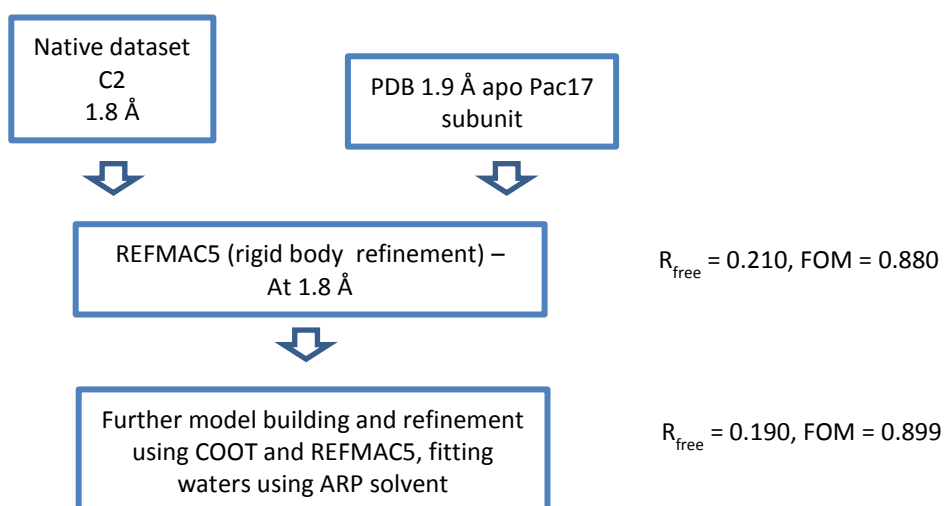


Figure 4.10 Summary of the solving of the higher resolution apo structure of Pac17.

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.3 Final structural parameters of apo Pac17 structure.

Resolution Range (Å)	44.16 – 1.81
R _{work}	0.1656
R _{free}	0.1897
Ramachandran plot: favoured/allowed (%)	98.9/99.8
R.M.S bond distance deviation (Å)	0.015
R.M.S bond angle deviation (°)	1.48
No. of residues in protein: Chain A/B/C/D	477/478/483/482
No. of water molecules/K ⁺ /glycerol	1933/4/4
Mean B factors (Å ²): protein/water/K ⁺ /glycerol/overall	15.8/24.9/33.1/38.4/16.9

4.2.4 Evaluating Pac17 apo model quality

Once model building and refinement were complete, the quality of the model was determined by analysis of the model with the MOLPROBITY server (<http://molprobity.biochem.duke.edu/>) (Chen et al., 2010). MOLPROBITY analyses a number of steric and geometric parameters to determine the quality of the model and designates it a percentile score based on a comparison of the parameters determined by all current entries in the PDB (Figure 4.11).

Chapter 4 – Structural Investigation of the Pac17 Protein

All-Atom Contacts	Clashscore, all atoms:	5.12	96 th percentile* (N=816, 1.81Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	0.28%	Goal: <1%
	Ramachandran outliers	0.21%	Goal: <0.2%
	Ramachandran favored	98.80%	Goal: >98%
	C β deviations >0.25Å	4	Goal: 0
	MolProbity score [^]	1.27	98 th percentile* (N=11302, 1.81Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

Figure 4.11 MOLPROBITY output for the Pac17 structure.

From the analysis by MOLPROBITY of the finished model of apo-Pac17, it was determined that the quality of the model was among the 96th percentile of protein models deposited in the PDB. MOLPROBITY also identified each of the Met 199 residues in the tetramer as Rhamachandran ‘outliers’.

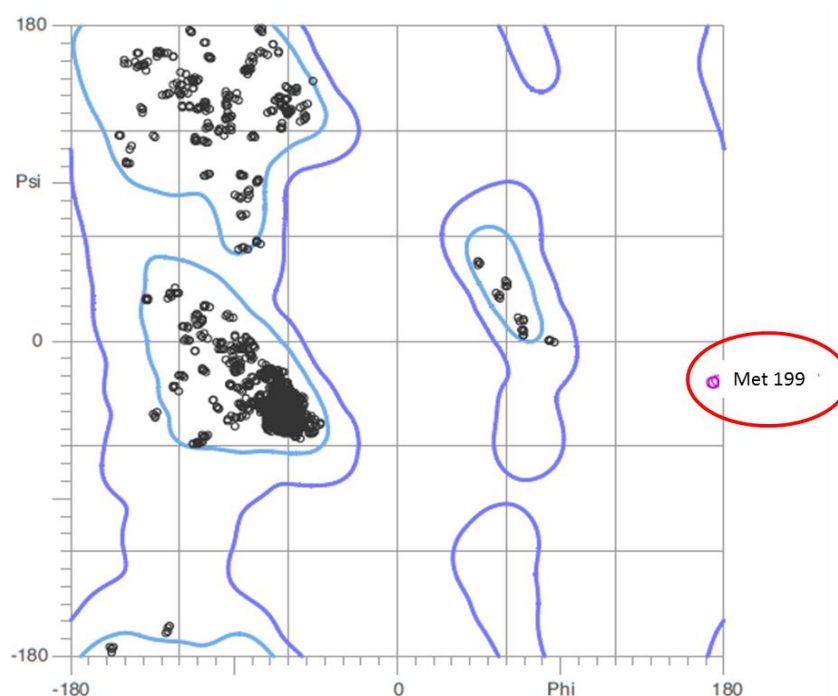


Figure 4.12 The Ramachandran plot produced by MOLPROBITY with the four outliers highlighted, these outliers corresponding to methionine (MET) 199 in each chain A, B, C and D.

Chapter 4 – Structural Investigation of the Pac17 Protein

Further investigation was carried out on these residues using COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 1997). Removing the side chains of these residues in COOT and refining the structure confirmed the electron density is definitely present and not biased by the presence of the amino acid in the model (Figure 4.13). This suggests that the conformation of the methionines in the model are correct and that they are genuine outliers.

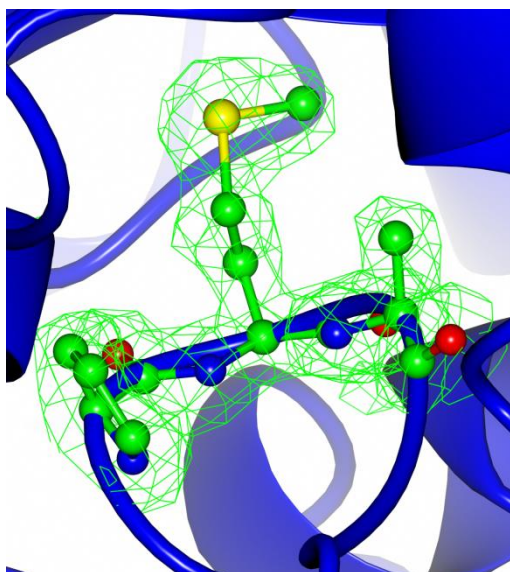


Figure 4.13 Showing $F_{\text{obs}} - F_{\text{calc}}$ density (omit map) for residues Pro 198 – Ala 200, confirming that Met 199 is a genuine Ramachandran outlier; resolution 1.81 Å, contour level 3.0 σ .

Further investigation of the environment in which these Ramachandran outliers exist shows that they are at the centre of a GXMVG motif (X being a proline and alanine, respectively) and at the interface of two of the four subunits of the Pac17 tetramer. The methionine appears to be at the turn of a looped region of Pac17, the glycines allowing there to be more flexibility in the region making the conformation of this methionine more energetically favourable, the proline possibly providing the chain with the ability to ‘kink’.

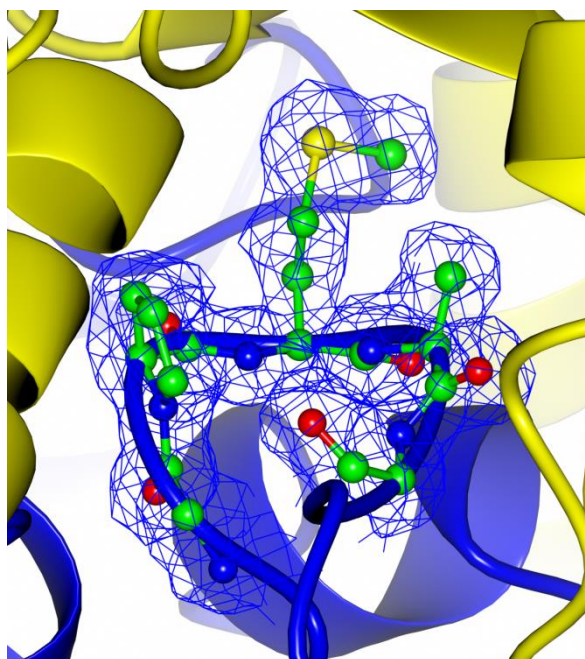


Figure 4.14 showing the environment the methionine outlier is present in, the presence of glycines either side allowing the region to have more flexibility and the presence of a proline giving the region the opportunity to 'kink'. $2F_{obs} - F_{calc}$ map; resolution 1.81 Å, contour level 1.0 σ .

4.2.5 Analysis of the Pac17 apo- structure

Pac17 exists as a tetramer in its crystallographic form. The structure consists of mainly α -helices with a total of 20 long α -helices (5 from each monomer) coming together at the core of the tetramer. A secondary structure analysis of the protein using the online server PDBSUM (<http://www.ebi.ac.uk/pdbsum/>) (Laskowski et al., 1997) (Figure 4.15) deduced that each monomer within the tetramer consists of 22 α -helices between amino acid residues; 11-18, 20-30, 32-48, 54-66, 69-73, 83-94, 108-144, 164-189, 208-214, 224-229, 232-258, 270-272, 286-309, 318-31, 328-348, 353-362, 367-378, 382-399, 409-418, 426-429, 436-441 and 451-490. 8 short β -strands are also present in each monomer, 4 of which form 2 β -hairpins; 150-153, 157-162 and 274-275, 283-284.

Chapter 4 – Structural Investigation of the Pac17 Protein

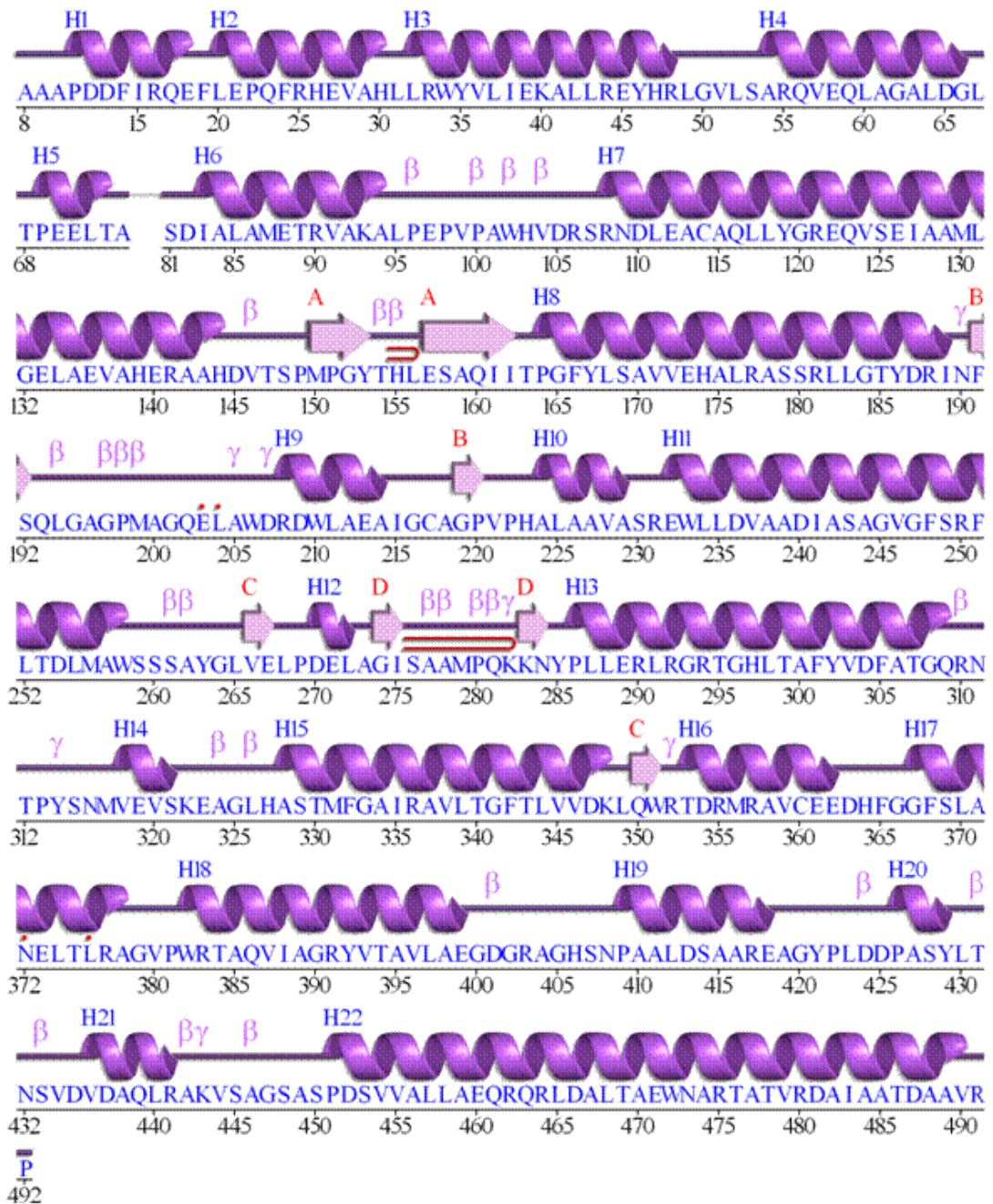


Figure 4.15 PDBSUM output showing a graphical representation of the secondary structure of Pac17.

PDBSUM suggests that between residues 11 and 49 in each of the monomers, there are three helices, however, on closer inspection of the structure in COOT (Emsley and Cowtan, 2004) it is likely to be one long α -helix which is slightly bent (see Figure 4.16), which explains the discrepancy.

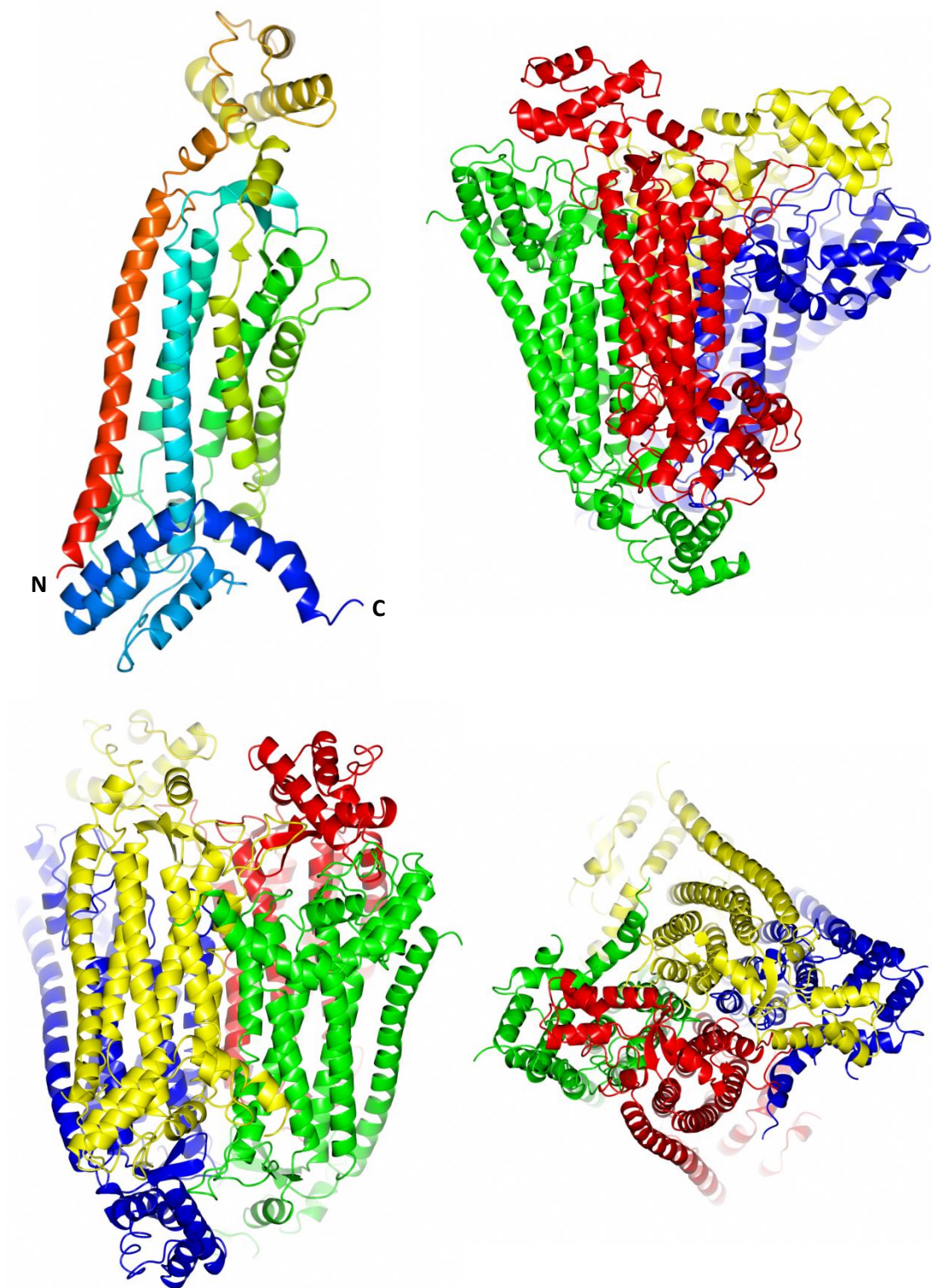


Figure 4.16 Images of Pac17, showing the Pac17 monomer (top left) coloured from the N (red) to C (blue) terminus and three images of the tetramer at different orientations. The bent helix previously discussed can easily be seen in the monomer image in red.

Chapter 4 – Structural Investigation of the Pac17 Protein

The Pac17 model was used as a template in the PDBeFOLD server (www.ebi.ac.uk/pdbe/ssm) (Krissinel and Henrick, 2004) to identify the closest structural homologs. On interrogation of the results, it was found that although model 2E9F was the closest primary structure homolog, it was not present in the PDBeFOLD analysis using the default parameters and only two models (1TJU; duck delta crystallin mutant and 1K62; human argininosuccinate lyase mutant) met the parameters of the search. Superimposition of both the 2E9F and 1TJU models onto the Pac17 model (Figure 4.17) show that the 1TJU model is more closely related to Pac17 in 3D space, however 2E9F is also similar which is likely to be why molecular replacement was successful. Re-submitting the Pac17 model into PDBeFOLD and changing the parameters to accept lower matches resulted in the 2E9F model being present in the search results along with PDB entry 3R6V which has aspartate bound as a ligand. Table 4.4 summarises the comparison of Pac17 to these structural homologs.

Table 4.4 Selected structural homologs of Pac17

Protein	PDB code	Source	PDBeFold output	
			Z-score	Identity (%)
Delta2	1TJU	<i>Anas platyrhynch</i>	11.5	24.0
Argininosuccinate lyase	2E9F	<i>Thermus thermophilus</i>	9.9	27.9
AspB	3R6V	<i>Bacillus sp. YM55-1</i>	8.3	14.2

Protein	Resolution (Å)	R.m.s deviation (Å)/aligned residues		Reference
		Subunit	tetramer	
Delta2	2.1	1.79/409	2.30/1620	(Sampaleanu et al., 2004)
Argininosuccinate lyase	2.8	1.86/401	2.16/1590	Goto et al. (unpublished)
AspB	2.6	2.39/335	2.72/1355	(Fibriansah et al., 2011)

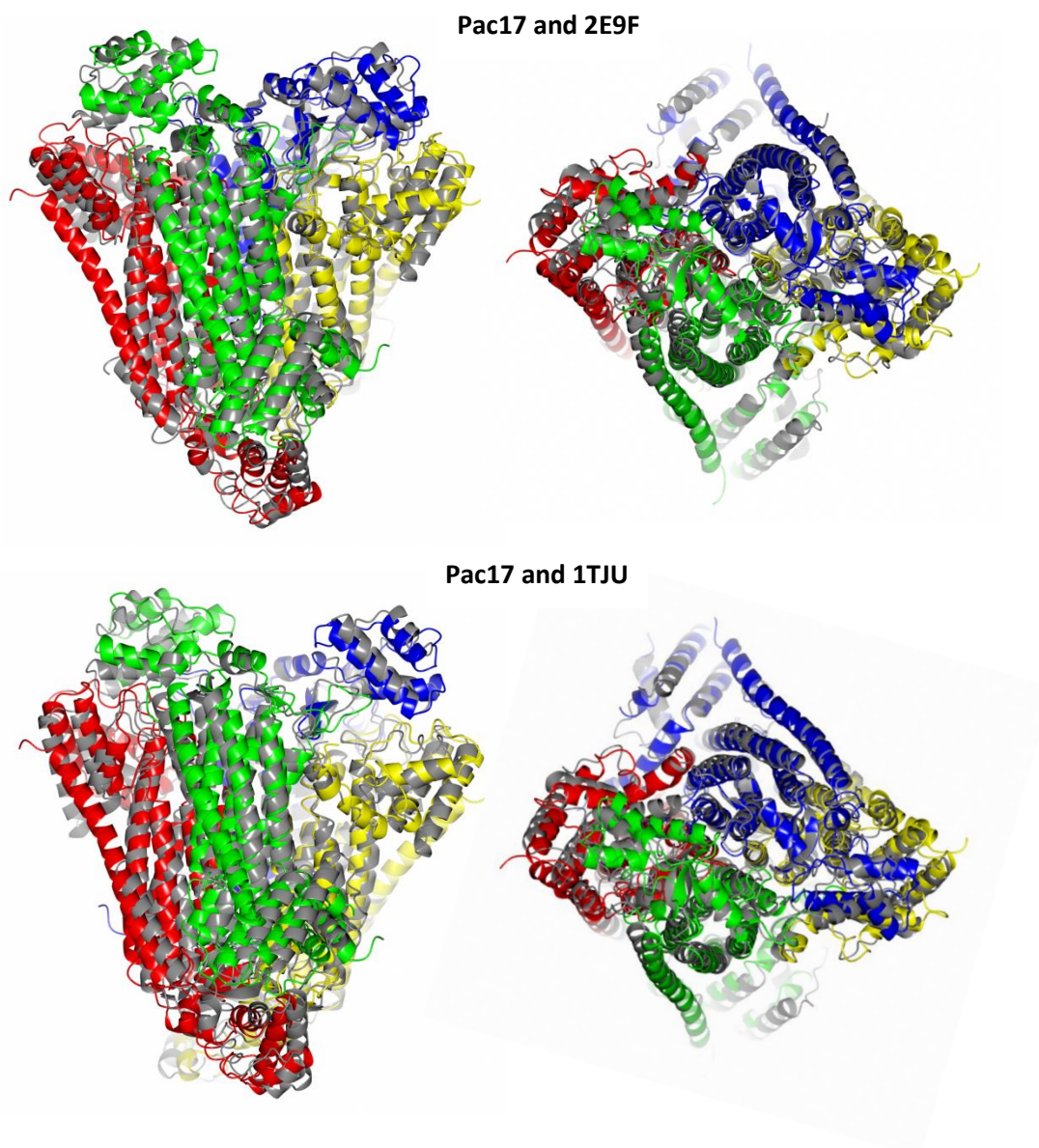


Figure 4.17 Superimpositions of the structure of Pac17 with 2E9F (top); the structure used in molecular replacement and 1TJU (bottom); the highest hit from analysis by PDBeFOLD.

4.2.6 Conservation of residues in the Pac17 structure

Following the determination of the structure of Pac17, an amino acid alignment was performed of Pac17 with proteins with a sequence similarity > 50% using CLUSTAL OMEGA. The alignment was used in CCP4 QTMG to map conservation onto the surface model of Pac17 (Figure 4.18).

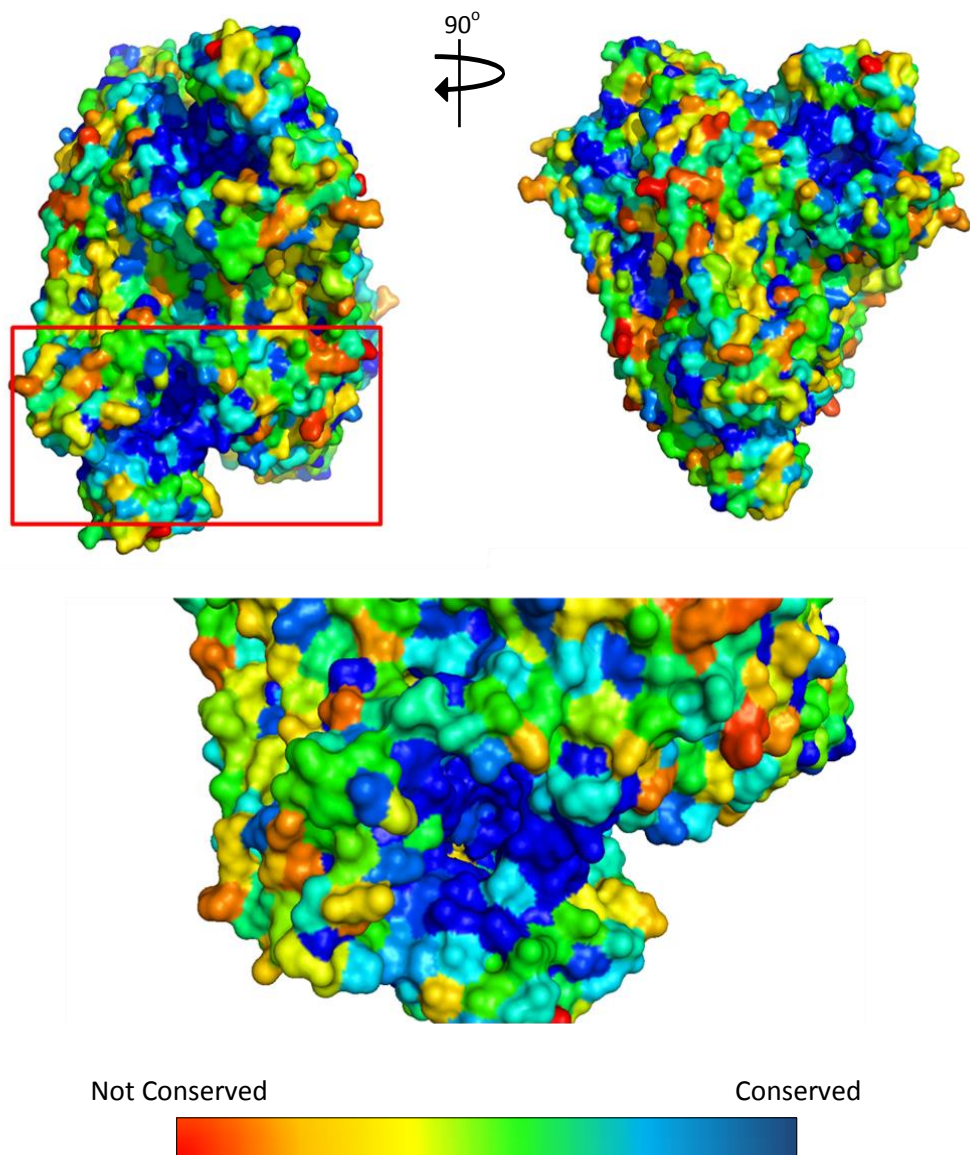


Figure 4.18 mapping the conservation of residues of sequence similar proteins to Pac17 with a region highlighted as having a high level of conservation relative to the rest of the structure. Picture generated using PYMOL.

A large region of conservation can be seen in four positions on the surface of the tetramer. The conserved region was also found to be at the interface of three of the chains of the tetramer which is a common position for active sites to be present. Further evaluation of the conservation between the proteins shows that Pac17 has very little other conservation on the surface of the structure other than at the postulated active site region. Further to this, if the substrate for Pac17 is aspartate (as hypothesised through bioinformatics

analysis) it would suggest that the conserved region around the active site is much larger than one would assume necessary, which may suggest the protein interacts or forms complexes with other proteins or uses a co-substrate.

4.2.7 Crystallisation of Pac17 with its hypothesised substrate

To test the hypothesis that aspartate is the substrate of Pac17, 1 mM of L-aspartate was added to the crystallisation conditions in the hope that the protein would co-crystallise co-ordinated to the ligand. Initial crystal optimisations produced crystals in all co-crystallisation conditions, however, the addition of a cryo-protectant in preparation for flash cooling caused the crystals to breakdown. A new set of screens were set-up which included 10 % (v/v) glycerol in the condition and all conditions produced crystals. The best crystals were found in condition; 10 % (v/v) glycerol, 20 % (w/v) PEG 3350 , 0.2 M potassium sodium tartrate, 0.1 M bis tris propane pH 7.5. Crystals were mounted onto litho-loops, flash-cooled by plunging into liquid nitrogen and stored prior to transport to the synchrotron. For data collection, a single crystal was transferred to the goniostat of beamline I04-1 at the Diamond Light Source (Oxfordshire), maintaining the temperature at -173 °C. A total of 1000 images were collected in a single sweep of the crystal in 0.2° oscillations. The data were merged and scaled using XDS (Kabsch, 2010) and SCALA (Evans, 2006), respectively.

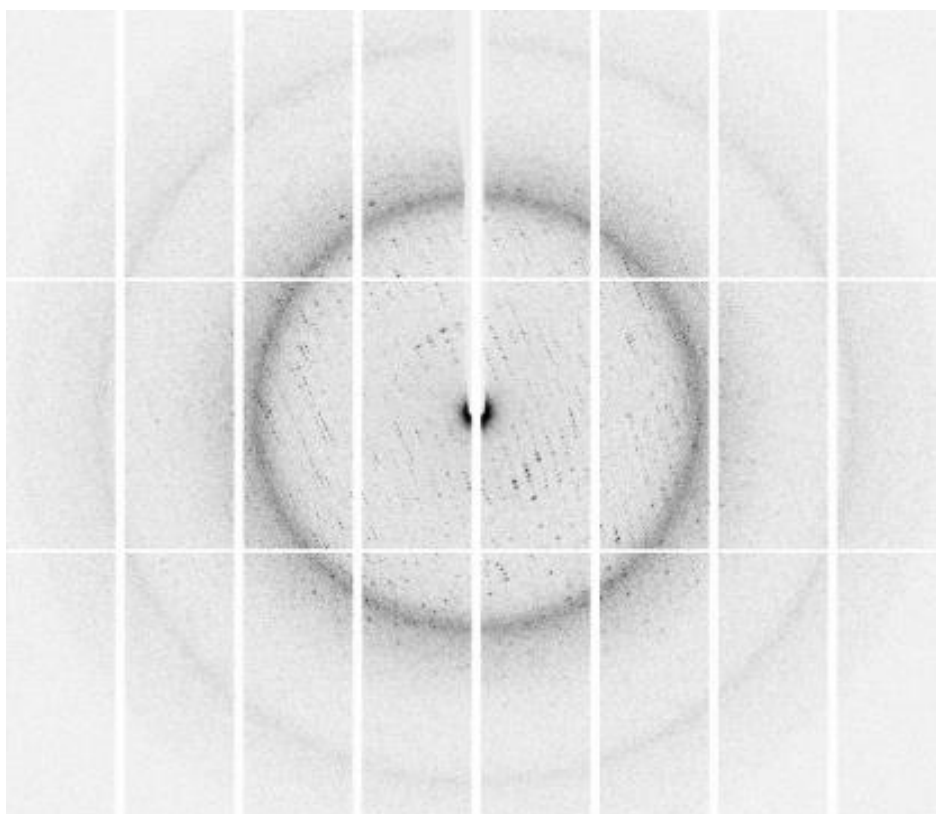


Figure 4.19 Image of the Pac17-Asp crystal in the litho-loop that data were collected on (top) and an example of the diffraction data collected (bottom).

4.2.7.1 Solving the structure of co-crystallised Pac17

The data collected indexed to a different space group ($C222_1$) to that of the apo-Pac17 structure which was confirmed by analysis of the self-rotation function (Figure 4.20) produced in MOLREP (Vagin and Teplyakov, 1997). The solvent content analysis suggested that it was most likely that a dimer was present in the ASU. The structure was solved by molecular replacement in MOLREP (Vagin and Teplyakov, 1997) using a subunit of the

Chapter 4 – Structural Investigation of the Pac17 Protein

apo-Pac17 structure as the search model. As expected, two monomers were located in the ASU.

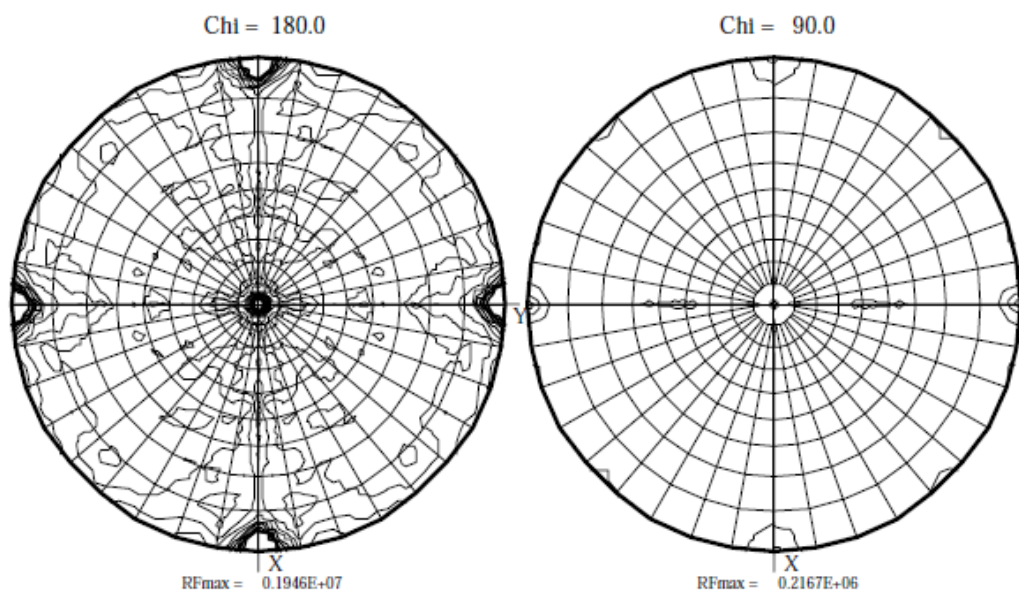


Figure 4.20 Self rotation function generated in MOLREP for Pac17-Asp structure showing Chi = 180.0° and Chi = 90.0. The four peaks aligning with the axes at Chi = 180.0° suggests the space group of the unit cell is C222₁. On the Chi = 180° section, only peaks corresponding to the crystallographic twofold axes are seen. This indicates that the twofold axes of the Pac17 tetramer are aligned with these axes.

The statistics for this dataset were:

Table 4.5 Statistics for X-ray data collection of Pac17 co-crystallised with 1 mM aspartate.

Number of crystals	1
Beamline	I04-1, Diamond Light Source, UK
Wavelength (Å)	0.9173 Å
Detector	Pilatus 2M
Crystal-to-detector distance (mm)	265.8
Rotation range per image (°)	0.2
Exposure time per image (s)	0.2

Chapter 4 – Structural Investigation of the Pac17 Protein

Beam transmission (%)	47
Total rotation range (°)	140.0
Resolution range (Å)	53.00 – 2.30 (2.36 – 2.30)
Space Group	C222 ₁
Cell parameters (Å)	a = 71.08, b = 211.94, c = 141.74
Total no. of measured intensities	216943 (13269)
Unique reflections	47079 (3392)
Multiplicity	4.6 (3.9)
Mean $I/\sigma(I)$	6.8 (2.0)
Completeness (%)	98.3 (97.7)
$R_{\text{merge}}^{\dagger}$	0.148 (0.579)
$R_{\text{meas}}^{\ddagger}$	0.185 (0.745)
Wilson B value (Å ²)	20.20

$^{\dagger} R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. $^{\ddagger} R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Rounds of building and refinement using COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 1997) were used to complete the ligand bound structure. The structure solution and the final structural parameters of the model are summarised in Figure 4.21 and Table 4.6, respectively.

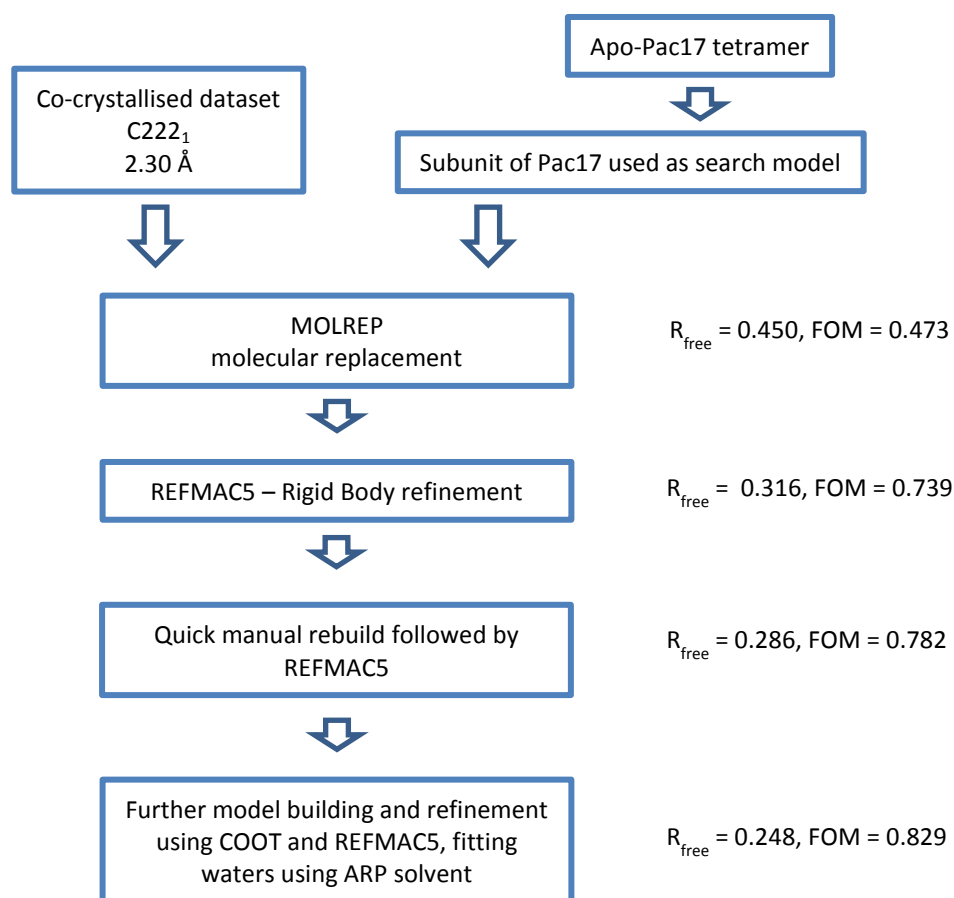


Figure 4.21 Summary of the solving of the ligand bound Pac17 structure.

Table 4.6 Final structural parameters of ligand bound Pac17 structure.

Resolution Range (Å)	53.0-2.30
R_{work}	0.1947
R_{free}	0.2482
Ramachandran plot: favoured/allowed (%)	97.4/99.8
R.M.S bond distance deviation (Å)	0.012
R.M.S bond angle deviation (°)	1.45
No. of residues in protein: Chain A/B	468/483
No. of water molecules/Asp	443/1
Mean B factors: protein/water/Asp/overall	29.6/25.0/34.5/29.4

Chapter 4 – Structural Investigation of the Pac17 Protein

The completed model of ligand bound Pac17 was validated using validation tools in COOT and running the model through the online MOLPROBITY server, the server's analysis summary is shown below.

All-Atom Contacts	Clashscore, all atoms:	6	99 th percentile * (N=355, 2.30Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	2.97%	Goal: <1%
	Ramachandran outliers	0.21%	Goal: <0.2%
	Ramachandran favored	97.35%	Goal: >98%
	Cβ deviations >0.25Å	2	Goal: 0
	MolProbity score [^]	1.81	96 th percentile * (N=8909, 2.30Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

Figure 4.22 MOLPROBITY output for the Pac17-Asp structure. Again the Met 199 of each subunit is a Rhamachndran outlier.

4.2.8 The Aspartate bound structure of Pac17

Interrogation of the resulting structure showed that when crystallographic symmetry was applied to the dimer in the ASU, a tetramer was formed which correlated with that of the apo-Pac17 model. Overlaying the apo and co-crystallised models of Pac17 suggested that there were a number of conformational differences between the structures. Further to this, these regions of conformational change were found around the region of conserved residues of Pac17 (when compared to its closest sequence homologs).

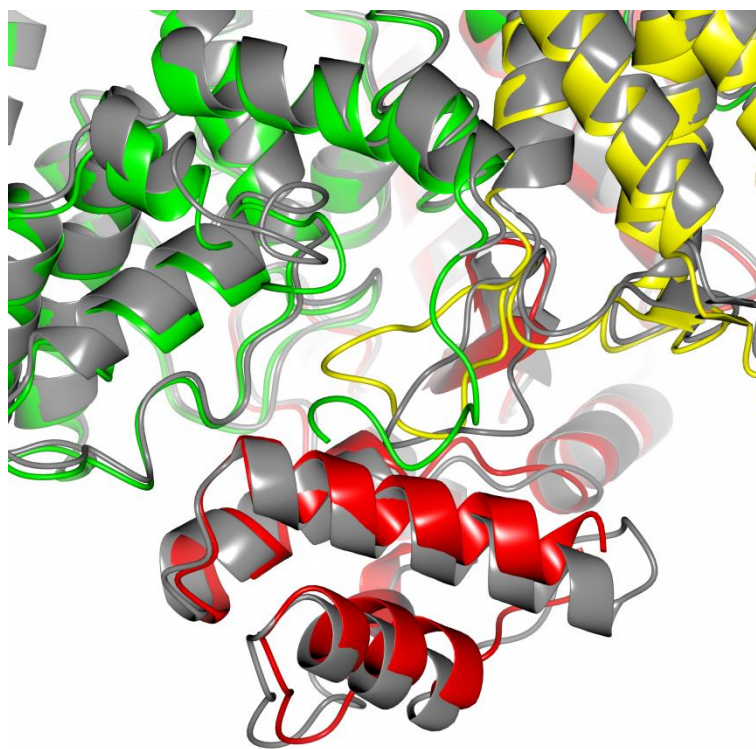


Figure 4.23 Superposition of apo-Pac17 (grey) and Pac17-Asp (coloured) where the two structures differ.

There are two main areas in this region that move; A mobile loop (centre of Figure 4.23 in yellow), this loop moves a significant distance from its apo to ligand bound form, the largest distance moved by an alanine residue (Ala278) which moves a distance of 5.6 Å (comparing position of C α backbone) and an α -helix (bottom of Figure 4.23 in red). These regions are readily recognised if the difference in distance is plotted between the apo and ligand bound structure of Pac17. As there are only two chains in the asymmetric unit of the ligand bound structure, the change in distances were compared to their counterpart in the apo structure of Pac17. It was observed that there is a 'dynamic' chain (chain A) and a 'static' chain (chain B) (Figure 4.24), chain A showing the two conformational changes that have previously been highlighted.

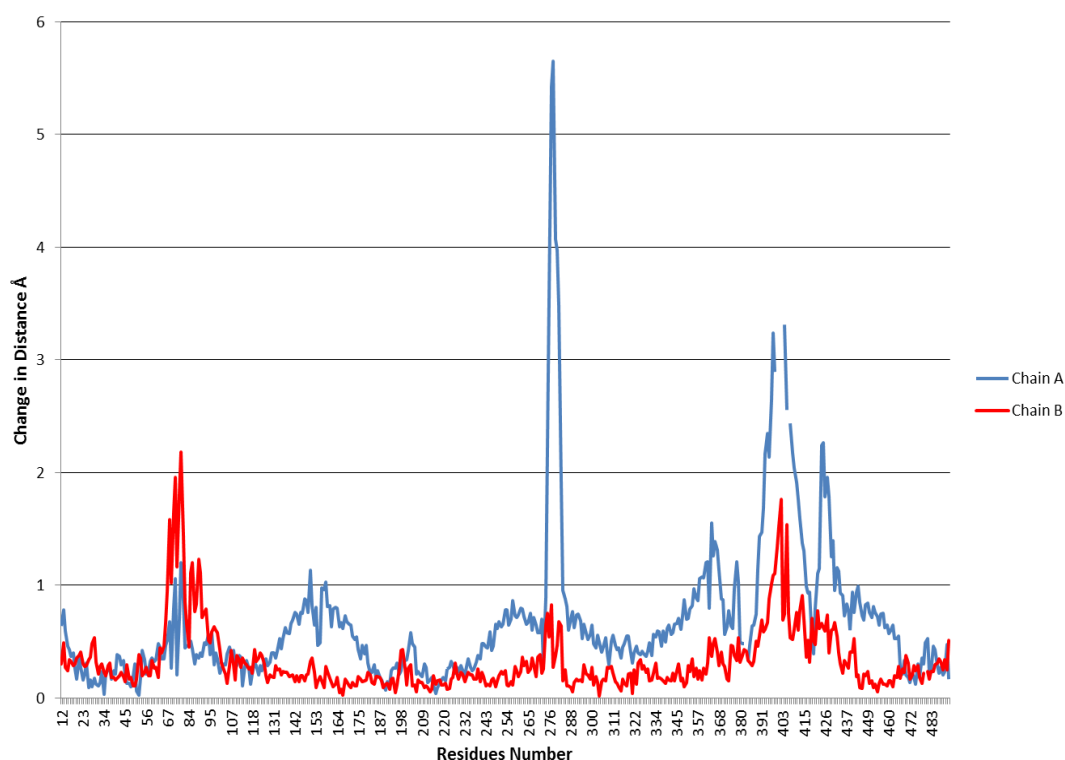


Figure 4.24 Graph showing the difference in distances of chains A and B in the ligand bound model of Pac17 in comparison to their chain counterparts in the apo-Pac17 model. Two regions in chain A can be seen to move significantly; the first approximately between residues 270 and 284 and the second between 390 and 410.

In the centre of this region of conformational change, there is extra density in the ligand bound structure in comparison to that of the apo-Pac17 structure. This density is likely to be that of the ligand; aspartate. Aspartate was fitted into this density to determine whether the ligand was responsible for the extra density (Figure 4.25). Aspartate fits the density well which suggests it is this that is causing the extra density in the Pac17-Asp map.

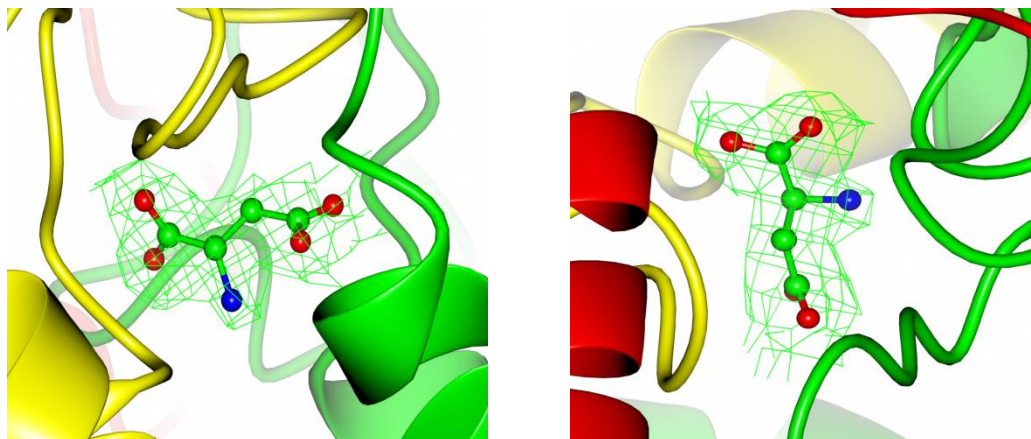


Figure 4.25 Two images showing the ligand (aspartate) modelled into its $F_{\text{obs}} - F_{\text{calc}}$ density map; resolution 2.30 Å, contour level 3.0σ .

Evaluation of this site suggests that the active site of Pac17 is at the interface of three of the four subunits of the tetramer, a common position for an active site. Another noticeable feature of the ligand bound structure is that although in the tetramer there are four active sites, only two of the four active sites are occupied by the ligand, suggesting that the only two active sites of the tetramer are active at any one time.

4.2.9 Further analysis of the ligand bound structure of Pac17

Ligand binding is co-ordinated by seven amino acids from three chains of the tetramer situated around the active site. From chain A; Ser276, Lys282 and Asn284, from chain B; Thr154 and His155, and from chain C; Arg108, Asn109 and Asn316. Ser276, Lys282 and Asp284 are found on the mobile loop that closes over the ligand to allow for the co-ordination between the residues in the loop and the ligand. Thr154 and His155, are found on the α -helix which shows a conformational change between the apo and ligand bound models. The other amino acid residues that are found to co-ordinate with the ligand are found on chain A which does not seem to change conformation from the apo and ligand bound form of Pac17. The co-ordination of these residues with aspartate is depicted in Figure 4.26.

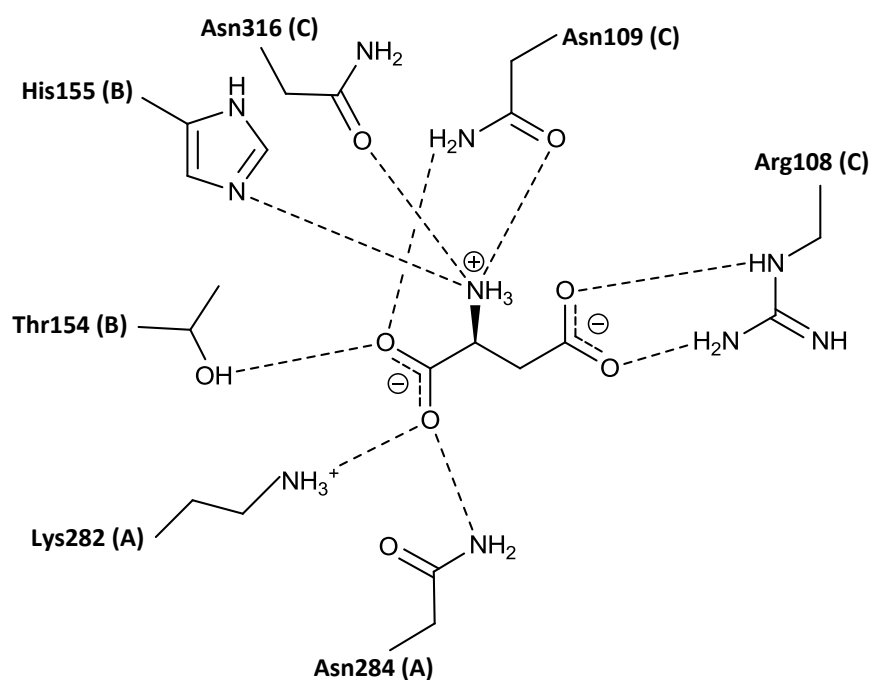
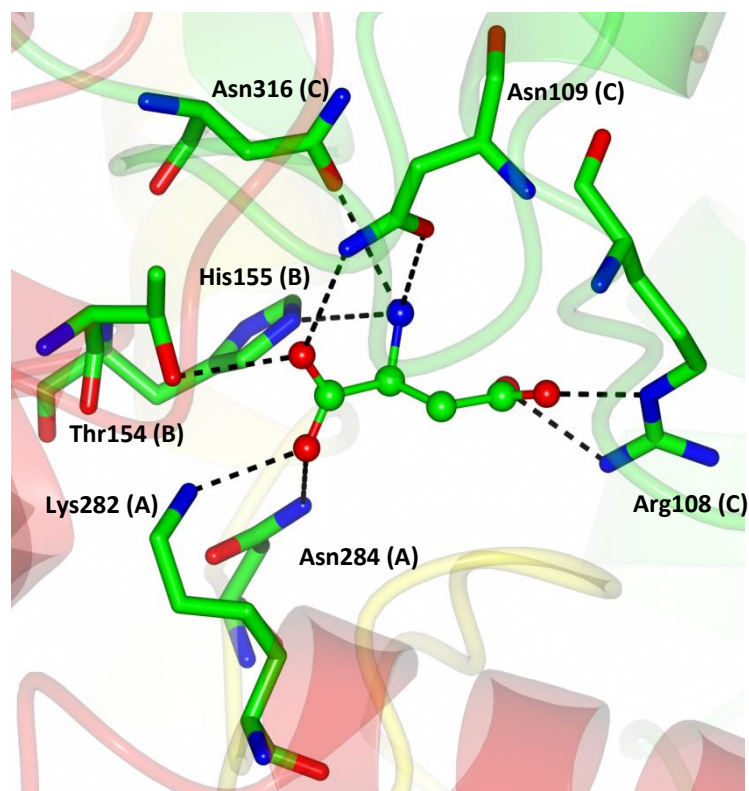


Figure 4.26 The co-ordination of the ligand (aspartate) with key residues in the active site of Pac17 in the context of the model (top) and diagrammatically (bottom).

Interestingly, Asn109 is also co-ordinated to Pro198, which is the amino acid next to the Ramachandran outlier Met199 (Figure 4.27). The fact that a residue that co-ordinates with

the ligand also co-ordinates with a residue next to an outlier is often a reason for the outlier being present.

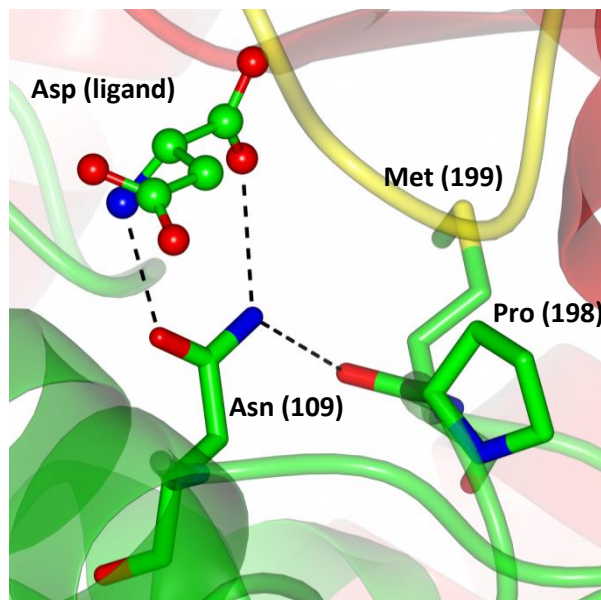


Figure 4.27 The co-ordination of asparagine 109 with the ligand (aspartate) present in the active site and with other residues in the Pac17 active, particularly proline 198, which is the residue next to the Ramachandran outlier methionine 199.

4.2.10 Comparison of Pac17 to AspB

Since aspartate was specifically recognised by Pac17, the hypothesis that the protein is an aspartase is further supported. Referring to the PDBeFOLD analysis shows that an aspartase (AspB from *Bacillus sp.* YM55-1) is present in the search results with PDB accession number 3R6Q. The apo-protein crystals were also soaked with 100 mM aspartate (a different method to that used for Pac17 which was the co-crystallisation of the ligand with the protein); PDB 3R6V (Fibriansah et al., 2011). AspB is reported to be a tetramer, however, it was observed that the 3R6V structure contained one aspartate per tetramer whereas two aspartate ligands are present in each tetramer of ligand bound Pac17. Fibriansah et al. (2011) report that the aspartase contains a characteristic signature sequence of GSSxPxKxN for the Aspartase/Fumarase superfamily members which they designate the SS loop. Further to this they report that this loop region is highly mobile in the apo structure. Interrogation of the Pac17 sequence showed that the 'SS loop' in Pac17 is characterised by the sequence ISAxxPxKxN, I and N being residues 275 and 284, respectively. This sequence differs from the AspB sequence at residues one and three, again this loop being disordered

Chapter 4 – Structural Investigation of the Pac17 Protein

in the apo structure and containing three amino acids that are found to co-ordinate with the ligand. Fibriansah et al. (2011) also report that the second residue (serine) in the sequence is essential for the enzymatic activity, the serine being responsible for the deprotonation of the aspartate to form an enediolate intermediate, which, via an electron rearrangement, causes the release of ammonia from aspartate, the other product being fumarate. In the Pac17 structure, the putative catalytic serine candidate is Ser276. Further analysis of Ser276 shows that the residue is perfectly placed (the C β oxygen of Ser276 being 3.1 Å away from the C β of the aspartate ligand) to be the catalytic residue. The serine is on the highly mobile active site loop and the C α backbone of this residue moves a distance of 3.5 Å between its apo and ligand bound position, the C β oxygen moving even further (4.3 Å). This movement is shown in Figure 4.28.

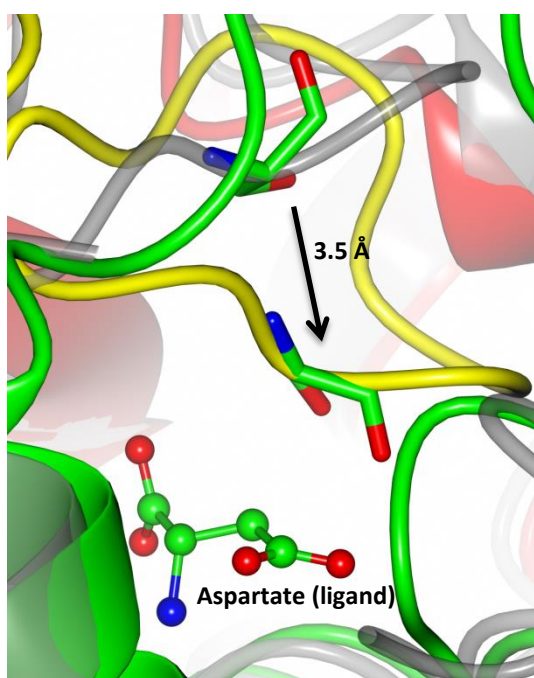


Figure 4.28 Overlay of the apo and Asp bound model of Pac17, showing the position of the ligand in the bound structure and the movement of Ser276 from its apo position (on the grey chain) to its ligand bound position (yellow chain).

Comparison of the amino acids that coordinate with the ligand in Pac17 and AspB are shown in Table 4.7.

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.7 Comparison of the active site residues in Pac17 and AspB.

Pac17 (chain)	AspB (chain)
Serine 276 (A) (catalytic serine)	Serine 318 (C)
Lysine 282 (A)	Lysine 324 (C)
Asparagine 284 (A)	
Threonine 154 (B)	Threonine 187 (A)
Histidine 155 (B)	Histidine 188 (A)
Asparagine 316 (C)	
Asparagine 109 (C)	Asparagine 142 (B)
Arginine 108 (C)	Serine 319 (C), Threonine 101 (B), Threonine (B) and Serine 140 (B)

AspB does not appear to have a residue in the active site that shows similar co-ordination as residues Asn284 and Asn316 in Pac17. Further to this there are differences in the coordination of active site residues with the carboxy group of the amino acid ligand. The ligand bound Pac17 structure shows that Arg108 coordinates with each of the oxygens in the C β carboxylic acid, whereas AspB appears to coordinate with these atoms of the ligand with a series of threonine and serine residues. The co-ordination of arginine with this group would seem to be a better candidate for the subsequent reaction, allowing for better stabilisation of the enediolate intermediate that is postulated to be part of the enzymatic mechanism according to Fibriansah et al. (2011). The mechanism suggested by Fibriansah et al. (2011) is depicted in Figure 4.29 in the context of the Pac17 active site.

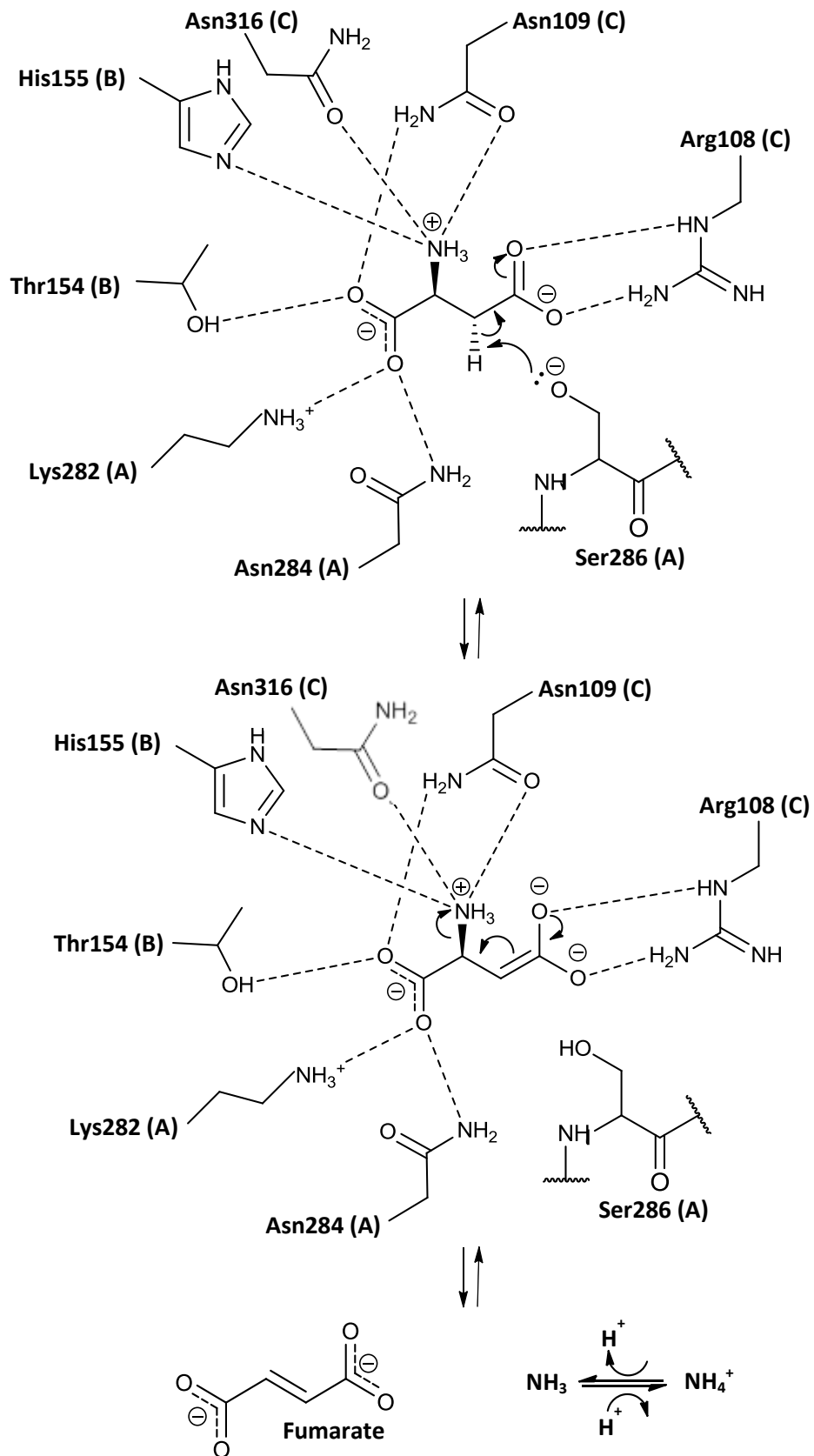


Figure 4.29 Postulated enzymatic mechanism for Pac17 (based on Fibriansah et al., 2011).

4.2.11 Further interrogation of Pac17 substrate

As will be described in chapter 6, *in vitro* enzymatic studies of Pac17 failed to detect enzyme activity. It was decided that further substrate interrogation of Pac17 may be possible using protein crystallography. As a conformational change was seen when Pac17 was co-crystallised with aspartate, and with the bioinformatics evidence that the protein was likely to be an argininosuccinate lyase or aspartate lyase, further co-crystallisations were set up using structurally similar amino acids to aspartate (asparagine, glutamine and glutamate) and argininosuccinate. The conditions used were identical to those for the co-crystallisation of Pac17 with aspartate using the same concentration of ligand (1 mM). Crystals formed in the asparagine (data shown in the apo-Pac17 structure earlier in this chapter), glutamate and glutamine screens, however, co-crystallisation of Pac17 with argininosuccinate failed to produce any crystals. Crystals were mounted onto litho-loops, flash cooled by plunging into liquid nitrogen and stored prior to transport to the synchrotron. For data collection, a single crystal was transferred to the goniostat of station I04-1 at the Diamond Light Source (Oxfordshire), maintaining the temperature at -173 °C. The data collection statistics are reported in Table 4.8.

Table 4.8 Statistics for X-ray data collection of Pac17 co-crystallised with L-glutamate and L-glutamine.

Co-crystallised with	1 mM L-glutamate	1 mM L-glutamine
Number of crystals	1	1
Beamline	I04-1	I04-1
Wavelength (Å)	0.9173	0.9173
Detector	Pilatus 2M	Pilatus 2M
Crystal-to-detector distance (mm)	340.7	297.1
Rotation range per image (°)	0.2	0.2
Exposure time per image (s)	0.4	0.3
Beam transmission (%)	100.0	100.0
Total rotation range (°)	240.0	200.0

Chapter 4 – Structural Investigation of the Pac17 Protein

Resolution range (Å)	60.70-2.32 (2.38-2.32)	53.67-2.00 (2.05-2.00)
Space Group	C2	C2
Cell parameters (Å/°)	a = 214.8, b = 71.2, c = 142.4, β = 92.86	a = 214.96, b = 71.32, c = 142.41, β = 92.87
Total no. of measured intensities	417400 (29065)	530030 (30015)
Unique reflections	92148 (6791)	141427 (9388)
Multiplicity	4.5 (4.3)	3.7 (3.2)
Mean $I/\sigma(I)$	8.4 (2.7)	12.1 (3.4)
Completeness (%)	96.9 (95.9)	93.8 (82.3)
$R_{\text{merge}}^{\dagger}$	0.122 (0.500)	0.069 (0.290)
$R_{\text{meas}}^{\ddagger}$	0.156 (0.639)	0.093 (0.392)
Wilson B value (Å ²)	20.7	15.5

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. [‡] $R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Inspection of the refined datasets of the co-crystallisation of Pac17 and L-glutamate and L-glutamine showed that Pac17 was present in its apo conformation (the models were isomorphous), and the maps produced showed no evidence for bound ligands. These datasets were therefore not considered any further. The absence of these structurally similar amino acids to aspartate in the co-crystallisation datasets further supports that aspartate is likely to be the natural substrate for Pac17.

As a further study, the co-crystallisation of Pac17 with its likely enzymatic product, fumarate was also attempted. The conditions used for crystallisation were the same as that for the other co-crystallisation attempts. Although crystals were obtained in this optimisation, no useful data was obtained at the synchrotron.

Chapter 4 – Structural Investigation of the Pac17 Protein

4.2.12 Site directed mutagenesis (SDM)

From the structural determination of the active site of Pac17, a number of active site mutants were produced to determine the importance of the presence of these residues. Primers were designed to mutate residues in the active site to alanine residues (method described in chapter 2). The table below shows the residues intended to be mutated and the primers used in the SDM method to produce the alanine mutants.

Table 4.9 Showing primers used to produced Pac17 active site mutants.

Mutant	Forward Primer	Reverse Primer
Arginine 108 to Alanine	Pac17_108Arg/Ala_F	Pac17_108Arg/Ala_R
Asparagine 109 to Alanine	Pac17_109Asn/Ala_F	Pac17_109Asn/Ala_R
Serine 276 to Alanine	Pac17_276Ser/Ala_F	Pac17_276Ser/Ala_R
Methionine 279 to Alanine	Pac17_279Met/Ala_F	Pac17_279Met/Ala_R
Lysine 282 to Alanine	Pac17_282Lys/Ala_F	Pac17_282Lys/Ala_R
Asparagine 284 to Alanine	Pac17_284Asn/Ala_F	Pac17_284Asn/Ala_R

Mutagenesis of four of the residues were successful (Asn109, Met279, Lys282 and Asn284), determined by DNA sequencing, but mutagenesis was unsuccessful on the hypothesised catalytic serine (276) and arginine (108). The mutants were expressed and purified then used in crystal trials, using the same conditions as used for the ligand bound structure (including 1 mM L-aspartate). The addition of the aspartate was to determine whether Pac17 could still bind the ligand when one of the amino acid mutations were present. Crystals formed in all of the mutant trials. The crystals were mounted onto litho-loops, flash cooled by plunging into liquid nitrogen and stored prior to transport to the synchrotron. For data collection, a single crystal was transferred to the goniostat of either station I04 or I04-1 at the Diamond Light Source (Oxfordshire), maintaining the temperature at -173 °C. The data collection statistics and the final model parameters for each of the datasets are reported in the subsequent tables.

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.10 Statistics for X-ray data collection of Pac17 mutants Asn109Ala and Met279Ala.

Mutant	Asn109Ala	Met279Ala
Number of crystals	1	1
Beamline	I04	I04-1
Wavelength (Å)	0.9795	0.9173
Detector	ADSC Q315r	Pilatus 2M
Crystal-to-detector distance (mm)	323.3	369.5
Rotation range per image (°)	0.2	0.4
Exposure time per image (s)	0.5	0.4
Beam transmission (%)	49.97	100.0
Total rotation range (°)	180.0	240.0
Resolution range (Å)	106-2.00 (2.05-2.00)	44.17-2.30 (2.36-2.30)
Space Group	C2	C2
Cell parameters (Å/°)	a = 214.0, b = 71.3, c = 142.0, β = 92.94	a = 214.50, b = 71.56, c = 143.34, β = 92.95
Total no. of measured intensities	490079 (19734)	373468 (14878)
Unique reflections	137089 (7254)	88533 (4685)
Multiplicity	3.6 (2.7)	4.2 (3.2)
Mean $I/\sigma(I)$	8.6 (2.4)	8.9 (5.7)
Completeness (%)	94.9 (68.8)	90.9 (65.5)
$R_{\text{merge}}^{\dagger}$	0.124 (0.590)	0.099 (0.114)
$R_{\text{meas}}^{\ddagger}$	0.175 (0.832)	0.127 (0.158)
Wilson B value (Å ²)	11.7	9.2

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. [‡] $R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Chapter 4 – Structural Investigation of the Pac17 Protein

Table 4.11 Final structural parameters of Pac17 mutants Asn109Ala and Met279Ala.

Mutant	Asn109Ala	Met279Ala
Resolution Range (Å)	106.0 – 2.0	44.17 – 2.30
R _{work}	0.1899	0.1662
R _{free}	0.2466	0.2260
Ramachandran plot: favoured/allowed (%)	98.1/99.8	97.9/99.8
R.M.S bond distance deviation (Å)	0.020	0.017
R.M.S bond angle deviation (°)	1.811	1.69
No. of residues in protein: Chain A/B/C/D	478/477/475/476	478/478/482/481
No. of waters/K ⁺ /glycerol	1953/4/3	1939/4/3
Mean B factors: protein/water /overall	15.4/24.0/17.3	10.7/20.9/12.2

Table 4.12 Statistics for X-ray data collection of Pac17 mutants Lys282Ala and Asn284Ala.

Mutant	Lys282Ala	Asn284Ala
Number of crystals	1	1
Beamline	I04-1	I04-1
Wavelength (Å)	0.9173	0.9173
Detector	Pilatus 2M	Pilatus 2M
Crystal-to-detector distance (mm)	383.8	369.5
Rotation range per image (°)	0.2	0.2
Exposure time per image (s)	0.5	0.3
Beam transmission (%)	100.0	100.0
Total rotation range (°)	200.0	200.0
Resolution range (Å)	51.06-2.50 (2.56-2.50)	60.45 – 2.80 (2.87-2.80)
Space Group	C2	C2

Chapter 4 – Structural Investigation of the Pac17 Protein

Cell parameters (Å/°)	a = 214.82, b = 71.39, c = 142.56, β = 92.93	a = 213.7, b = 71.3, c = 141.9, β = 92.86
Total no. of measured intensities	264696 (14296)	151951 (8192)
Unique reflections	72567 (4532)	48638 (3090)
Multiplicity	3.6 (3.2)	3.1 (2.7)
Mean $I/\sigma(I)$	7.0 (2.7)	7.3 (3.0)
Completeness (%)	90.8 (78.2)	91.9 (80.3)
$R_{\text{merge}}^{\dagger}$	0.123 (0.387)	0.131 (0.357)
$R_{\text{meas}}^{\ddagger}$	0.166 (0.514)	0.176 (0.480)
Wilson B value (Å ²)	20.3	18.5

[†] $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. [‡] $R_{\text{meas}} = \frac{\sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

Table 4.13 Final structural parameters of Pac17 mutants Lys282Ala and Asn284Ala.

Mutant	Lys282Ala	Asn284Ala
Resolution Range (Å)	51.06 – 2.50	60.45 – 2.80
R_{work}	0.1837	0.2115
R_{free}	0.2680	0.3250
Ramachandran plot: favoured/allowed	96.7/99.6	94.9/99.5
R.M.S bond distance deviation (Å)	0.013	0.013
R.M.S bond angle deviation (°)	1.48	1.48
No. of residues in protein: Chain A/B/C/D	479/479/482/481	482/481/480/479
No. of waters/K ⁺ /glycerol	1938/3/4	1930/2/3
Mean B factors: protein/water/overall	13.6/28.979/15.4	25.7/19.1/25.6

Chapter 4 – Structural Investigation of the Pac17 Protein

The residue mutations were confirmed through a combination of inspecting $F_{\text{obs(wild type)}} - F_{\text{obs(mutant)}}$ difference maps phased on the wild type model, and refining the structure with the introduced mutation against the mutant dataset.

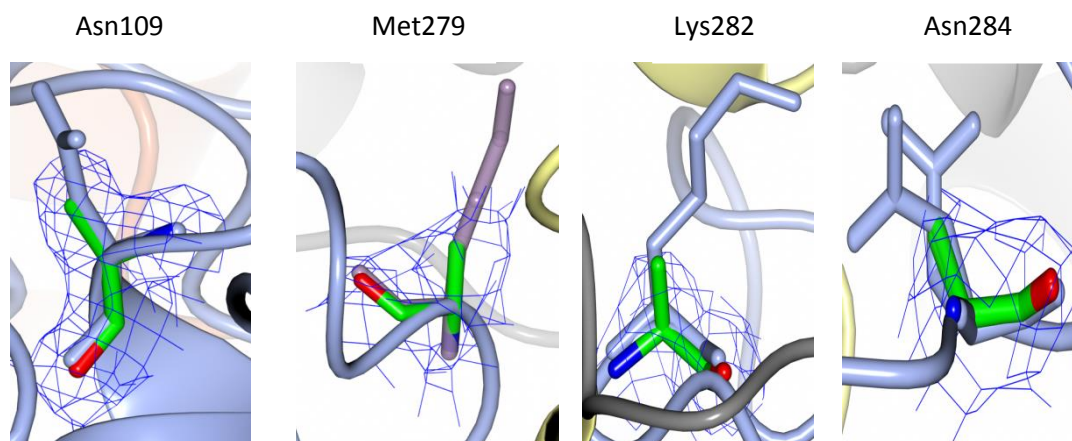


Figure 4.30 Showing $2F_{\text{obs}} - F_{\text{calc}}$ maps of the mutated residues; the full residue seen in blue showing that density for the non-mutated residue is absent.

Analysis of the four Pac17 mutants suggest that all are in the apo conformation, suggesting that these four residues are essential for the successful co-ordination of aspartate with the protein. Interestingly, Met279 does not directly co-ordinate with the ligand, however hydrogen bonding is present between this residue and Lys282. The later does co-ordinate with the ligand and as such may be being stabilised by the presence of Met279.

4.3 Conclusions

The structure of the protein Pac17, a lyase encoded for by the *pac17* gene in the gene cluster for pacidamycin from *Streptomyces coeruleorubidus* has been determined. An L-aspartate bound structure of this protein has also been produced, which has enabled further analysis of the protein. Co-crystallisation of Pac17 under the same conditions used for the crystallisation of the protein with the ligand, with structurally similar ligands showed no change to the apo structure. Furthermore, site-directed mutagenesis of residues (when co-crystallised with L-aspartate) within the postulated 'active' site also resulted in the protein taking the apo conformation rather than the ligand bound, suggesting that these residues are essential for the recognition and/or the co-ordination of

Chapter 4 – Structural Investigation of the Pac17 Protein

the ligand with the protein. The comparison of Pac17 to AspB (Fibriansah et al., 2011) showed a number of similarities and differences, particularly in the stoichiometry of number of ligands per protein tetramer, and also the residues that co-ordinate with the ligand in the active site. This comparison however has shed some light on the likely mode of action for the protein, highlighting a serine residue that is likely to be the catalytic residue. A mechanism for the enzymatic reaction has been suggested (adapted from Fibriansah et al (2010)); the catalytic serine causing the deprotonation of the aspartate, forming an enediolate intermediate, which, via a rearrangement of electrons, causes the release of ammonia from the intermediate to form fumarate, the by-product of the reaction. A number of questions do arise over this suggested mechanism such as how the deprotonation of the serine would occur due to serine's high pKa (approximately 9.5) and whether the deprotonation of the aspartate ligand is via a 1,2-elimination (Mohrig, J., 2012)

Site direct mutagenesis attempts on the 'likely' catalytic serine (Ser276) and an active site arginine (Arg108) were unsuccessful in this study. This would represent a good starting place for continuing the study of the protein, as well as more attempts to obtain a ligand bound structure of the protein with the enzymatic reaction product, fumarate. Additionally, *in vitro* activity studies of this protein to obtain kinetic data for the native protein and the mutants produced would also complement the study. This *in vitro* study has been attempted and will be discussed further in chapter 6 of this thesis.

Pacidamycin Cluster Gene Disruption Studies

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

5.1 Introduction

In vivo analysis of genes can offer fundamental insights into the biosynthesis of metabolites and natural products.

PCR mediated gene disruption (Gust et al., 2003) was considered a suitable approach for the disruption of candidate genes within the pacidamycin biosynthetic gene cluster of *Streptomyces coeruleorubidus*. This method allows for the replacement of a gene on a bacterial chromosome with a selectable marker and the accurate replacement of genes *in vivo*, removing the need to carry out laborious *in vitro* methods.

A number of genes within the pacidamycin biosynthetic gene cluster were of interest in this study; the four genes hypothesised to be involved in the biosynthesis of DABA (*pac17* through *20*), the gene cluster's four hypothetical proteins (*pac1*, *2*, *7* and *13*) and a N-methyltransferase (*pac22*) responsible for the methylation of the DABA residue. Previous studies into pacidamycin biosynthesis had determined that disruption of the 5' half of *pac19* eliminated the production of pacidamycin in the native producer *S. coeruleorubidus* (Rackham et al., 2010). The study by Rackham et al. (2010) also showed that the chemical complementation of this disruption with the diamino acid reinstated pacidamycin production. It has previously been stated that a function for *pac2* has been elucidated (Zhang et al., 2011), however, prior to publication of this paper, a disruption of this gene was attempted (along with the 5' and 3' genes of the cluster; *pac1* and *22*, respectively) but was unsuccessful.

In this chapter, the disruption of the DABA biosynthetic genes and two hypothetical protein genes (*pac7* and *pac13*) will be reported. Disruption in the native producer (*S. coeruleorubidus*) and in the cloned gene cluster expressed heterologously (in *S. lividans* TK24) will be discussed. A number of considerations will also be raised including, the translational coupling of *pac17-19* and the implications this has on the disruption study, the polar effects of genes in close proximity, and the essentiality of the genes being studied to pacidamycin biosynthesis.

5.2 Results

5.2.1 Gene disruption studies in the pacidamycin native producer *S. coeruleorubidus*

To ensure that the methods for growth, metabolite extraction and detection reported by Rackham et al. (2010) were suitable for this study, an initial test using wild type *S. coeruleorubidus* was carried out as described in chapter 2. After metabolite extraction, both LC-MS and LC-MS/MS were carried out. LC-MS analysis consisted of monitoring the detection of five of the most abundant pacidamycins produced by *S. coeruleorubidus* and included; pacidamycin 1 ($m/z = 875$), pacidamycin 4 ($m/z = 804$), pacidamycin 4N ($m/z = 816$), pacidamycin 5T ($m/z = 781$) and pacidamycin D ($m/z = 712$). LC-MS/MS analysis consisted of isolation of ions with $m/z = 712$ in MS_1 (pacidamycin D parent ion) and analysis of the parent ion fragmentation by MS_2 .

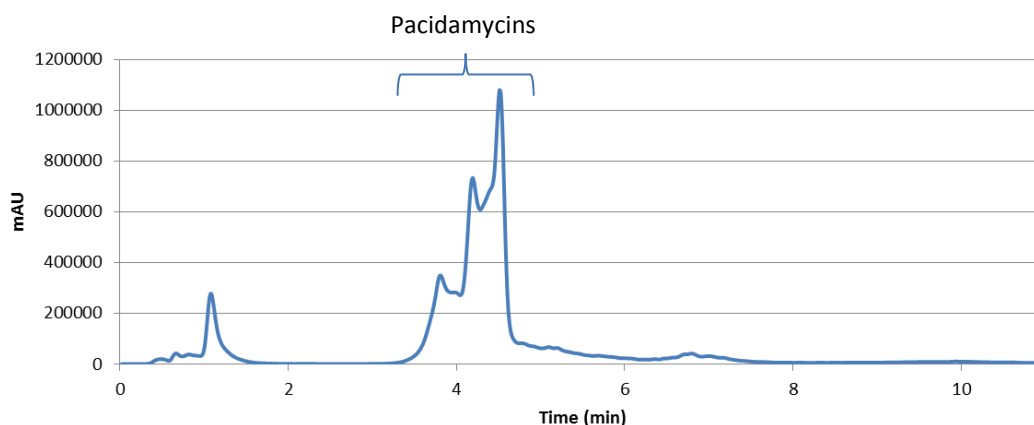


Figure 5.1 The LC profile produced by extraction of metabolites from wild type *S. coeruleorubidus*. The profile shows the UV absorbance produced by the ion masses of pacidamycins 1, 4, 4N, 5T and D. The pacidamycins have an elution time of approximately 3.7 – 4.7 min using the method described in chapter 2.

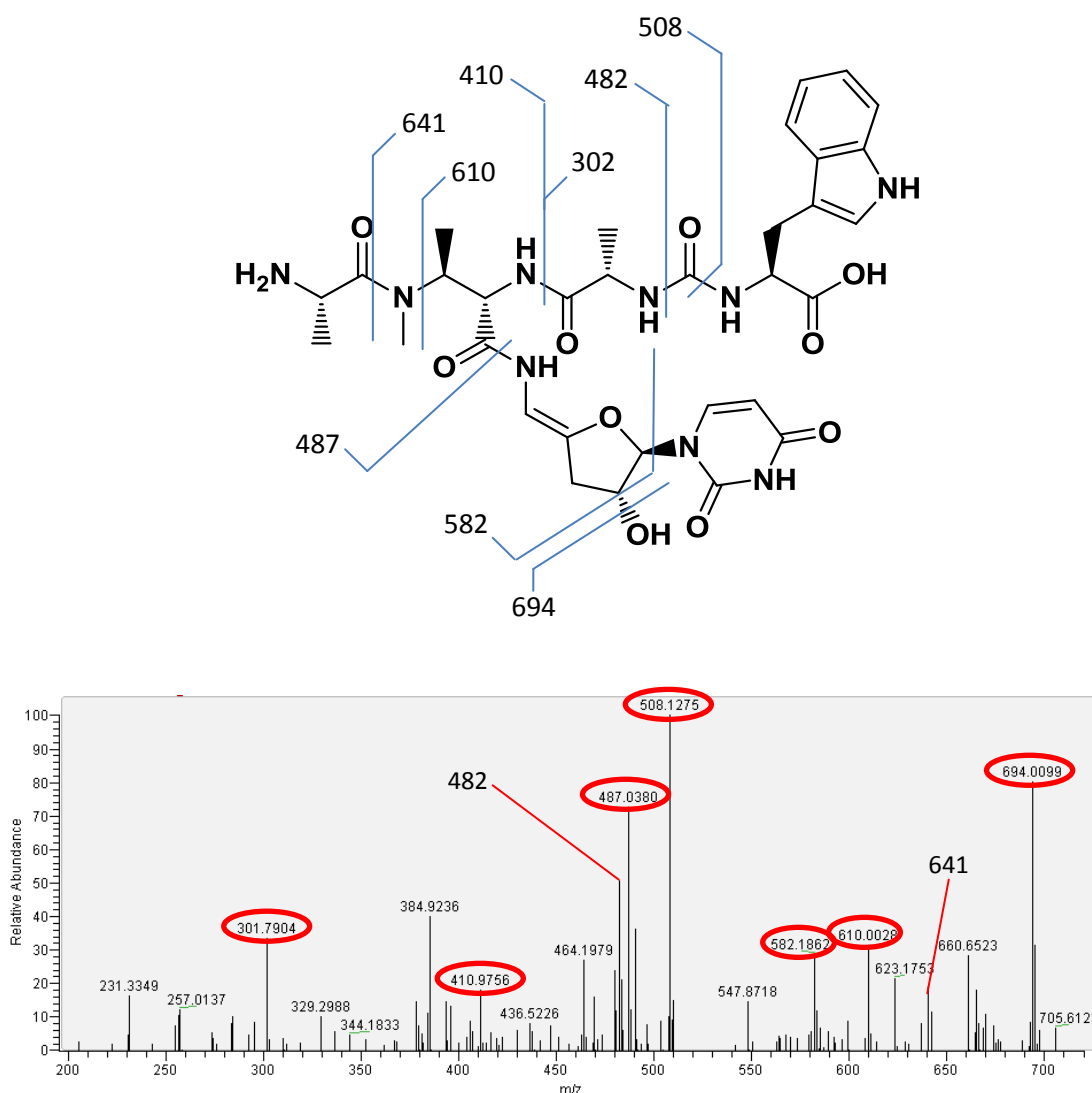


Figure 5.2 The fragments predicted to be observed by MS₂ analysis of pacidamycin D (top) and those observed upon MS₂ analysis, the major peaks in the experimental profile (highlighted in red) match those predicted to be present.

From analysis of the metabolite profile of the wild type of *S. coeruleorubidus*, the methods outlined by Rackham et al. (2010) were found to be robust and suitable for use in this gene disruption study.

5.2.1.1 Gene disruption studies of hypothetical protein genes

As discussed in the introduction of this chapter, the disruption of all four postulated hypothetical genes was initially attempted, however, although a number of attempts to disrupt *pac1* and *pac2* were undertaken, each recombination attempt of the gene of interest with that of the selectable marker failed. The study of these two genes was

therefore not taken further. As discussed in chapter 3, the function of *pac2* has subsequently been identified (Zhang et al., 2011) and the function of a *pac1* homolog in the sansamycin gene cluster also reported (Li et al., 2013).

5.2.1.2 Disruption of *pac7* and *pac13* in *S. coeruleorubidus*

Primers were designed as per the procedure outlined in chapter 2 for replacement of the entire *pac7* and *pac13* (no part of the gene would remain in the *S. coeruleorubidus* genome). Primers Redirect_Pac7_F and Redirect_Pac7_R and Redirect_Pac13_F and Redirect_Pac13_R were used for amplification of the cassette for the disruption of *pac7* and *pac13*, respectively. The cassettes were introduced into *E. coli* strain BW25113. Recombination of the cassettes in place of the gene of interest on the 2H-5 cosmid was selected for by growth of the *E. coli* strains on selective medium and PCR analysis. The disrupted cosmids were isolated from the *E. coli* strains by alkaline lysis and the cosmid introduced into the non-methylating *E. coli* ET12567/pUZ8002. *E. coli* clones containing the mutagenised cosmids were selected for on selective medium and used for conjugal transfer of the mutant cosmid into *S. coeruleorubidus*. Exconjugates were restreaked onto selective medium to ensure isolation of single exconjugates, and spore stocks were produced. The resulting *S. coeruleorubidus* strains were designated strain ID's of DT-001 and DT-002 for the *pac7* and *pac13* disruptions, respectively.

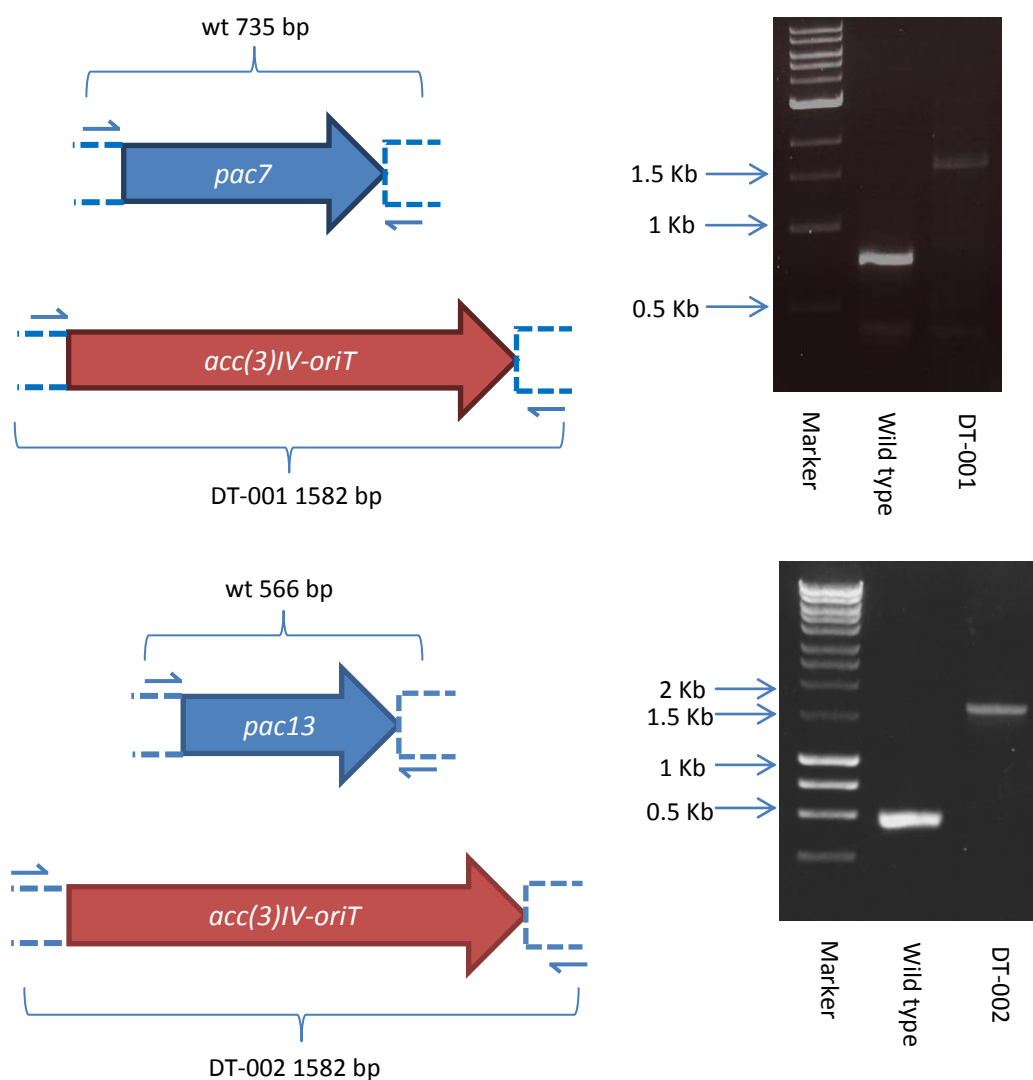


Figure 5.3 The predicted size of the PCR products produced by the amplification of the chromosomal regions of the *pac7* and *pac13* genes in the wild type and disrupted strains along with the PCR analysis.

5.2.1.3 Liquid growth and LCMS analysis of the extracted metabolites of strains *S. coeruleorubidus* DT-001 and DT-002

The *S. coeruleorubidus* DT-001 and DT-002 strains were cultured in ISP₂ liquid medium as per the procedure outlined in chapter 2. Once cultured, the cell free extracts were used in a metabolite extraction method that is outlined in chapter 2 which allowed for the extraction of the pacidamycins for subsequent analysis. After metabolite extraction of the cultures, LC-MS analysis was carried out to confirm the presence or absence of pacidamycin

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

production in the different *S. coeruleorubidus* strains. The LC-MS parameters were identical to those described in the method in chapter 2.

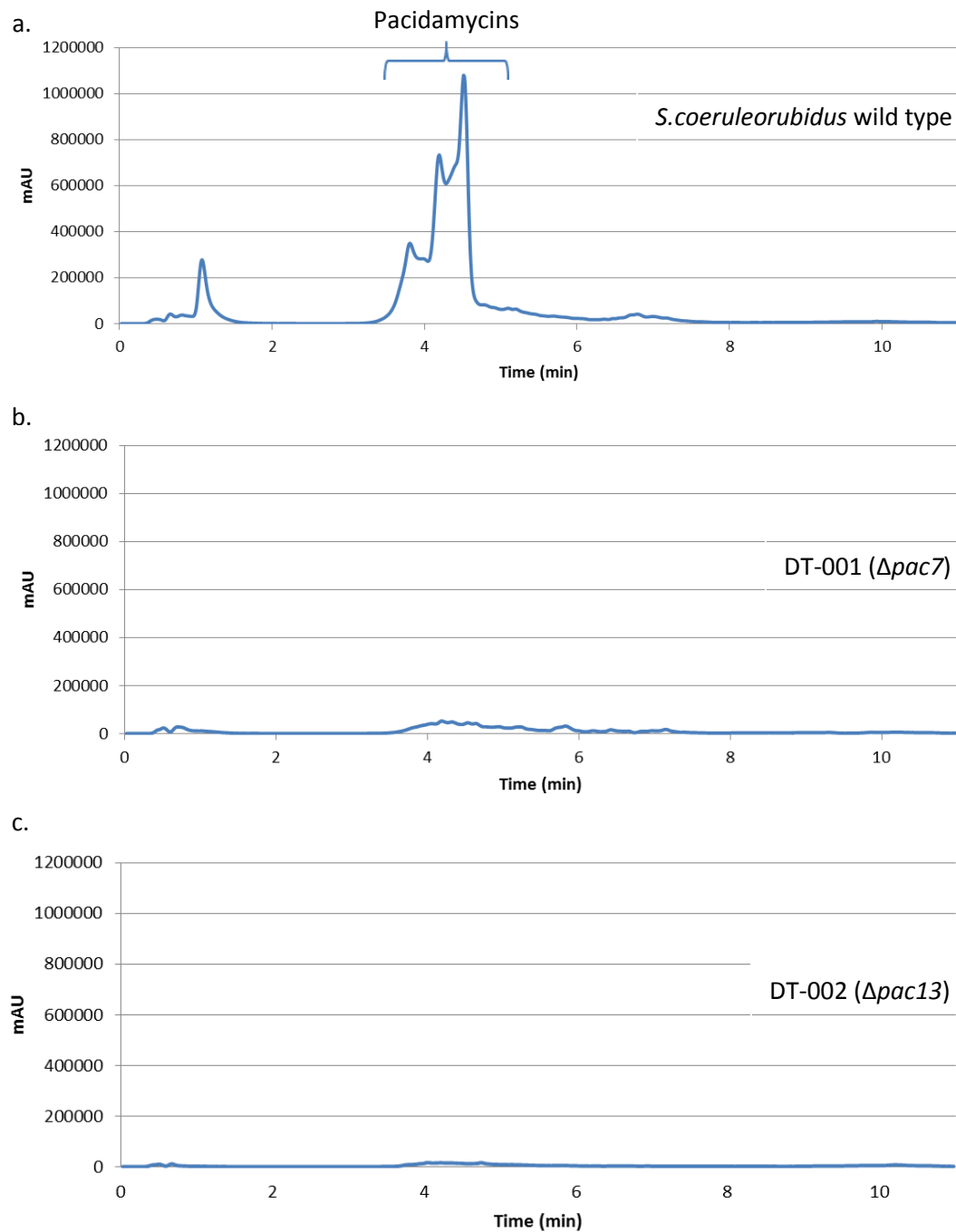


Figure 5.4 The LC profiles for (a) wild type *S. coeruleorubidus* and the two disrupted strains; (b) *S. coeruleorubidus* DT-001 and (c) *S. coeruleorubidus* DT-002. Comparison of the three profiles shows that pacidamycin production is eliminated by the disruption of either *pac7* or *pac13*.

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

produced pacidamycins, corresponding to the hydrated form of the molecules. LC-MS/MS analysis confirmed that in fact, the *S. coeruleorubidus* DT-002 did produce hydrated pacidamycin species and this was further supported with *in vitro* work carried out by A. Ragab (Ragab et al., 2011). Thus, *pac13* was deemed to be responsible for the dehydration of the furan ring.

5.2.1.5 Gene disruption studies of the DABA biosynthetic genes

To complement the biochemical and biophysical approaches to deduce the biosynthesis of DABA, a genetic approach was also undertaken. Rackham et al. (2010) had determined that replacement of the 5' half of *pac19* eliminated the production of pacidamycin in the native producer, *S. coeruleorubidus*. The approach to this study was to determine whether any of the other genes believed to be involved in the biosynthesis of DABA were also essential.

5.2.1.6 Gene disruption studies in *S. coeruleorubidus* of the DABA biosynthetic genes

The initial approach took into account the difficulty of the translational coupling of *pac17* through *pac19* and potential polar effects from single disruptions. The first part of the approach consisted of using PCR mediated gene deletion of the four genes believed to be responsible for the biosynthesis of the diamino acid (*pac17* through *pac20*)(Gust et al., 2003). The deletion of the four genes could then be complemented with different combinations of the four genes that had been removed and pacidamycin production monitored using LC-MS and LC-MS/MS analysis. This approach was initially taken as it allowed for the majority of the laborious genetic work to be carried out in *E.coli* strains rather than in *Streptomyces*; *E. coli* being a much easier organism to manipulate. The second part of the approach was to disrupt the stand alone gene, *pac20*, and also *pac19*, as disruption of the 3' portion of this gene should not have any polar effects on the upstream genes that *pac19* is translationally coupled to.

For the study of the disruption of the entire DABA biosynthetic gene set, primers Redirect_Pac17-20_F and Redirect_Pac20_R were designed as per the procedure outlined in chapter 2. The primers were used to amplify the resistance cassette using the PCR conditions specified in chapter 2. After successful amplification of the disruption cassettes, the fragments were introduced into *E. coli* BW25113 containing the recombination plasmid

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

pIJ790. Recombination of the cassette in place of genes of interest on the 2H-5 cosmid was selected for by growth on the selective medium and PCR analysis. The disrupted cosmid was isolated from a culture of the *E. coli* strain by alkaline lysis and the cosmid introduced into the non-methylating *E. coli* ET12567/pUZ8002 via electroporation to ensure successful conjugation. *E. coli* containing the mutagenesised cosmid was selected for on selective medium and used for conjugal transfer into *S. coeruleorubidus* on SFM agar containing no antibiotics. Exconjugates were streaked onto antibiotic selective SFM agar and finally DNA agar to produce spore stocks from a single exconjugate. Spore stocks were collected of the isolated *S. coeruleorubidus* strain containing the desired replacement and designated *S. coeruleorubidus* DT-005.

The same procedure as described above was also used for the replacement of *pac19* and *pac20*, but using primers Redirect_Pac19_F and Redirect_Pac19_R for the *pac19* replacement and Redirect_Pac20_F and Redirect_Pac20_R for the *pac20* replacement. *S. coeruleorubidus* strains produced during the replacement of *pac19* and *pac20* were designated DT-003 and DT-004, respectively.

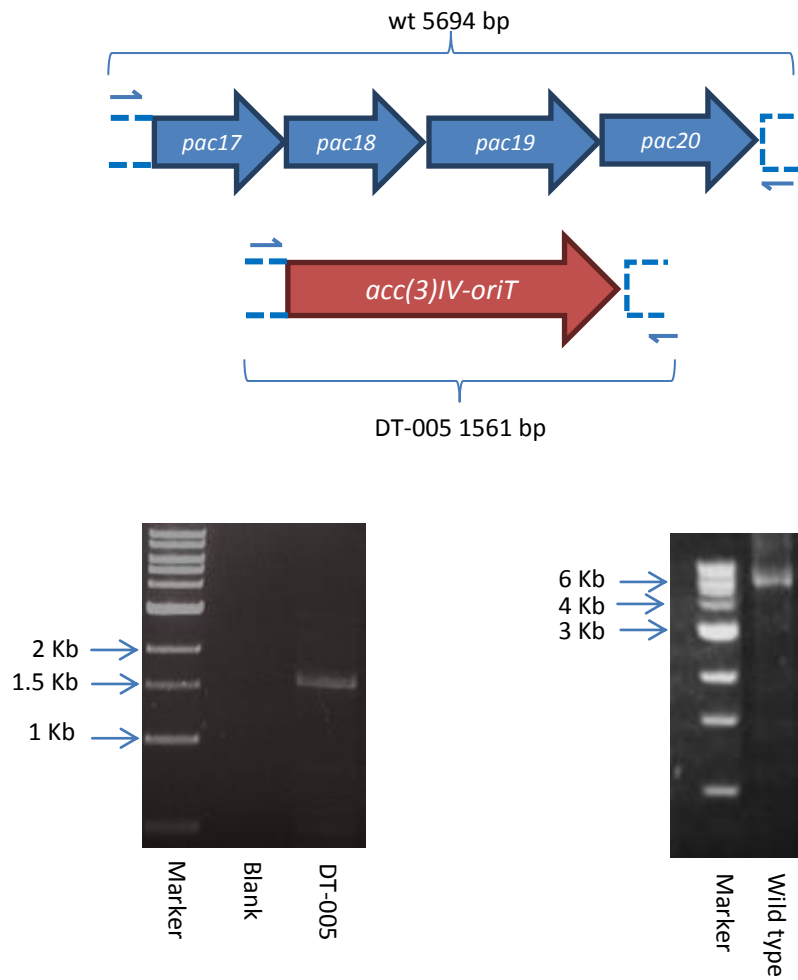


Figure 5.6 The predicted size of the PCR products produced by the amplification of the chromosomal region of the *pac17-pac20* genes in the wild type and disrupted strains along with the PCR analysis.

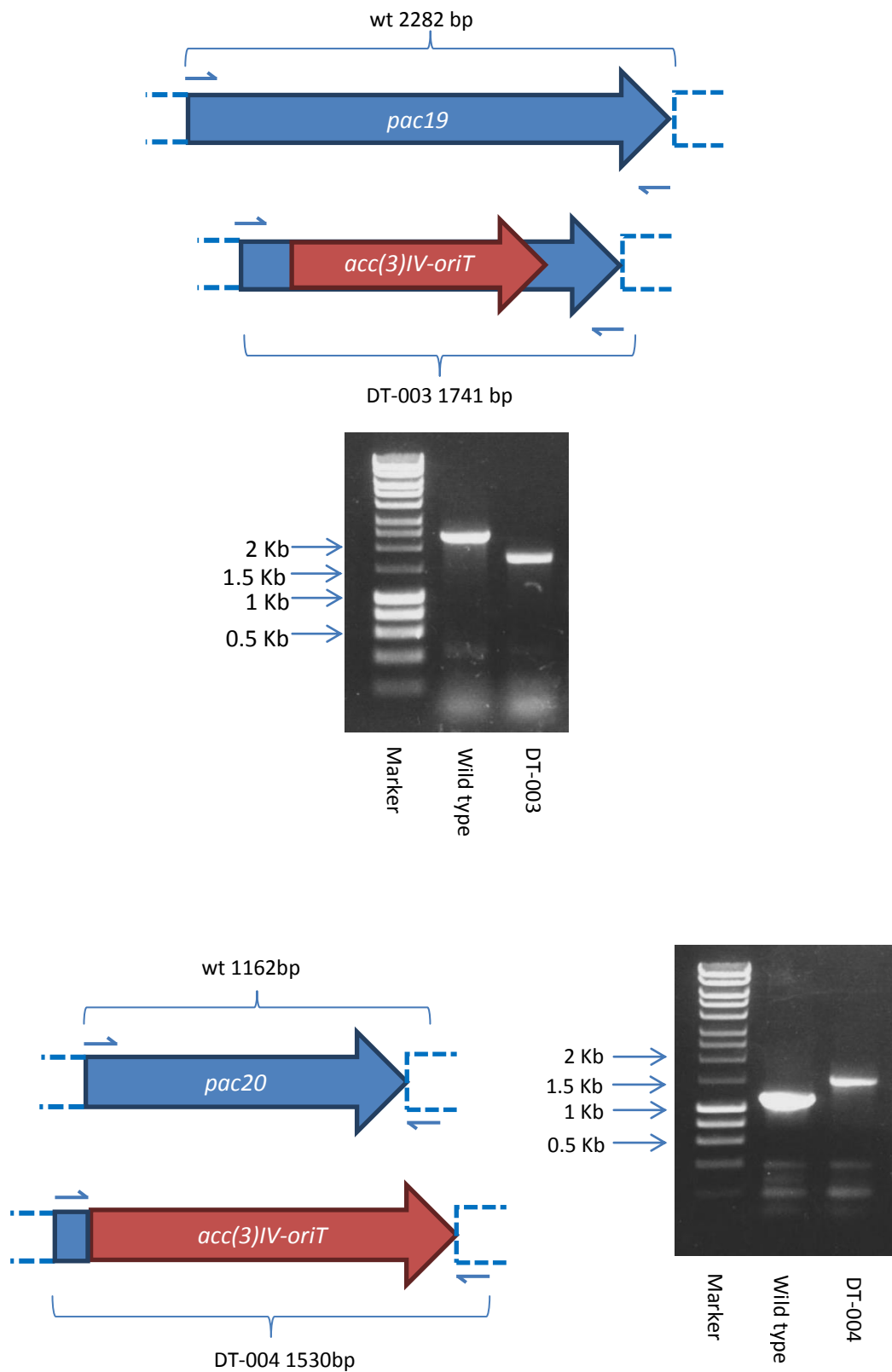


Figure 5.7 The predicted size of the PCR products produced by the amplification of the chromosomal regions of *pac19* and *pac20* in the wild type and disrupted strains along with the PCR analysis.

5.2.1.7 Liquid growth, metabolite extraction and LC-MS analysis of the *S. coeruleorubidus* DT-003, DT-004 and DT-005 and metabolite extraction

The *S. coeruleorubidus* DT-003, DT-004 and DT-005 strains were cultured in ISP₂ liquid medium as per the procedure outlined in chapter 2. Once cultured, the cell free extracts were used in a metabolite extraction method that is outlined in chapter 2 which allowed for the extraction of the pacidamycins for subsequent analysis. After metabolite extraction of the cultures, LC-MS analysis was carried out to confirm the presence or absence of pacidamycin production in the different *S. coeruleorubidus* strains. The LC-MS parameters were identical to those described in the method in chapter 2.

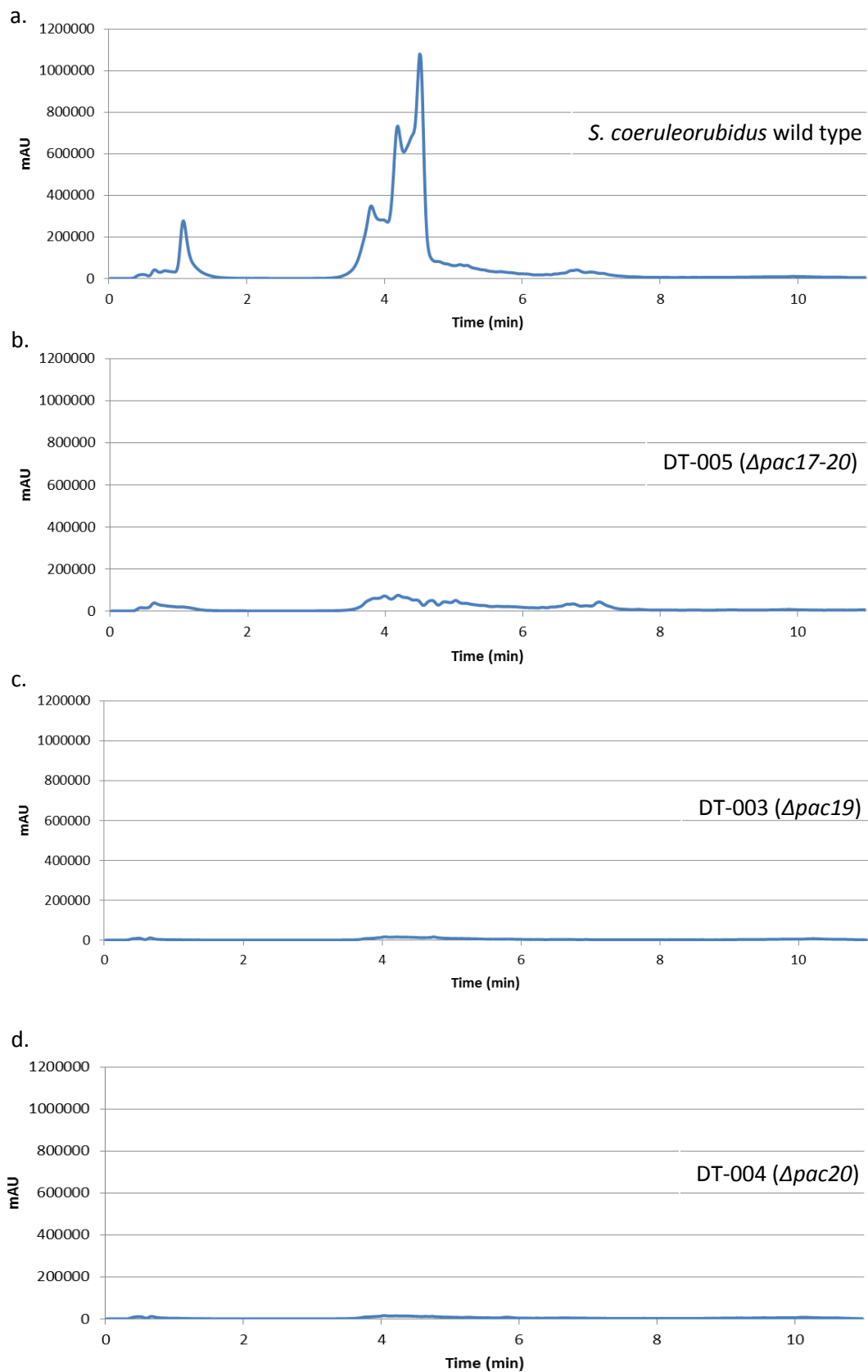


Figure 5.8 LC profiles for (a) wild type *S. coeruleorubidus*, (b) *S. coeruleorubidus* DT-005, (c) *S. coeruleorubidus* DT-003 and (d) *S. coeruleorubidus* DT-004.

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

LC-MS analysis of *S. coeruleorubidus* strains DT-003, DT-004 and DT-005 suggested that deletion of the four hypothesised DABA biosynthetic genes of the pacidamycin gene cluster (*pac17-20*) or the disruption of *pac19* or *pac20* eliminated the ability of *S. coeruleorubidus* to produce pacidamycins.

5.2.1.8 Chemical complementation analysis

As no production of pacidamycin was seen in the *pac19*, *pac20* or *pac17-20* deletion mutants, an attempt to complement these mutants with potential products of the genes or structurally similar compounds as the gene products was undertaken. Rackham et al. (2010) had previously shown that either protected or unprotected DABA could complement the deletion of the 5' half of *pac19*. Unfortunately for this study, no 2*S*,3*S*-diaminobutyric acid was available for this investigation. Chemical complementation of *pac19* and the *pac17-pac20* was attempted using the closest structural homolog to DABA; DAP, which is a C3 diamino acid whereas DABA is a C4 diamino acid. Chemical complementation of *pac20* disruption was attempted using L-threonine and L-allo-threonine, which, from bioinformatics analysis, were the two likely products of the gene's activity (Altschul et al., 1990). None of these chemical complementation attempts were successful. Potential reasons for this include: the compounds were unable to pass into the cell, the compounds were not pathway intermediates or the compounds were metabolised by other cellular enzymes.

5.2.2 Gene disruption studies in the heterologous host *S. lividans* TK-24

As described in chapter 2, heterologous expression in streptomycetes requires the conjugation into bacteria of either a plasmid that can replicate in streptomycetes or a plasmid that can integrate into the genome of the organism. For the heterologous expression of the 2H-5 cosmid, which contains the minimal pacidamycin gene cluster, the backbone of the 2H-5 cosmid needs to be targeted with a *SspI* fragment (as outlined in chapter 2) and integrated into the Φ C31 *attB* site of the *S. lividans*TK24 genome. When integrating the 2H-5 cosmid with disrupted genes present, another step must be carried out prior to the replacement of the 2H-5 backbone with the *SspI* fragment. The disruption cassette, which in this case contains the apramycin resistance gene, needs to be replaced with a 81 bp scar so that the replacement of the backbone with the *SspI* fragment can be

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

selected for (the SspI fragment also contains the apramycin resistance gene as the selectable marker). Replacing the disruption cassette with the scar uses a flippase enzyme and the protocol is outlined in chapter 2.

5.2.2.1 Integration of the non-disrupted 2H-5 cosmid into *S. lividans* TK24 and metabolite extraction and analysis

To ensure that integration of the 2H-5 cosmid into *S. lividans* TK24 was possible, the backbone of the 2H-5 cosmid was replaced with the SspI fragment. As described in chapter 2, the cosmid was conjugated into the heterologous host (*S. lividans* TK24) and the integration of the cosmid into the Φ C31 site of the *S. lividans* genome confirmed by resistance to apra and PCR analysis, using colony PCR analysis as described in chapter 2.

PCR analysis confirmed integration of the cosmid into the correct location and the strain was designated the name *S. lividans* DT-006. *S. lividans* DT-006 was cultured as described in chapter 2, the metabolites extracted as described in chapter 2, and LC-MS and LC-MS/MS analysis carried out to determine whether pacidamycin D was produced, using *S. lividans* TK24 as the negative control. During heterologous expression, only a subset of the pacidamycins (predominately pacidamycin D) are produced due to the absence of *phhA*, which is found elsewhere in *S. coeruleorubidus* genome. *phhA* has been found to be responsible for the synthesis of the meta-tyrosine found in a number of the pacidamycins (Gruschow et al., 2011).

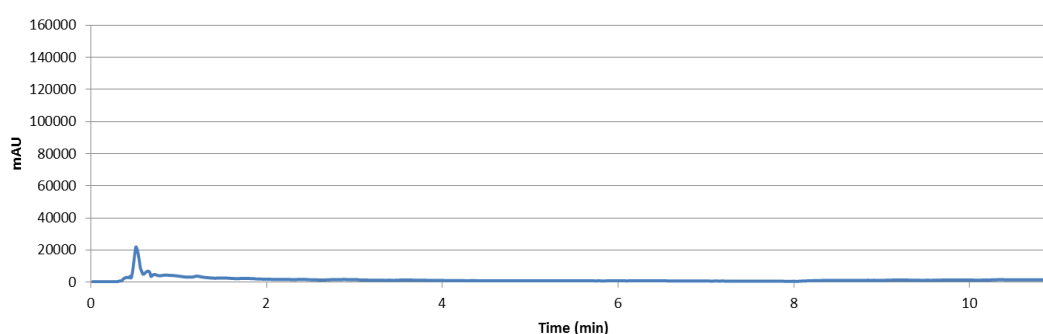


Figure 5.9 The LC profile for the negative control strain (*S. lividans* TK24). The profile shows the absorbance for ions of mass $m/z = 712.0$.

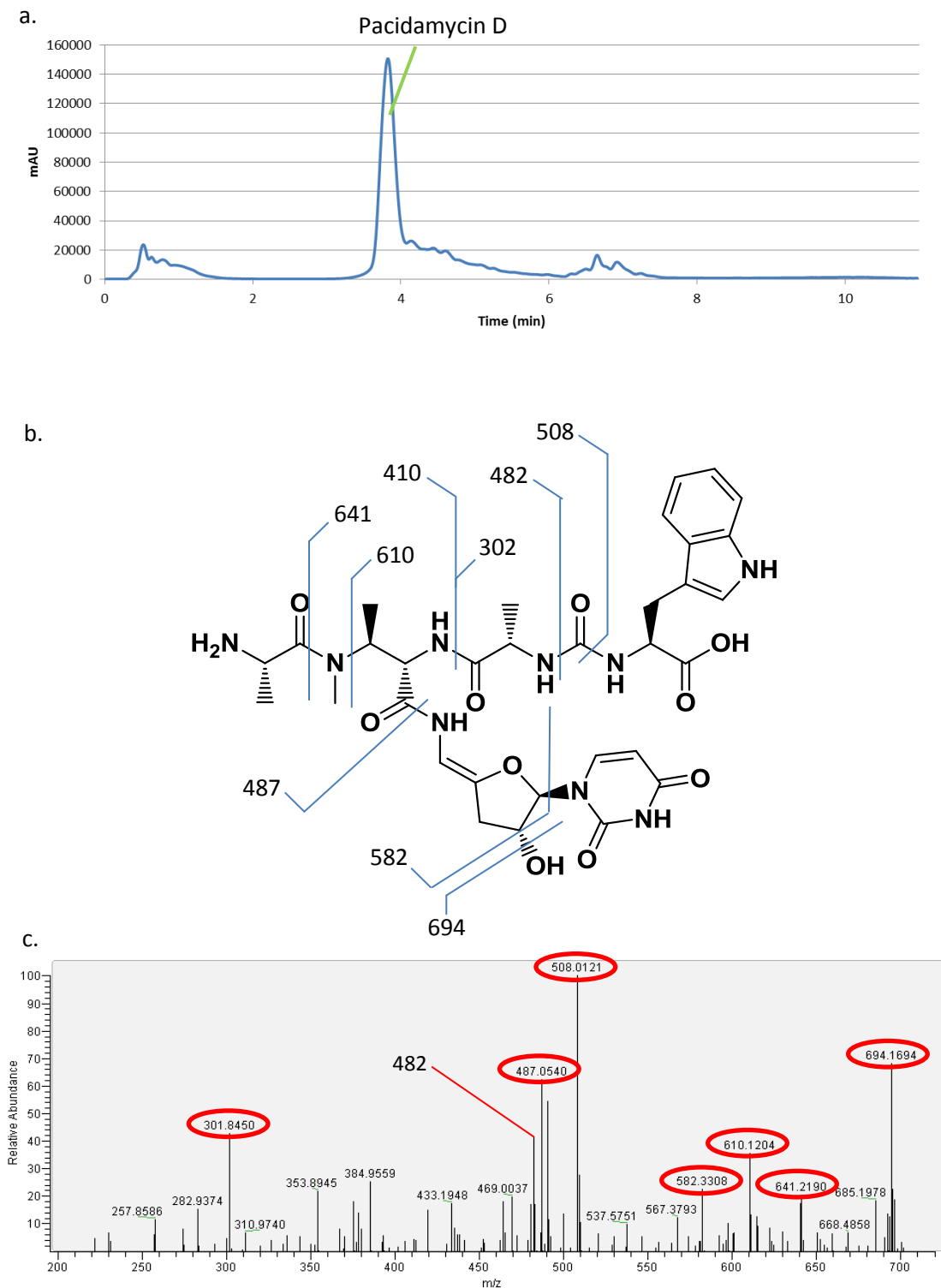


Figure 5.10 (a) The LC profile of metabolites extracted from *S. lividans* when heterologously expressing the minimal gene cluster of pacidamycin. The profile shows the absorbance for ions of mass $m/z = 712.0$ (pacidamycin D), (b) the predicted fragmentation pattern of pacidamycin D and (c) the MS_2 spectrum produced by the analysis; the major peaks in this profile (highlighted in red) correspond to the peaks predicted to be present.

5.2.3 New approach to analysis of gene function

Due to the translational coupling of *pac17*, *pac18* and *pac19* it was decided that the gene disruption of *pac17* and *pac18* would need to be carried out in a heterologous host such as *Streptomyces coelicolor* or *Streptomyces lividans*.

An approach was taken whereby a 'scar' (adaptation from Gust et al., 2003) would be introduced into *pac17* and *pac18* such that the mutated gene and the wild type gene would be of the same size (to minimise the chances of polar effects during transcription or translation), but with the function of the gene abolished.

The method involved using the same strategy as previously described in this chapter, however, as the scar that is introduced is only 81 bp in size, only 81 bp from each of the genes would be removed. This method posed the problem of which 81 bp to remove as the removal of these bases needed to eliminate the gene's functionality.

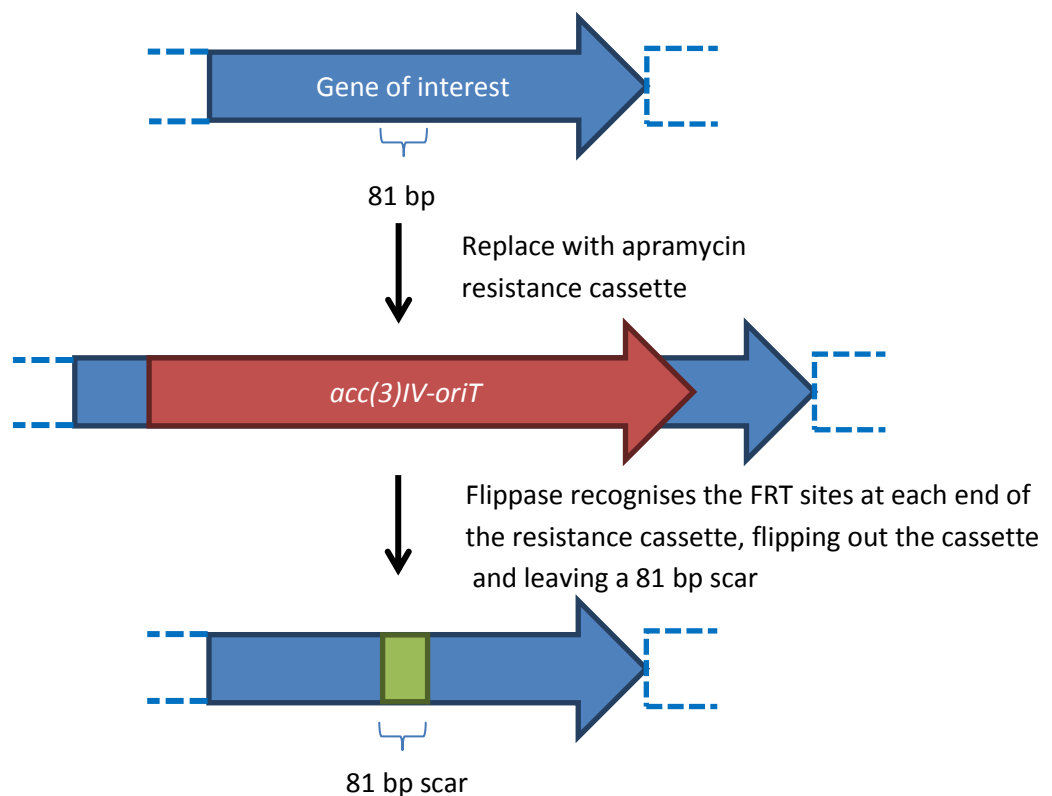


Figure 5.11 The strategy for mutation of *pac17* and *pac18*, replacing a 81 bp sequence with a scar of the same size, to reduce the chances of polarity.

5.2.3.1 Construction of the disrupted *pac17* and *pac18* integration cosmids

Determining the ideal 81 bp to replace in *pac17* and *pac18* was carried out in two different ways. For *pac17*, the structure of the protein had previously been determined (described in chapter 4). The structural elucidation of the protein had shed some light on the active site and which residues were key for the activity of the enzyme. The section of the gene that encoded for this key region was therefore identified and primers designed (Redirect_Pac17_F and Redirect_Pac17_R) to enable the replacement of 81 bp of this region with the resistant cassette and ultimately with the FRT scar on cosmid 2H-5, the cosmid then integrated into the Φ C31 *attB* site of the *S. lividans*TK24 genome as previously described, being designated strain number *S. lividans* DT-007.

For *pac18*, the amino acid sequence was aligned with sequence homologs using BLAST and a well conserved region of 27 amino acids identified (Altschul et al., 1990). Using cosmid 2H-5, the corresponding nucleotide sequence (81 bp in total) of *pac18* was replaced with the resistance cassette and ultimately by the FRT scar and the cosmid integrated in the Φ C31 *attB* site of the *S. lividans*TK24 genome as previously described (primers used: Redirect_Pac18_F and Redirect_Pac18_R), being designated strain number *S. lividans* DT-008.

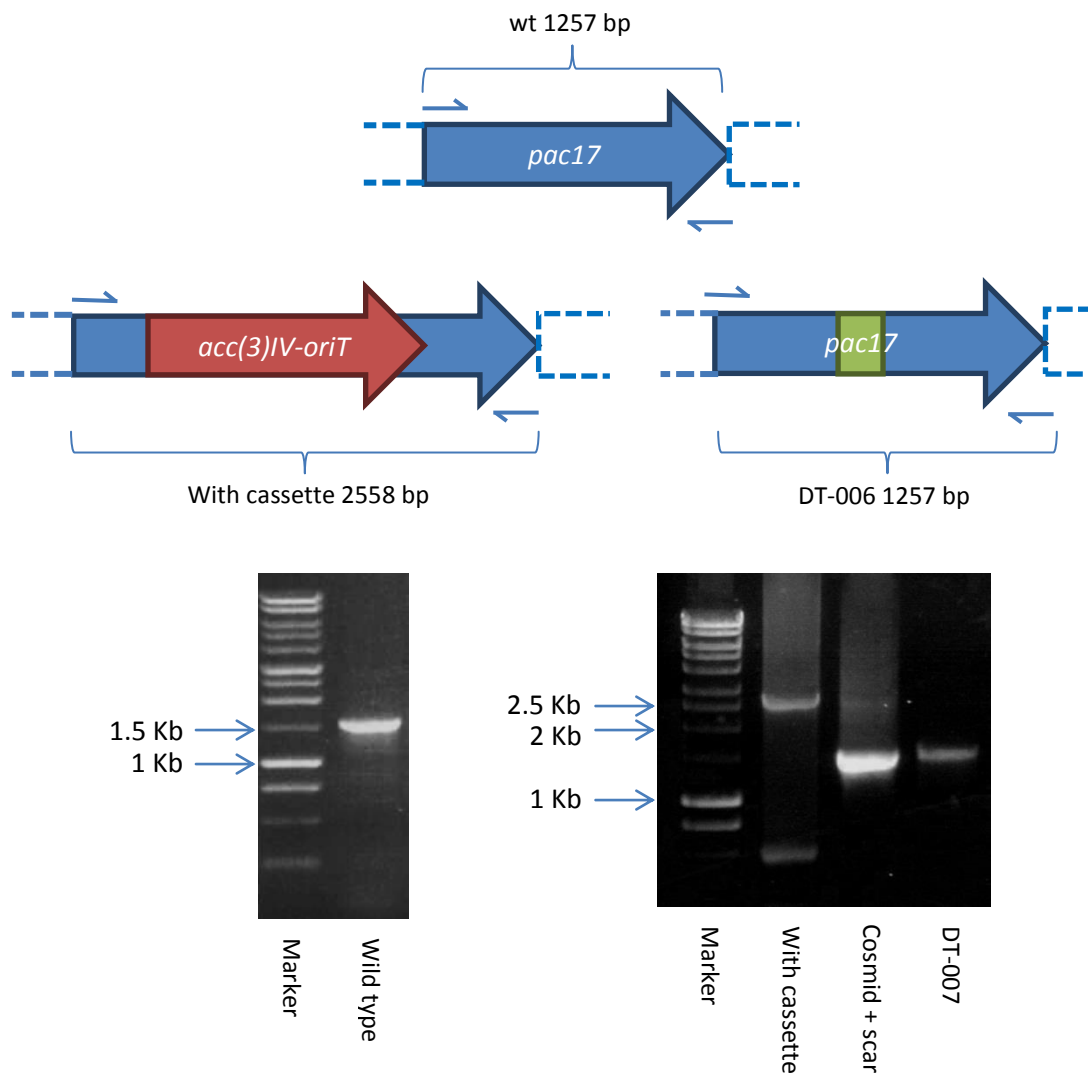


Figure 5.12 The predicted size of the PCR products produced by the amplification of the *pac17* wild type and disrupted strains along with the PCR analysis.

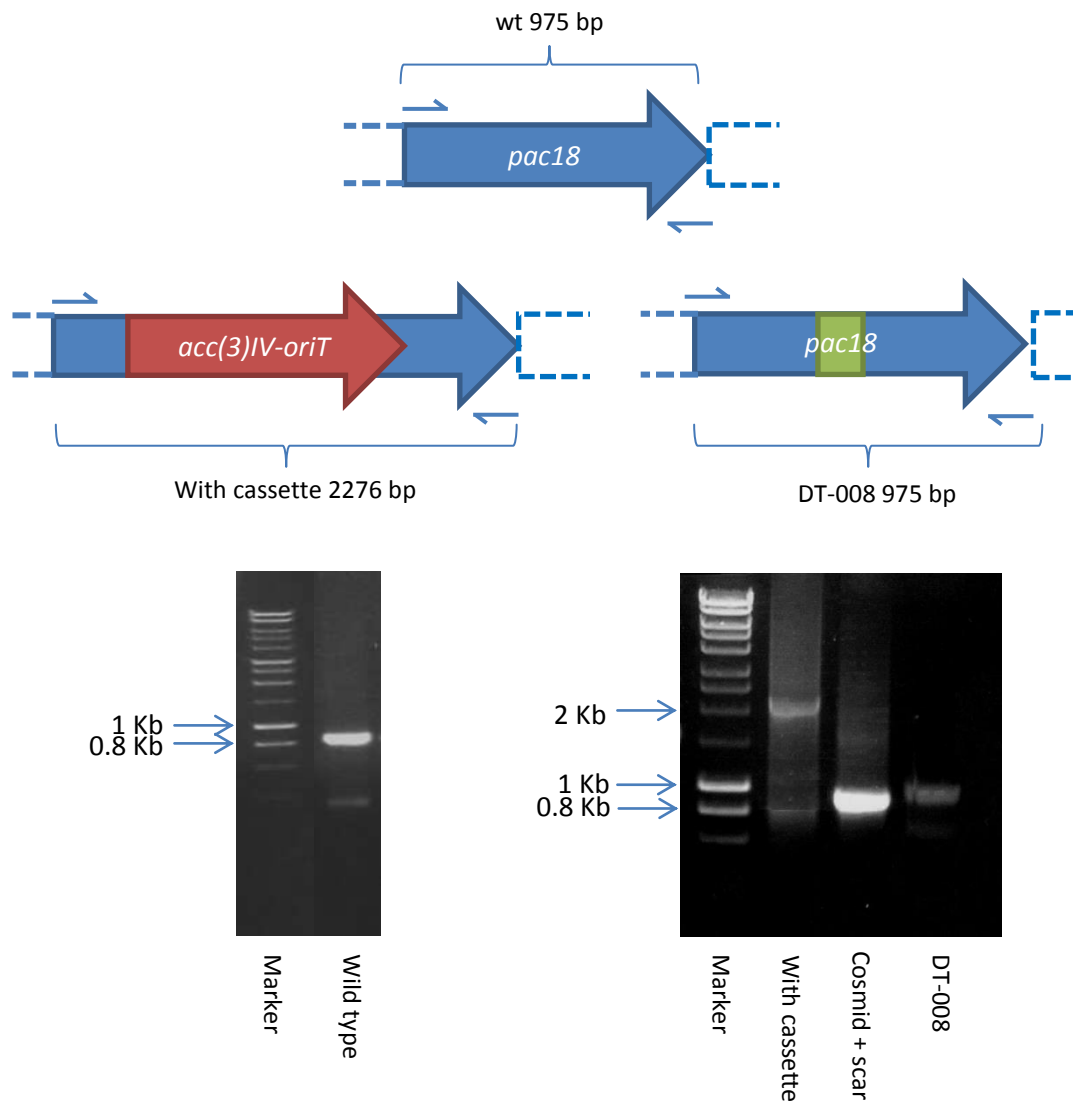


Figure 5.13 The predicted size of the PCR products produced by the amplification of the *pac17* wild type and disrupted strains along with the PCR analysis.

5.2.3.2 Growth of *S. lividans* strains DT-007 and DT-008, metabolite extraction and analysis

The *S. lividans* strains were incubated in R5 liquid medium as described in chapter 2. The metabolites were extracted as previously described and analysed by LC-MS analysis, searching for pacidamycin D ions (m/z 712) using *S. lividans* DT-006 as a positive control and *S. lividans* TK24 as a negative control.

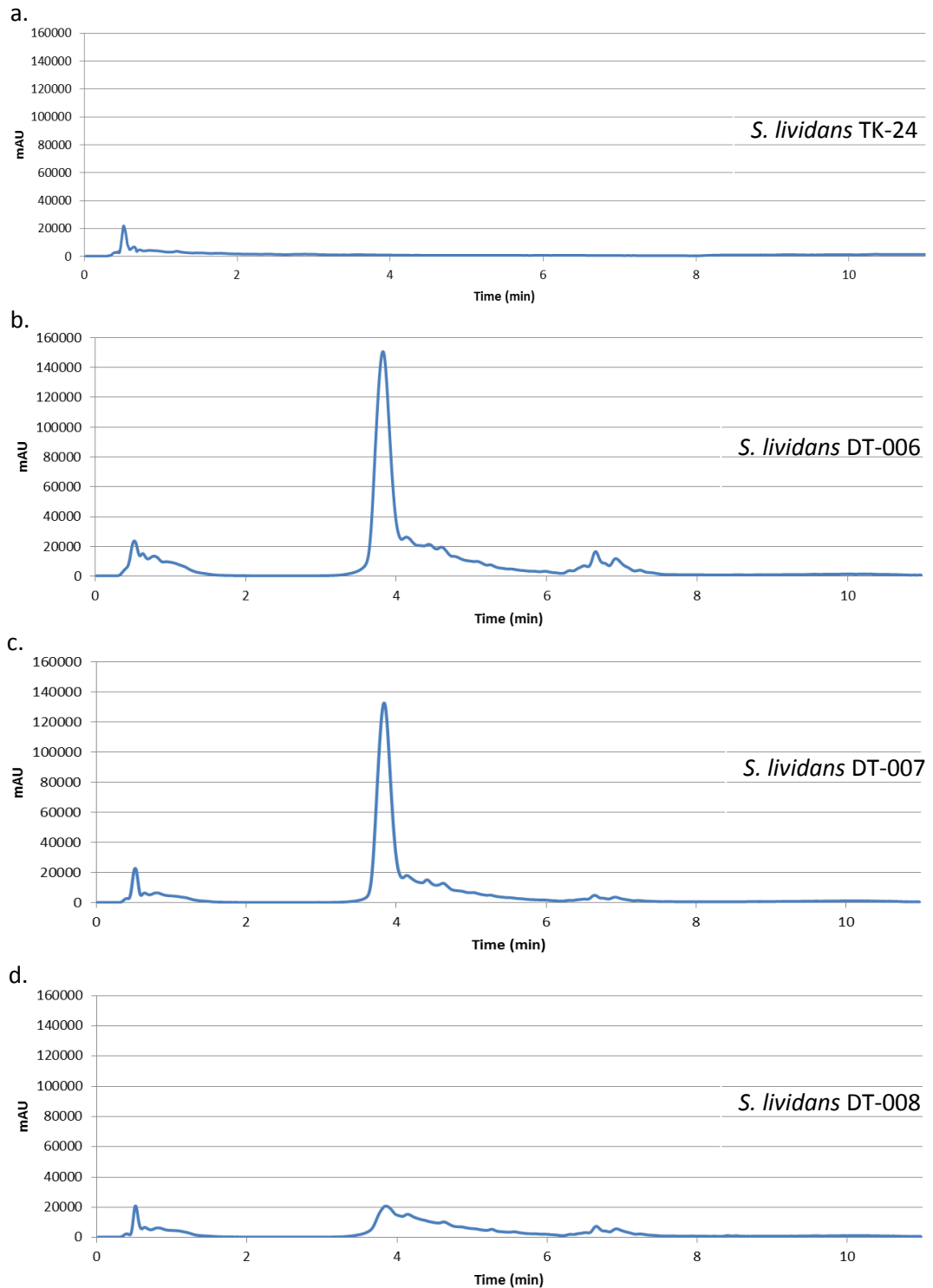


Figure 5.14 The LC profiles for (a) negative control (*S. lividans* TK24), (b) positive control (*S. lividans* DT-006), (c) *S. lividans* DT-007 and (d) *S. lividans* DT-008

The spectra suggested that disruption of *pac17* had very little effect on production of pacidamycin D in the heterologous host. However, it is extremely likely that the Pac17

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

protein was complemented by other proteins present in the organism. A BLAST analysis of the Pac17 amino acid sequence suggests that homologs are common among the streptomycetes and bacteria in general (Altschul et al., 1990). LC-MS/MS analysis of the *S. lividans* DT-008 metabolite extraction was undertaken as there appeared to be absorbance at the correct retention time for pacidamycin D. LC-MS/MS analysis showed that in fact pacidamycin D was still being produced by the heterologous host when *pac18* had been disrupted, although at a much lower level compared to the positive control.

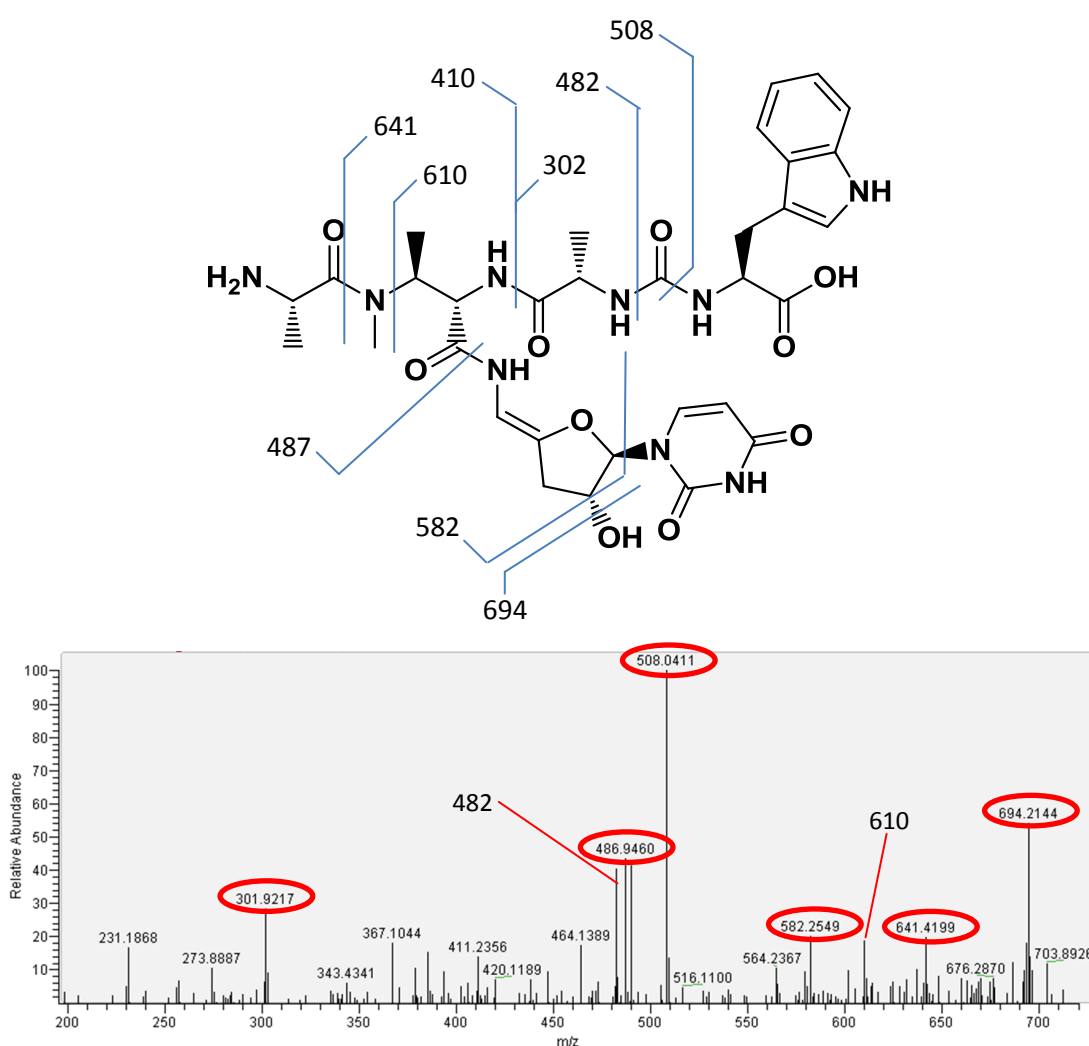


Figure 5.15 Showing the fragments predicted to be observed by MS₂ analysis of pacidamycin D (top) and the MS₂ spectra produced by the analysis of *S. lividans* DT-008.

To determine if *S. lividans* TK24 contains a protein that could be complementing the gene disruption, a nucleotide search was carried out on the partial nucleotide sequence of the *S. lividans* TK24 genome (supplied by G. Chandra, JIC) using the nucleotide sequence of *pac18* as the query sequence. Unfortunately no 'hits' were found in this search. As *S. lividans* TK24 is extremely homologous to *S. coelicolor*, it was decided that a protein blast on the known *S. coelicolor* sequence could be carried out and any 'hits' could be confidently assumed to also be present in *S. lividans* TK24 (Leblond et al., 1993). A BLAST search resulted in the identification of a protein produced by *S. coelicolor* with a 83 % coverage and 30 % identity to Pac18. This protein could be complementing the disruption of *pac18*, allowing for the continued production of pacidamycin even when the gene is disrupted. Further to this, the pacidamycin biomachinery may still be able to produce pacidamycin even when Pac18 is not present. As Pac18 is predicted to be a kinase which phosphorylates threonine at the β -hydroxy position to create a better leaving group for the downstream reaction, it is plausible that Pac19 is able to process both the phosphorylated and non-phosphorylated forms of threonine. It must be reiterated that although Pac18 does not appear to be essential for pacidamycin production or it is being complemented by other cell machinery, the amount of pacidamycin produced when the gene is disrupted appears to be greatly reduced, suggesting that although the protein is not essential, it has a drastic impact on the quantity of pacidamycin produced by the organism.

5.2.4 Gene disruption studies of the DABA biosynthetic genes *pac19* and *pac20* in the heterologous host *S. lividans* TK-24

5.2.4.1 Construction of the disrupted *pac19* and *pac20* integration cosmids

For the construction of the integrated cosmids containing disruptions of *pac19* and *pac20*, the cosmid constructs 2H-5 Δ *pac19* and 2H-5 Δ *pac20* were used. The apramycin cassette was replaced with an 81 bp scar and the backbone was replaced with the SspI fragment to produce cosmids 2H-5 Δ *pac19*-integration and 2H-5 Δ *pac20*-integration, respectively. *S. lividans* TK-24 containing the integrated 2H-5 cosmid with the scar in place of *pac19* or *pac20* were designated strain numbers DT-009 and DT-010, respectively. The integration of the cosmids into the *S. lividans* TK-24 Φ C31 *attB* site was confirmed by the ability of the transformants to grow on apramycin supplemented medium and by PCR analysis.

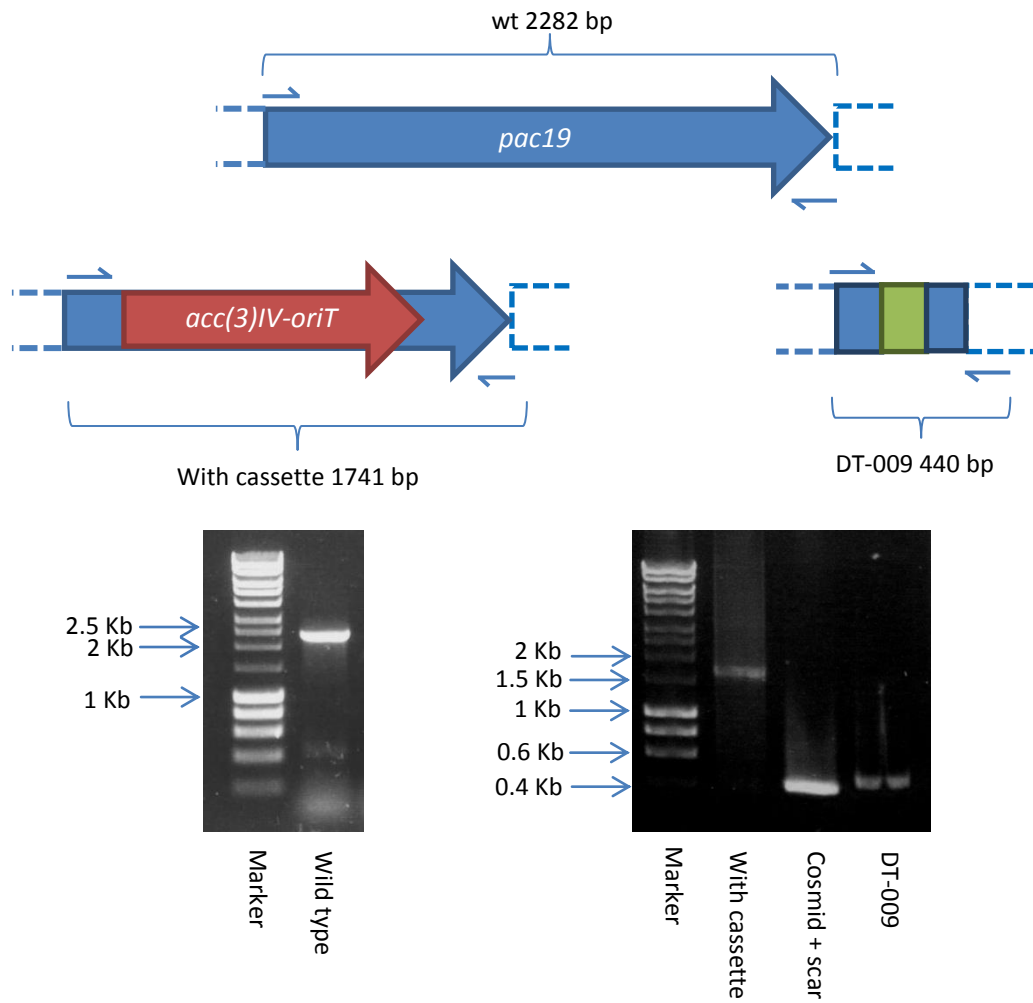


Figure 5.16 The predicted size of the PCR products produced by the amplification of the *pac19* wild type and disrupted strains along with the PCR analysis.

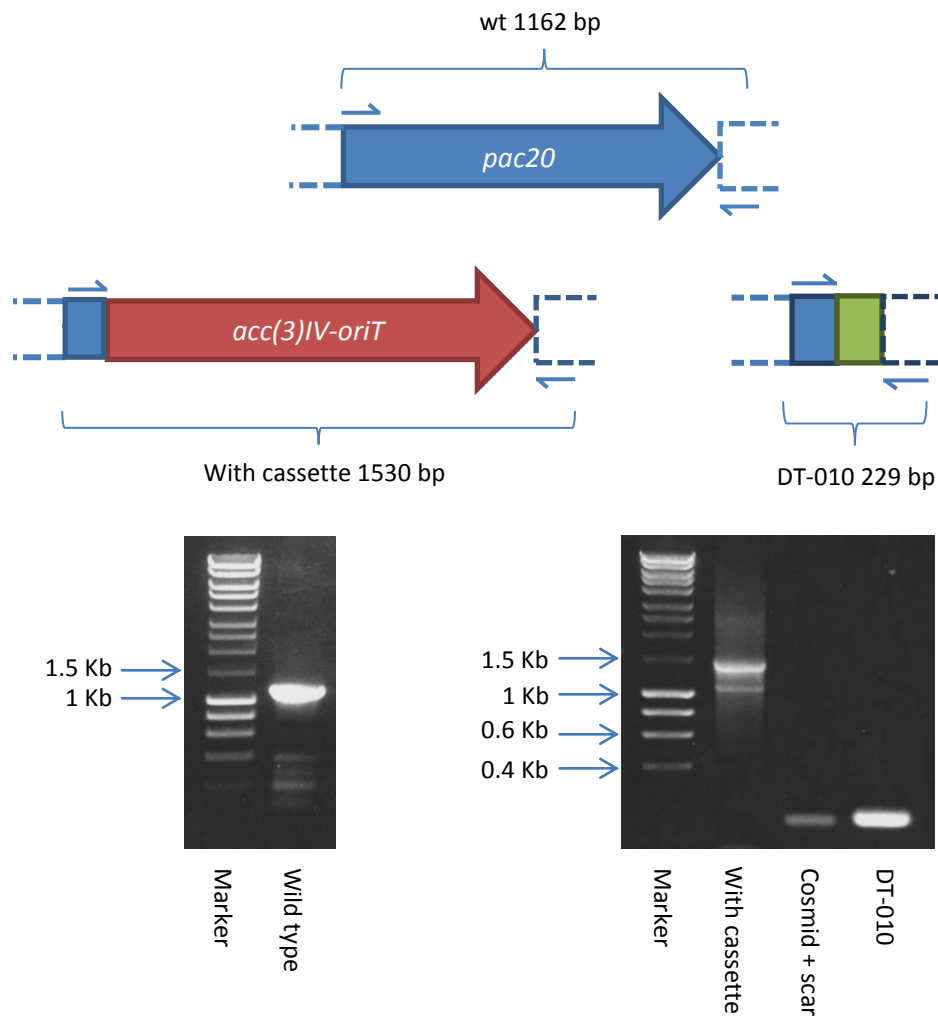


Figure 5.17 The predicted size of the PCR products produced by the amplification of the *pac20* wild type and disrupted strains along with the PCR analysis.

5.2.4.2 Growth of *S. lividans* strains DT-009 and DT-010, metabolite extraction and analysis

The *S. lividans* strains were incubated in R5 liquid medium as described in chapter 2. The metabolites were extracted as previously described and analysed by LC-MS analysis, searching for pacidamycin D ions (m/z 712) using *S. lividans* DT-006 as a positive control and *S. lividans* TK24 as a negative control.

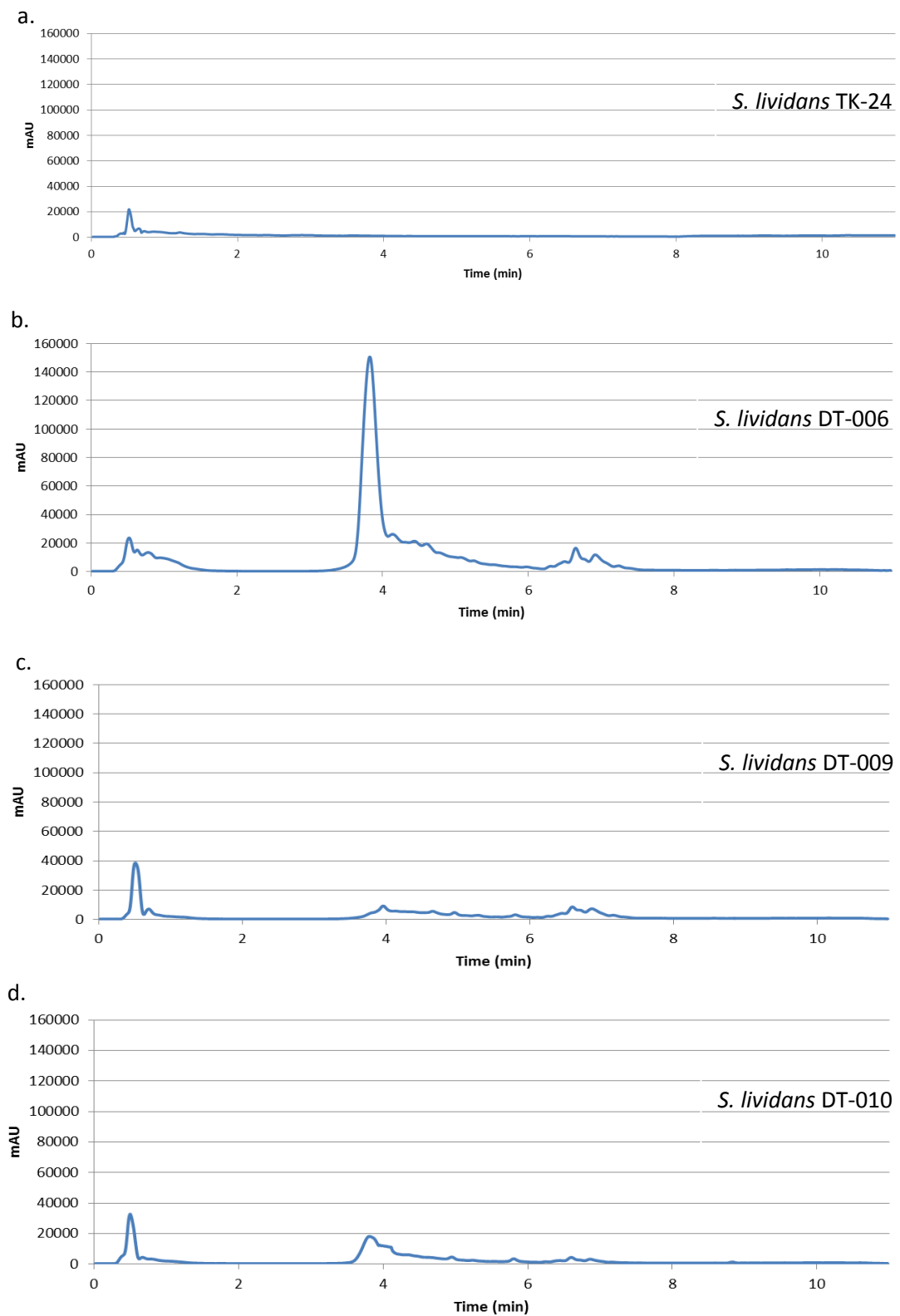


Figure 5.18 LC profiles for (a) negative control (*S. lividans* TK24), (b) positive control (*S. lividans* DT-006), (c) *S. lividans* DT-009 and (d) *S. lividans* DT-010.

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

From the LC-MS analysis of *S. lividans* DT-009 and DT-010 which heterologously express the minimal pacidamycin gene cluster with disruptions of *pac19* and *pac20*, respectively, the first observation is that the peak corresponding to pacidamycin D is absent. Further investigation of the spectra suggested that pacidamycin D is being produced by these strains, but at a much lower level. LC-MS/MS analysis was carried out to determine whether fragmentation of the parent $m/z = 712$ ions resulted in a pattern that was consistent with the expected fragmentation of pacidamycin D. LC-MS/MS analysis showed that no pacidamycin D was being produced by *S. lividans* DT-009 (disrupted *pac19* gene), which is consistent with the previous studies in *S. coeruleorubidus* and by Rackham et al. (2010). LC-MS/MS analysis of *S. lividans* DT-010 showed that pacidamycin D was being produced by this strain (Figure 5.19), which is inconsistent with the parallel study carried out in *S. coeruleorubidus* which is reported earlier in this chapter. These results suggest that *pac20* is complemented in the heterologous host, therefore allowing for the continued production of pacidamycin D, whereas, the disruption of this gene cannot be complemented in the native pacidamycin producer.

A nucleotide search was carried out on the partial nucleotide sequence of the *S. lividans* TK24 genome (supplied by G. Chandra, JIC) using the nucleotide sequence of *pac20* as the query sequence. Unfortunately no 'hits' were found in this search. As *S. lividans* TK24 is extremely homologous to *S. coelicolor*, it was decided that a protein BLAST on the known *S. coelicolor* sequence could be carried out and any 'hits' could be confidently assumed to also be present in *S. lividans* TK-24 (Leblond et al., 1993). The BLAST analysis retrieved two proteins produced by *S. coelicolor* that showed maximum identities of 30 and 26 % to the amino acid sequence of *pac20*. These corresponded to a L-allo-threonine aldolase and aldolase, respectively (Altschul et al., 1990). Interestingly, the highest 'hit' was a L-allo-threonine aldolase, which may suggest that the starting substrate for the production of pacidamycin is in fact the L-allo rather than the L- form of threonine. This could also explain why a gene that is highly homologous to one associated with primary metabolism is found in a secondary metabolism gene cluster. Even so, it was previously reported in this chapter that the feeding of both L-threonine and L-allo-threonine to *S. coeruleorubidus* DT-004 failed to reinstate pacidamycin production in the natural producer, however, this could be a result of the organism being unable to take up the compounds from the liquid medium.

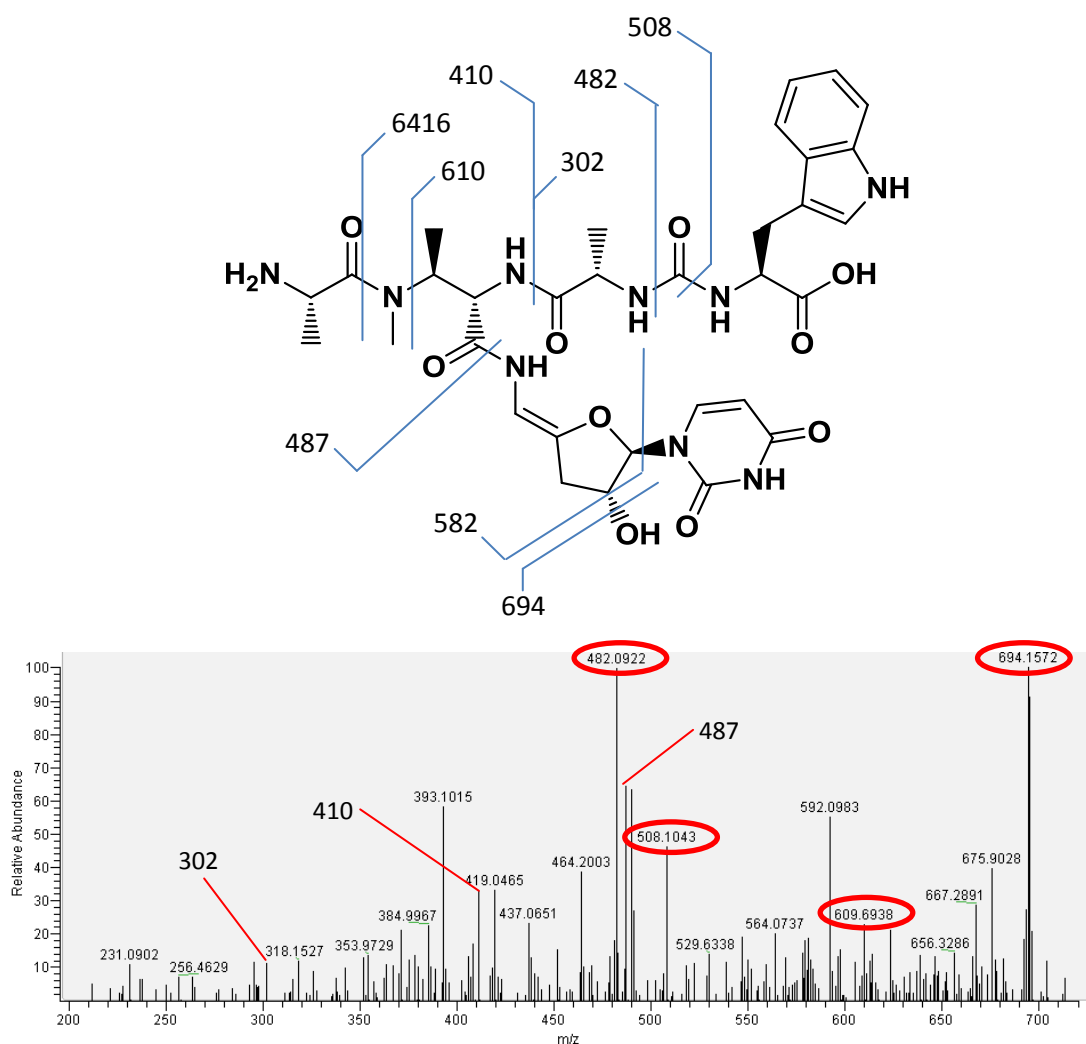


Figure 5.19 The fragments predicted to be observed by MS₂ analysis of pacidamycin D (top) and the MS₂ spectra produced by the analysis of *S. lividans* DT-010.

5.3 Conclusion

A number of conclusions can be drawn from these *in vivo* studies of the hypothetical proteins and putative DABA biosynthesis genes present in the pacidamycin gene cluster.

This study, in collaboration with A. Ragab was able to determine the function of *pac13*, a gene previously identified as a hypothetical protein. It is responsible for the dehydration of the furan ring in the uridyl nucleoside moiety found within the pacidamycin structure. *pac7* has also been found to be important in the biosynthesis of pacidamycin, as disruption of this gene appears to eliminate the production of the antibiotic in the native producer *S. coeruleorubidus*. Unfortunately, to date the function of this gene remains unknown.

Chapter 5 – Pacidamycin Cluster Gene Disruption Studies

Current analysis of the role of *pac17* – *pac20* in the biosynthesis of DABA and pacidamycin suggest that both *pac19* and *pac20* are essential for pacidamycin production in the native producer. The disruption of *pac19* has previously been found to be complemented by feeding the disrupted strain a combination of DABA and protected DABA (Rackham, 2010). The disruption of *pac17* in a heterologous host does not show a significant change in the organism's ability to produce pacidamycin, however, a BLAST search suggests that there are potential candidate genes that could complement this disruption. In order to fully assess the role of *pac17* in pacidamycin production it will be necessary to take a different approach which allows for the study of gene in the context of the native producer. The disruption of either *pac18* or *pac20* does not seem to eliminate pacidamycin production in a heterologous host, however, disruption of *pac20* in *S. coeruleorubidus* does appear to stop pacidamycin production. However, in both cases in the heterologous host, the ability of the organism to produce pacidamycin is drastically reduced. BLAST analysis of the amino acid sequence of the proteins expressed by these genes suggest that both have homologs in *S. coelicolor* and are therefore likely to have homologs in the heterologous producer *S. lividans* TK24. Although the disruption of *pac20* in the heterologous host appears to be complemented in some way, feeding the *pac20* disrupted *S. coeruleorubidus* strain with potential products of *pac20*, i.e. L-threonine and L-allo-threonine did not appear to reinstate the production of pacidamycin.

To complement this current study a number of future investigations could be carried out. Firstly, using a new approach to allow the disruption of all four DABA biosynthesis genes in the natural producer, *S. coeruleorubidus*, would allow for a more robust analysis to study the function and the importance of the genes. As previously explained in this chapter, a heterologous study approach was carried out on *pac17* and *pac18* to overcome the issue of *pac17-pac19* being translationally coupled. Further to this, genetic complementation of the gene disruptions would enable a more in depth analysis, providing greater confidence that the observed chemotypes are direct effects of the gene disruptions. Genetic complementation was attempted within this study, however, due to the lack of available plasmids that can be integrated into *S. coeruleorubidus* (being only pIJ10257), the attempts failed either during the cloning or conjugal transfer stages.

***In Vitro* Studies of the DABA
Biosynthesis Genes**

6.1 Introduction

When investigating the biological function of a gene, the gold standard is to observe the activity of the protein encoded for by the gene *in vitro*. This allows for the investigation of the protein in a simpler environment, away from other cellular functions which would otherwise make the results difficult to interpret. A potential danger of *in vitro* studies is that a protein may become separated from some component that is essential for its correct activity, eg. an enzyme might lose its cofactor with a concomitant loss of catalytic activity.

It has previously been described in this thesis that a number of the proteins postulated to be involved in the biosynthesis of DABA were successfully cloned, produced and purified. Further to this, studies have been undertaken into the structure of the proteins produced (most successfully with Pac17) and the study of the genes that encode the proteins has been investigated *in vivo*, both in the natural host and by heterologous expression.

In this chapter, the attempts to obtain *in vitro* activity for Pac17, Pac19 and Pac20 will be discussed. A number of different approaches will be discussed and possible future developments also suggested.

As previously described, the activities of the DABA biosynthetic enzymes found in the mureidomycin gene cluster have been investigated (Lam et al., 2008). Further to this, work has been carried out on the structurally similar amino acid DAP (Wang and Gould, 1993, Beasley et al., 2011). Other than the results of these studies, little is known about the biosynthesis of DABA.

6.2 Pac17

6.2.1 Introduction

Pac17 has been the most studied of the four DABA biosynthetic genes within this study. Its structure, obtained through an X-ray crystallographic approach has already been discussed (chapter 4). Prior to a more in-depth crystallographic study, a number of attempts were made to elucidate the activity of the protein *in vitro*. An analysis of the gene cluster of pacidamycin reported by Rackham et al. (2010), had postulated the protein was likely to be a lyase, with its closest sequence homologs being argininosuccinate lyases, fumarases and aspartases. From this work it was hypothesised that Pac17 was likely to be an aspartase, catalysing the lysis of aspartate into fumarate and ammonia, the ammonia subsequently

being used by Pac19 to convert a stereoisomer of threonine to DABA (Rackham et al., 2010).

Further to this, the study into the structure of Pac17 also shed some light into the likely substrate of the protein. Co-crystallisation of the protein with aspartate resulted in the observation of a ligand bound structure, with aspartate being found in what appears to be the active site. Further analysis of this structure using the PDBeSUM server found a structural homolog which also has aspartate bound as a ligand (Laskowski, 2001, Fibriansah et al., 2011). The study into this homolog of Pac17, along with other studies, reported *in vitro* activity of the protein as an aspartase and the importance of a number of active site residues (Weiner et al., 2008, Veetil et al., 2009). The studies used a simple but effective method for observing activity. Since the absorbances of aspartate and fumarate are different, one is able to measure the increase or decrease of either the substrate or product over time using a spectrophotometric assay. Weiner et al. (2008) report that the formation of fumarate can be measured at a wavelength of 240 nm. The study suggests that the activity is best observed at the alkaline pH of 8.5 (Weiner et al., 2008). An alkaline pH would be expected due to the postulated catalytic mechanism for the lyase that was discussed in chapter 4. The basic conditions being required for the deprotonation of the catalytic serine in preparation for the subsequent downstream steps in the mechanism (Fibriansah et al., 2011).

6.2.2 Results of the investigation into the activity of Pac17

To ensure that the method for measuring fumarate concentration was robust, two standards (one of L-aspartic acid and one of fumarate) were prepared and diluted to 1 mM concentration in 50 mM NaH₂PO₄ buffer pH 8.5. The absorbance of the two solutions was measured across a wavelength range of 200 – 500 nm. The resulting spectra suggested that measuring the formation of fumarate at 240 nm was a viable approach.

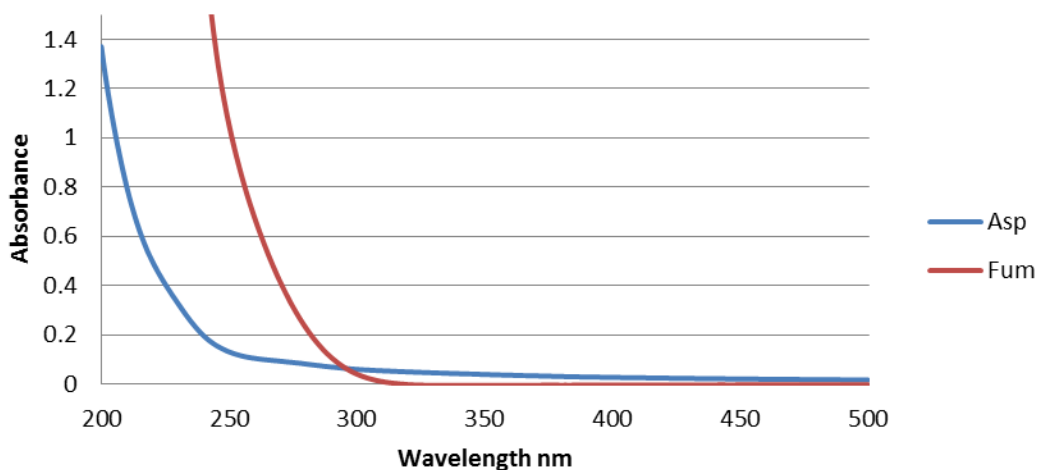


Figure 6.1 Spectra produced by the 1 mM L-aspartate and fumarate standards to determine whether measuring the formation of fumarate at 240 nm was appropriate for the study.

The continuous assay for measuring the activity of Pac17 was set-up as per the parameters outlined by Weiner et al. (2008) and reported in chapter 2. Unfortunately at the basic pH of 8.5, no formation of fumarate was observed by Pac17.

In an attempt to obtain any such activity from Pac17, a systematic approach was undertaken which included investigating the effects of pH, potential co-factors and varying the concentration of the substrate added (between 1 mM and 500 mM) to the reaction. It was previously reported that magnesium may play a role in the activity of the aspartase, AspB, and potassium was found bound to Pac17 during structural studies, therefore the addition of different metals was also considered. Table 6.1 shows the different conditions which were investigated in an attempt to observe Pac17 activity. Unfortunately, none of these conditions gave rise to measurable Pac17 activity.

Chapter 6 – In Vitro Studies of the DABA Biosynthesis Genes

Table 6.1 List of the different conditions that were used in an attempt to observe Pac17 activity using the spectrophotometric assay described in chapter 2. The concentration of co-factors and metals used was 10 mM.

Buffers	pH	Cofactors	Metals
50 mM NaH ₂ PO ₄	6.0	PLP	Potassium (KCl)
50 mM KH ₂ PO ₄	6.5	ATP	Sodium (NaCl)
20 mM Hepes	7.0		Iron (FeCl ₂)
	7.5		Magnesium (MgCl ₂)
	8.0		Cobalt (CoSO ₄)
	8.5		Zinc (ZnCl ₂)
	9.0		Manganese (MnCl ₂)
			Nickel (NiCl ₂)

Another approach for observing the activity of Pac17 is by studying the reverse reaction (i.e. the formation of L-aspartate), which may be more favourable under the experimental conditions used. Crystallographic observations found that when co-crystallised with L-aspartate, Pac17 was found bound to the ligand. The fact that aspartate and not fumarate was found bound suggests that aspartate was not being turned over by the protein, and that maybe the formation of aspartate from fumarate and ammonia was more favourable. Weiner et al. (2008) report measuring the kinetic parameters for the addition of ammonia onto fumarate (as well as the breakdown of aspartate), but using different experimental conditions (outlined in chapter 2). This method was also attempted, but all attempts failed to show Pac17 activity (by measuring the depletion of fumarate at 270 nm).

6.2.3 GC-MS

Since no activity for Pac17 was observed using spectrophotometric methods, a mass spectrometric approach was attempted which was hoped to allow for a more sensitive analysis. For GC-MS analysis, which was carried out by the Metabolomics Facility (JIC), Pac17 was incubated overnight in NaH₂PO₄ buffer pH 8.5 and 1 mM of either L-aspartic acid or fumarate and ammonia (in the form of ammonium chloride). GC-MS analysis was then carried out on the two incubations in an effort to determine whether any turnover of the substrates had occurred. Unfortunately this analysis did not indicate any activity in either direction.

6.2.4 Use of cell free extract and incubation with *S. coeruleorubidus*

Two crude methods were undertaken in an attempt to obtain *in vitro* activity of Pac17. These crude approaches may give the protein access to an unusual cofactor or environment that may have been omitted from the assays performed with purified enzyme. The first was to use a cell free extract from *E. coli* pDT005 (the Pac17 expression construct). The second approach was to incubate purified Pac17 with a *S. coeruleorubidus* extract. The detection method was the same as previously described. Neither of these approaches appeared to help in observing enzyme activity, however, due to the complexity of the mixtures, there is a possibility that the activity could have been missed.

6.3 Pac19

6.3.1 Introduction

Pac19 (DABA synthase), has been shown via genetic analysis to be essential for the formation of pacidamycin (Rackham et al., 2010). Work into the *in vitro* activity of DABA biosynthetic enzymes in the mureidomycin producer, *S. flavidovirens* had determined that DABA biosynthesis proceeds by an ammonia-dependent β -replacement reaction, rather than the other hypothesised pathway of oxidation of threonine to a 3-keto-2-aminobutyric acid followed by a transamination event (Lam et al., 2008). Lam et al. (2008) used a stopped colorimetric assay to determine formation of DABA, whereby addition of phenylglyoxal at an acidic pH derivatises diamino acids such as DABA to form a 2,3-dihydropyrazine derivative that can be measured at a wavelength of 345 nm.

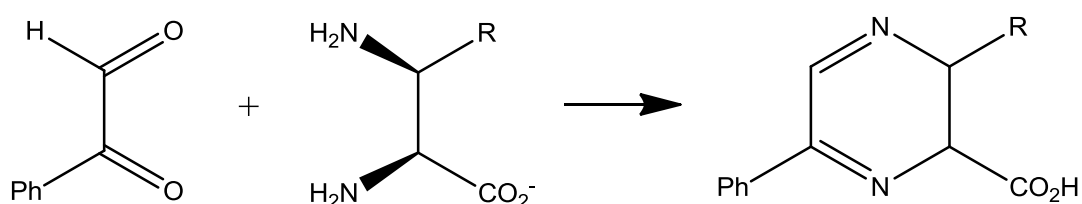


Figure 6.2 The derivatisation of DABA with phenylglyoxal forms a 2,3-dihydropyrazine derivative that can be measured at 345 nm. R = CH₃ in DABA and H in DAP

6.3.2 Results of the investigation into the activity of Pac19

To determine whether using the derivatisation of DABA method previously reported was viable, a small quantity of DABA was obtained from A. Fayad (UEA) and the derivatisation undertaken as described in chapter 2 (Lam et al., 2008, Tripier et al., 2003). The derivatisation appeared to be successful, as indicated by the appearance of a reddish colour. Unfortunately, no more DABA was readily available to confirm the derivatisation.

DAP is structurally similar to DABA, containing one less methyl group. As DAP is readily available from commercial sources, it was decided that a standard curve could be produced using DAP, which could then be used as a reference for the subsequent experiments into the activity of Pac19. A standard curve was produced by using set concentrations of DAP and derivatising DAP with the phenylglyoxal solution as outlined in chapter 2. The resulting standard curve is shown in Figure 6.3.

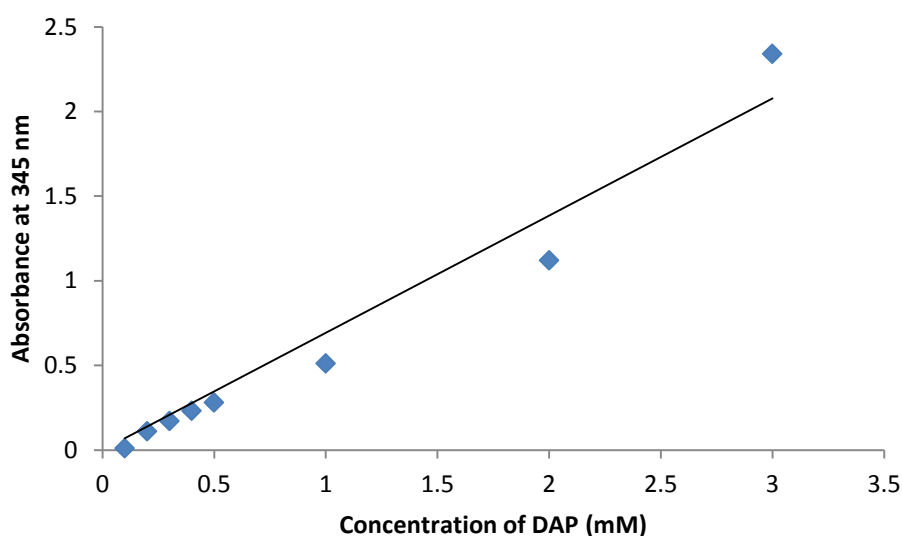


Figure 6.3 Standard curve produced by derivatising known concentrations of DAP

To determine activity of Pac19, assays were set up as outlined in chapter 2 (Lam et al., 2008). Initially the assays were carried out exactly as reported by Lam et al. (2008), however, no activity was seen after numerous attempts. As a result, a number of modifications were made to the method, including using L-allo-threonine and O-phospho-L-threonine as starting substrates, along with different sources of ammonia,

Chapter 6 – In Vitro Studies of the DABA Biosynthesis Genes

including; ammonium bicarbonate, ammonium chloride and ammonium acetate. A range of pH values were also tested between pH 6.0 and 8.0, along with the addition of PLP as a cofactor (PLP being postulated to be essential for the reaction) (Lam et al., 2008). Unfortunately DABA formation was not detected in any of these experiments.

To determine whether any turnover had occurred, an assay solution that had been incubated overnight at 30 °C was submitted to the Metabolomics Facility (JIC) for analysis. This found no evidence of DABA in the mixture.

6.3.3 Use of cell free extract and incubation with *S. coeruleorubidus*

Two crude approaches were undertaken in an attempt to obtain *in vitro* activity of Pac19. These crude approaches may give the protein access to an unusual cofactor or environment that may have been omitted from the assays performed with purified enzyme. The first was to use a cell free extract from *E. coli* pDT007. The second approach was to incubate purified Pac19 with a *S. coeruleorubidus* extract. Neither of these approaches demonstrated enzyme activity.

6.3.4 Coupling of Pac17 and Pac19

In a final attempt to obtain *in vitro* enzymatic activity from either Pac17 or Pac19, a coupled approach was undertaken. The rationale behind this approach was that as *pac17-19* are translationally coupled, perhaps the expressed proteins need to work together or even form a complex. Purified Pac17 and Pac19 were incubated together for 1 h and a series of assays set up which covered the spectrum of potential substrates, co-factors and pH values previously reported in this chapter. Two measurements were taken from these assays. Initially a continuous assay was undertaken in an attempt to observe the formation or depletion of fumarate (and in turn Pac17 activity). The second was a stopped assay measurement, adding the phenylglyoxal mixture which would derivatise any DABA present and therefore measure the activity of Pac19. Again, no activity was observed with either approach.

6.4 Pac20

It has previously been stated that through bioinformatics analysis that Pac20 is postulated to be a threonine aldolase. Genetic investigation reported in this study showed that within

the native producer, disruption of Pac20 appears to eliminate pacidamycin production. Further to this, heterologous expression studies showed that the gene disruption can be complemented by a gene present in *S. lividans*. A number of candidates for this complementation were found including a threonine aldolase and an allo-threonine aldolase. These findings leave a question mark over the product produced by Pac20 as feeding L-threonine or L-allo-threonine to the native producer did not reinstate pacidamycin production.

An assay for threonine aldolase was previously reported by Karasek et al. (1956), however due to the instability of the purified Pac20 (reported in chapter 3), no assays could be carried out.

6.5 Discussion

In conclusion, all attempts at obtaining *in vitro* activity from any of the proteins from the pacidamycin gene cluster hypothesised to be involved in the biosynthesis of DABA were unsuccessful. As previously discussed, formation of DABA has been reported during studies into the mureidomycin biosynthetic cluster (Lam et al., 2008). This study differed from the *in vitro* studies undertaken in this investigation in that Lam et al. (2008) used a cell free extract from *S. flavidovirens*. However, this method did not determine what was present in the active fraction used. This raises a number of questions about their study, for example - was the activity observed attributed to a single protein (such as a homolog of Pac19) or a collection of proteins? Further to this, due to the method used by Lam *et al.* (2008), the exact environmental conditions of the reaction are not known, for example, what the optimal pH is or whether there were co-factors involved (other than additives that they showed to increase DABA production, such as PLP). As previously discussed, activity for a structural homolog of Pac17 has been reported. From the structural analysis it was determined that Pac17 is likely to have a similar catalytic mechanism to that of the structural homolog, however, attempts to reproduce the method used by Weiner et al. (2008) with Pac17 gave no activity.

There are a number of explanations to why the *in vitro* study of the DABA biosynthetic proteins of pacidamycin failed. The fact that Pac17 through 19 are translationally coupled could suggest that the three proteins require each other for activity. That said, the current evidence from genetic studies suggest that only the disruption of *pac19* appears to

Chapter 6 – In Vitro Studies of the DABA Biosynthesis Genes

eliminate pacidamycin production. Unfortunately, as reported in chapter 3, this study was unsuccessful in obtaining stable and purified Pac18 (the middle protein encoded for on the *pac17-19* mRNA). The successful production of this protein may be the key to unlocking the activity of either Pac17 or Pac19. Further to this, the substrates used in these studies may not have been the natural substrates, for example, aspartyl-CoA may be the substrate of Pac17 as this would allow for a more stable intermediate during its breakdown. This hypothesis was not tested due to aspartyl-CoA not being commercially available.

A number of new approaches could be used in an attempt to observe *in vitro* activity of the DABA biosynthetic proteins of pacidamycin. One approach would be to assume that there is some activity taking place but at such a low rate that the current methodology for detection is not sensitive enough. An assay using radiolabelled substrate would enable low levels of turnover to be detected. Another approach would be to re-clone the enzymes, perhaps in a different expression system or with different affinity tags. This approach may allow for Pac18 to be expressed in a soluble form and improve the stability of Pac20, for example. Furthermore, the removal of affinity tags may be necessary as they could be hindering activity. This approach could also allow for the co-expression of the translationally coupled genes *pac17-19*, which may be important to the activity of the proteins. Another approach would be to express the proteins in a *Streptomyces* host. This may improve the stability of the proteins, as it would reduce issues with codon bias, and also supply the proteins with *Streptomyces* specific chaperones and cofactors that may also be important to protein stability and activity. Finally, new methodologies to observing activity could be explored, for example, using coupled assays. This approach may remove any form of product inhibition that may be occurring, by converting the products of the DABA enzymes into other products. This method would also allow for other means of detecting activity, for example the reduction of NAD⁺.

General Discussion and Conclusions

Chapter 7 – General Discussion and Conclusions

The work presented in this thesis describes a series of investigations into 1. The function of the hypothetical protein genes found in the pacidamycin gene cluster and 2. The biosynthesis of the central DABA residue of the pacidamycin structure. This work took a multidisciplinary approach, tackling the questions at a biochemical, genetic and structural level.

Pacidamycin is an uridyl peptide antibiotic which is produced by *S. coeruleorubidus*. The antibiotic shows specific activity against the Gram negative pathogen *Pseudomonas aeruginosa* with its mode of action reported as acting as an inhibitor of the cell wall biosynthetic enzyme translocase I (Fernandes et al., 1989). The pacidamycins have an intriguing structure which contains a tetra- or pentapeptide backbone that is inverted twice by the presence of a non-proteinogenic diamino acid (DABA) and a urea motif. They also contain an uridyl nucleoside derivative which gives this class of antibiotic its name.

At the beginning of this study, there were four hypothetical protein genes annotated within the 22 ORF gene cluster of pacidamycin, produced by *S. coeruleorubidus*. These genes were *pac1*, *pac2*, *pac7* and *pac13*. During this investigation, functions have been discovered for two of these (i.e. *pac2* and *pac13*) and a function for a homolog of *pac1* also reported (Zhang et al., 2011, Ragab et al., 2011, Li et al., 2013). Genetic studies reported in this thesis, along with work carried out by A. Ragab (UEA) identified the function of *pac13*, the gene expressing the protein Pac13 which is responsible for a dehydration event that occurs across the furan ring of the uridine nucleoside derivative of pacidamycin (Ragab et al., 2011).

The function of *pac2* was reported by Zhang et al. (2012), the gene being found to be responsible for the incorporation of the glycyl or alanyl moiety at the N-terminus of the pacidamycin peptide backbone. Most recently the function of a homolog of *pac1*, *ssaA* from the sansamycin gene cluster of *Streptomyces* sp. SS, has been found to be a transcriptional regulator of a number of genes in this cluster that are homologous to genes within the pacidamycin gene cluster (Li et al., 2013).

The biosynthesis of DABA is essential for pacidamycin production. A disruption of a synthase found within the pacidamycin gene cluster (*pac19*) had been found to eliminate pacidamycin production in the native producer *S. coeruleorubidus* (Rackham et al., 2010). Further to this, it was postulated that the biosynthesis of DABA was the responsibility of four genes found in the pacidamycin gene cluster: *pac17* through *20*. These genes were

Chapter 7 – General Discussion and Conclusions

annotated to be a lyase, kinase, synthase and threonine aldolase, respectively, based on sequence alone. Moreover, *pac17* through *19* were shown to be translationally coupled. The postulated mechanism for the biosynthesis of DABA by these four genes is shown in Figure 7.1.

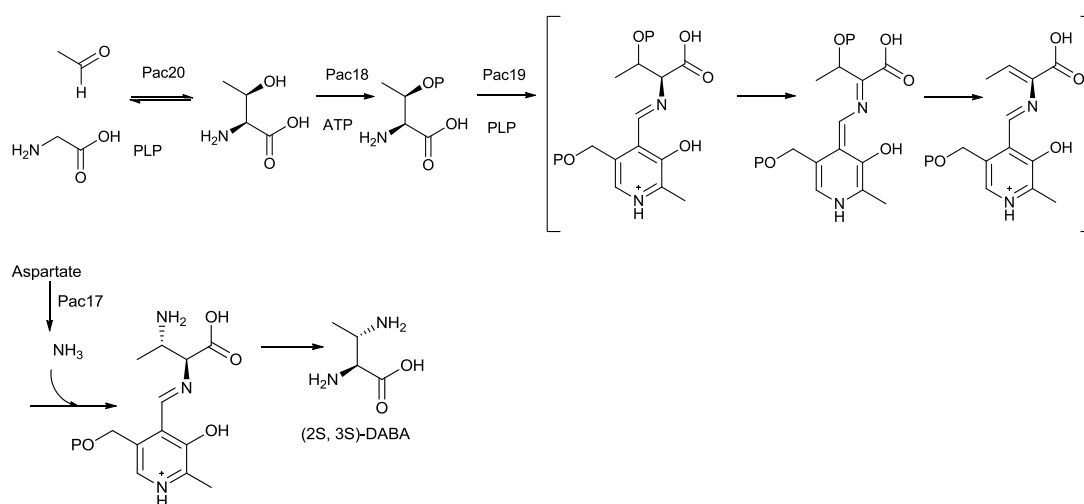


Figure 7.1 The postulated biosynthesis of DABA by Pac17 through 20, four proteins encoded for in the pacidamycin gene cluster.

Gene disruption studies showed that *pac19* and *pac20* were essential for the biosynthesis of pacidamycin in *S. coeruleorubidus*, however the disruption in *pac20* in the heterologous host, *S. lividans* TK24, appeared to be complemented by the organism. BLAST analysis of the homologous *S. coelicolor* genome suggested that there were two genes; a threonine aldolase and a L-allo-threonine aldolase, that were likely candidates for this complementation event (Altschul et al., 1990). Attempts to chemically complement these disruptions in *S. coeruleorubidus* with L-DAP for the *pac19* disruption, and L-threonine and L-allo-threonine for the *pac20* disruption, failed to reinstate pacidamycin production.

Gene disruption of *pac17* and *pac18* was carried out in the heterologous host *S. lividans* TK24 due to the likelihood of polar effects in the native producer. In both cases, the disruptions did not eliminate pacidamycin production, however, pacidamycin production appeared to be drastically reduced in the disruption of *pac18*. Bioinformatics analysis suggested that *pac17* was likely to be readily complemented by a number of candidates, and there were also potential candidates for the complementation of the *pac18* disruption.

Chapter 7 – General Discussion and Conclusions

Further to this, *pac18* is a kinase, and kinases are often used in biosynthetic pathways to produce substrates with better leaving groups (i.e. phosphates) for downstream reactions. The decrease in pacidamycin production seen in the *pac18* disrupted strain could therefore be due to the turnover of the non-phosphorylated form of the substrate.

The conclusions drawn from the genetic work which is reported in this thesis suggest that both *pac19* and *pac20* are essential in the native pacidamycin producer, however *pac20* is complemented in some way in the heterologous host *S. lividans*. This investigation also suggests that the roles of *pac17* and *pac18* are likely to be in maximizing the pathway throughput, whether by increasing the availability of a substrate for a particular step (i.e. Pac17) or by producing a substrate that has a better leaving group for the downstream reaction (ie. Pac18).

Pac17, Pac19 and Pac20 were successfully produced in an *E. coli* BL21 and purified using affinity chromatography. Pac20, however, was found by DLS analysis to be unstable. Further to this, attempts to overproduce Pac18 resulted in mainly insoluble protein.

Structural studies carried out in this investigation resulted in the determination of the structure of Pac17 to a resolution of 1.8 Å, using molecular replacement. Pac17, in its crystallographic form exists as a tetramer. Each monomer consists of 22 α-helices and 8 short β-strands. Along with the apo structure, a ligand-bound structure was determined by co-crystallisation of the protein with aspartate. This identified the active site of the protein. The aspartate bound Pac17 structure was found to show similarities to that of the protein AspB from *B. subtilis* (Fibriansah et al., 2011). Further structural studies found that aspartate was the likely substrate of Pac17 as structurally similar compounds, such as asparagine, glutamate and glutamine, were not found bound to the protein in equivalent co-crystallisation studies.

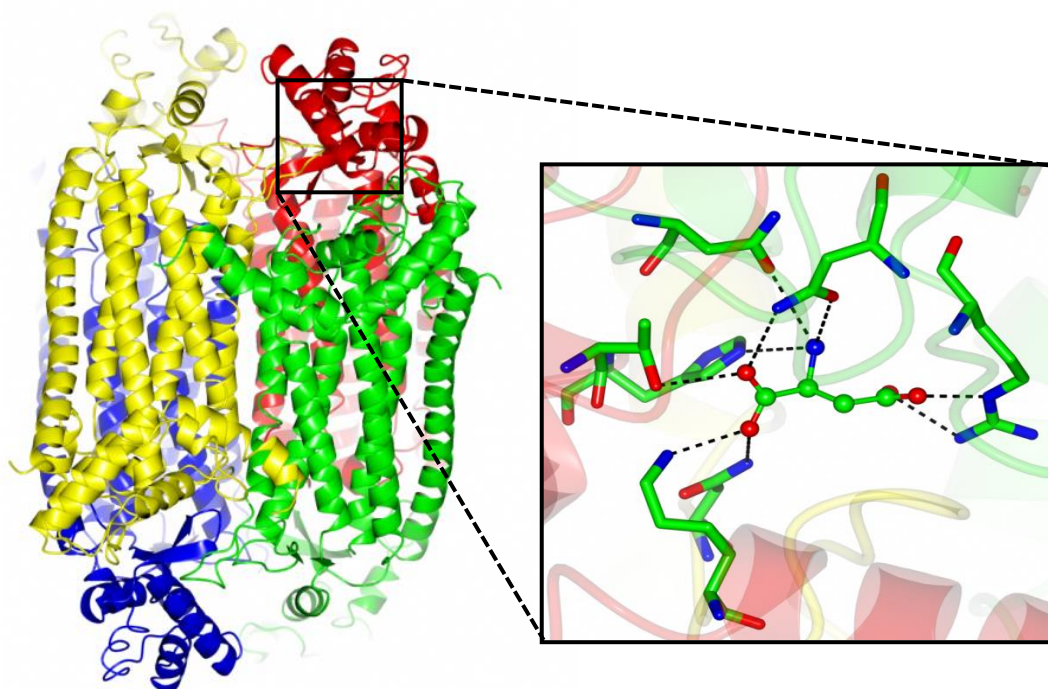


Figure 7.2 The structure of Pac17 and the active site of Pac17 with the ligand aspartate bound

Using site direct mutagenesis on amino acids within the active site, it was shown by further co-crystallisation studies that Asn109, Met279, Lys282 and Asn284 were also essential for ligand binding, these results being consistent with observations reported by Fibriansah et al. (2012).

This thesis describes attempts to observe *in vitro* activity for a number of the proteins implicated in the biosynthesis of DABA (namely Pac17 and Pac19). Unfortunately, although numerous attempts were made, no activity was seen.

In conclusion, this thesis reports a variety of techniques used in an attempt to elucidate the biosynthesis of DABA in *S. coeruleorubidus* and also investigations into the function of a number of hypothetical proteins annotated as being present in the pacidamycin gene cluster. Even though good progress has been achieved during this study, there is still further work that could be undertaken in the future to complete the story.

Further work needs to be carried out on producing more stable Pac20, along with producing soluble Pac18. Approaches that could be considered to tackle this include

Chapter 7 – General Discussion and Conclusions

truncating the proteins of interest to try and remove any unstable regions at the N- or C-termini of the proteins. Further to this, investigations into different protein expression systems, along with different affinity tags may be a beneficial approach to this problem. As *pac17* through *pac19* are translationally coupled, the co-expression of these genes may help in acquiring active protein and soluble Pac18. Further to this, the removal of the affinity tags may help to produce active proteins.

Genetic studies reported in this thesis indicate that *pac19* is essential for the production of pacidamycin and *pac20* is also essential in the native producer. These studies also reveal the function for one of the hypothetical proteins (*pac13*). This study also reported that a disruption of *pac7*, another postulated hypothetical protein, eliminates pacidamycin production. More work is necessary to determine the actual function for this gene being discovered. Further analysis of the DABA biosynthesis genes could shed more light on DABA biosynthesis. Firstly, as chemical complementation did not appear to reinstate DABA biosynthesis, it would be sensible to carry out genetic complementation of the *pac19* and *pac20* disruption to ensure that the chemotype observed is attributed to the expected disruption and is not caused by other effects, such as polarity on other gene functions. The development of an approach to study the disruption of *pac17* and *pac18* in *S. coeruleorubidus* may also be a sensible next step in this study. Observations in the heterologous host, *S. lividans* TK24, showed that neither disruption eliminated pacidamycin production. However, the disruption of *pac18* reduced production in comparison to the heterologous expression of the non-disrupted pacidamycin gene cluster. Since the chemotype of the *pac20* disruption differed between the native producer and heterologous host, this could also be happening in the *pac17* and *pac18* disruptions. Therefore a determination of function of these genes may only be possible in *S. coeruleorubidus*.

Attempts to observe activity of a number of the proteins believed to be involved in the biosynthesis of DABA reported in this thesis were unsuccessful. A number of explanations for this were discussed including the possibility of missing cofactors and that the translational coupling of *pac17-19* may indicate that all three proteins need to be present together for activity to be observed. Future work in this area may benefit from the use of different expression systems and purification protocols, as described above. Moreover, alternative assays could be used, perhaps using radiolabelled substrates.

Chapter 7 – General Discussion and Conclusions

To summarise, this thesis has reported a number of findings into aspects of the biosynthesis of the antibiotic pacidamycin, produced by the soil bacterium *S. coeruleorubidus*. In collaboration with A. Ragab (UEA), a function for *pac13* has been identified. Efforts to understand the biosynthesis of DABA have been discussed and an *in vivo* and *in vitro* approach taken. Furthermore, the structure of apo- and ligand-bound Pac17 has been reported.

The findings reported in this thesis, along with work carried out by others on the pacidamycin gene cluster and the gene clusters of over uridyl peptides, offer a wealth of knowledge into the biosynthesis of these compounds. A better understanding of these natural products may one day be fundamental in the fight against bacterial infection from *Pseudomonas aeruginosa*.

References

Al-Dabbagh, B., Henry, X., El Ghachi, M., Auger, G., Blanot, D., Parquet, C., Mengin-Lecreulx, D. and Bouhss, A. (2008) 'Active site mapping of MraY, a member of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily, catalyzing the first membrane step of peptidoglycan biosynthesis', *Biochemistry*, 47(34), 8919-8928.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), 403-10.

Asaduzzaman, S. M. and Sonomoto, K. (2009) 'Lantibiotics: diverse activities and unique modes of action', *Journal of Bioscience Bioengineering*, 107(5), 475-87.

Beasley, F. C., Cheung, J. and Heinrichs, D. E. (2011) 'Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*', *Bmc Microbiology*, 11.

Belshaw, P. J., Walsh, C. T. and Stachelhaus, T. (1999) 'Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis', *Science*, 284(5413), 486-489.

Bergendahl, V., Linne, U. and Marahiel, M. A. (2002) 'Mutational analysis of the C-domain in nonribosomal peptide synthesis', *European Journal of Biochemistry*, 269(2), 620-629.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) 'The Protein Data Bank', *Nucleic Acids Research*, 28(1), 235-42.

Bibb, M. J. (2005) 'Regulation of secondary metabolism in streptomycetes', *Current Opinions in Microbiology*, 8(2), 208-15.

Bouhss, A., Crouvoisier, M., Blanot, D. and Mengin-Lecreulx, D. (2004) 'Purification and characterization of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan biosynthesis', *Journal of Biological Chemistry*, 279(29), 29974-29980.

References

Bouhss, A., Mengin-Lecreulx, D., Le Beller, D. and van Heijenoort, J. (1999) 'Topological analysis of the MraY protein catalysing the first membrane step of peptidoglycan synthesis', *Molecular Microbiology*, 34(3), 576-585.

Bradford, M. M. (1986) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding', *Analytical Biochemistry*, 72, 248-254.

Brandish, P. E., Kimura, K., Inukai, M., Southgate, R., Lonsdale, J. T. and Bugg, T. D. H. (1996) 'Modes of action of tunicamycin, liposidomycin B, and mureidomycin A: Inhibition of phospho-N-acetylmuramyl-pentapeptide translocase from *Escherichia coli*', *Antimicrobial Agents and Chemotherapy*, 40(7), 1640-1644.

Brünger, A. T. (1993) 'Assessment of phase accuracy by cross validation: the free R value. Methods and applications', *Acta Crystallographica D Biological Crystallography*, 49(Pt 1), 24-36.

Bugg, T. D. H. and Walsh, C. T. (1992) 'Intracellular steps of bacterial-cell wall peptidoglycan biosynthesis - enzymology, antibiotics, and antibiotic-resistance', *Natural Product Reports*, 9(3), 199-215.

Bupp, K. and Vanheijenoort, J. (1993) 'The final step of peptidoglycan subunit assembly in *Escherichia coli* occurs in the cytoplasm', *Journal of Bacteriology*, 175(6), 1841-1843.

Carter, J. H., Dubus, R. H., Dyer, J. R., Floyd, J. C., Rice, K. C. and Shaw, P. D. (1974) 'Biosynthesis of viomycin. 1. origin of alpha, beta-diaminopropionic acid and serine', *Biochemistry*, 13(6), 1221-1227.

Castiglione, F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E. and Parenti, F. (2008) 'Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens', *Chemical Biology*, 15(1), 22-31.

References

Challis, G. L. and Naismith, J. H. (2004) 'Structural aspects of non-ribosomal peptide biosynthesis', *Current Opinions in Structural Biology*, 14(6), 748-56.

Chatterjee, S., Nadkarni, S. R., Vijayakumar, E. K. S., Patel, M. V., Ganguli, B. N., Fehlhaber, H. W. and Vertesy, L. (1994) 'Napsamycins, new Pseudomonas active antibiotics of the mureidomycin family from Streptomyces sp HIL Y-82,11372', *Journal of Antibiotics*, 47(5), 595-598.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallographica D Biological Crystallography*, 66(Pt 1), 12-21.

Chen, R. H., Buko, A. M., Whittern, D. N. and McAlpine, J. B. (1989) 'Pacidamycins, a novel series of antibiotics with anti-Pseudomonas aeruginosa activity. 2. isolation and structural elucidation', *Journal of Antibiotics*, 42(4), 512-520.

Chenna, R., Sugawara, T., Koike, R., Lopez, T. J., Gibson, D. G., Higgins and J. D. Thompson (2003) 'Multiple sequence alignment with the Clustal series of programs', *Nucleic Acids Research*, 31(13), 3497-3500.

Cirz, R. T., Chin, J. K., Andes, D. R., de Crecy-Lagard, V., Craig, W. A. and Romesberg, F. E. (2005) 'Inhibition of mutation and combating the evolution of antibiotic resistance', *Plos Biology*, 3(6), 1024-1033.

Clardy, J. and Walsh, C. (2004) 'Lessons from natural molecules', *Nature*, 432(7019), 829-837.

Cowtan, K. (2006) 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta Crystallographica D Biological Crystallography*, 62(Pt 9), 1002-11.

Desbois, A. P., Mearns-Spragg, A. and Smith, V. J. (2009) 'A Fatty Acid from the Diatom Phaeodactylum tricornutum is Antibacterial Against Diverse Bacteria Including Multi-resistant Staphylococcus aureus (MRSA)', *Marine Biotechnology*, 11(1), 45-52.

References

Dewick, P. M. (2003) 'Medicinal Natural Products (Second Edition)',

Dieckmann, R., Lee, Y. O., Vanliempt, H., Vondohren, H. and Kleinkauf, H. (1995) 'Expression of an active adenylate-forming domain of peptide synthetases corresponding to acyl-CoA-synthetases', *Febs Letters*, 357(2), 212-216.

Doering, G. and Pier, G. B. (2008) 'Vaccines and immunotherapy against *Pseudomonas aeruginosa*', *Vaccine*, 26(8), 1011-1024.

Driscoll, J. A., Brody, S. L. and Kollef, M. H. (2007) 'The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections', *Drugs*, 67(3), 351-368.

Eckardt, K. (1983) 'Tunicamycins, Streptovirudins, and corynetoxins, a special subclass of nucleoside antibiotics', *Journal of Natural Products*, 46(4), 544-550.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta Crystallographica D Biological Crystallography*, 60(Pt 12 Pt 1), 2126-32.

Evans, P. (2006) 'Scaling and assessment of data quality', *Acta Crystallographica D Biological Crystallography*, 62(Pt 1), 72-82.

Fan, C., Moews, P. C., Walsh, C. T. and Knox, J. R. (1994) 'Vancomycin resistance: structure of D-alanine:D-alanine ligase at 2.3 Å resolution', *Science*, 266(5184), 439-43.

Fernandes, P. B., Swanson, R. N., Hardy, D. J., Hanson, C. W., Coen, L., Rasmussen, R. R. and Chen, R. H. (1989) 'Pacidamycins, a novel series of antibiotics with anti-*pseudomonas aeruginosa* activity. 3. Microbiologic profile', *Journal of Antibiotics*, 42(4), 521-526.

Fibriansah, G., Veetil, V. P., Poelarends, G. J. and Thunnissen, A. M. (2011) 'Structural basis for the catalytic mechanism of aspartate ammonia lyase', *Biochemistry*, 50(27), 6053-62.

References

Fischbach, M. A. and Walsh, C. T. (2009) 'Antibiotics for emerging pathogens', *Science*, 325(5944), 1089-93.

Foulston, L. C. and Bibb, M. J. (2010) 'Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes', *Proceedings of the National Academy of Sciences U S A*, 107(30), 13461-6.

Friedmann, I. (1948) 'Staphylococcal infection due to penicillin-resistant strains', *British Medical Journal*, 1(4539), 27.

Fronko, R. M., Lee, J. C., Galazzo, J. G., Chamberland, S., Malouin, F. and Lee, M. D. (2000) 'New pacidamycins produced by *Streptomyces coeruleorubidus*, NRRL 18370', *Journal of Antibiotics*, 53(12), 1405-1410.

Gentle, C. A. and Bugg, T. D. H. (1999) 'Role of the enamide linkage of nucleoside antibiotic mureidomycin A: synthesis and reactivity of enamide-containing analogues', *Journal of the Chemical Society-Perkin Transactions*, 1(10), 1279-1285.

Gruschow, S., Rackham, E. J. and Goss, R. J. M. (2011) 'Diversity in natural product families is governed by more than enzyme promiscuity alone: establishing control of the pacidamycin portfolio', *Chemical Science*, 2(11), 2182-2186.

Gust, B., Challis, G. L., Fowler, K., Kieser, T. and Chater, K. F. (2003) 'PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin', *Proceedings of the National Academy of Sciences U S A*, 100(4), 1541-6.

Hanson, J. R. (2003) 'Natural Products: the secondary Metabolites', Cambridge, UK, The Royal Society of Chemistry.

Heise, E. R. (1982) 'Diseases associated with immunosuppression', *Environmental Health Perspective*, 43, 9-19.

References

Hollstein, U. (1974) 'Actinomycin. Chemistry and mechanism of action', *Chemical Reviews*, 74(6), 625.

Howard, N. I. and Bugg, T. D. H. (2003) 'Synthesis and activity of 5'-uridylyl dipeptide analogues mimicking the amino terminal peptide chain of nucleoside antibiotic mureidomycin A', *Bioorganic & Medicinal Chemistry*, 11(14), 3083-3099.

Inukai, M., Isono, F. and Takatsuki, A. (1993) 'Selective inhibition of the bacterial translocase reaction in peptidoglycan synthesis by the mureidomycins', *Antimicrobial Agents and Chemotherapy*, 37(5), 980-983.

Isono, F., Inukai, M., Takahashi, S., Haneishi, T., Kinoshita, T. and Kuwano, H. (1989) 'Mureidomycins A-D, novel peptidyl nucleoside antibiotics with spheroplast forming activity. II. Structural elucidation', *Journal of Antibiotics (Tokyo)*, 42(5), 667-73.

Jones, D. T. (1999) 'Protein secondary structure prediction based on position-specific scoring matrices', *Journal of Molecular Biology*, 292(2), 195-202.

Kabsch, W. (2010) 'XDS', *Acta Crystallographica D Biological Crystallography*, 66(Pt 2), 125-32.

Karwowski, J. P., Jackson, M., Theriault, R. J., Chen, R. H., Barlow, G. J. and Maus, M. L. (1989) 'Pacidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. 1. taxonomy of the producing organism and fermentation', *Journal of Antibiotics*, 42(4), 506-511.

Kaysser, L., Tang, X., Wemakor, E., Sedding, K., Hennig, S., Siebenberg, S. and Gust, B. (2011) 'Identification of a Napsamycin Biosynthesis Gene Cluster by Genome Mining', *Chembiochem*, 12(3), 477-487.

Kelley, L. A. and Sternberg, M. J. (2009) 'Protein structure prediction on the Web: a case study using the Phyre server', *Nature Protocols*, 4(3), 363-71.

References

Kerr, K. G. and Snelling, A. M. (2009) 'Pseudomonas aeruginosa: a formidable and ever-present adversary', *Journal of Hospital Infection*, 73(4), 338-344.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. and Hopwood, D. A. (2000) *Practical Streptomyces Genetics*, Norwich, UK: John Innes Foundation.

Kimura, K. and Bugg, T. D. H. (2003) 'Recent advances in antimicrobial nucleoside antibiotics targeting cell wall biosynthesis', *Natural Product Reports*, 20(2), 252-273.

Kleywegt, G. J. and Brünger, A. T. (1996) 'Checking your imagination: applications of the free R value', *Structure*, 4(8), 897-904.

Krissinel, E. and Henrick, K. (2004) 'Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions', *Acta Crystallographica D Biological Crystallography*, 60(Pt 12 Pt 1), 2256-68.

Lam, W.-H., Rychli, K. and Bugg, T. D. H. (2008) 'Cation of a novel beta-replacement reaction in the biosynthesis of 2,3-diaminobutyric acid in peptidyl nucleoside mureidomycin A', *Organic & Biomolecular Chemistry*, 6(11), 1912-1917.

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', *Chemistry & Biology*, 3(11), 923-936.

Lambert, P. A. (2002) 'Mechanisms of antibiotic resistance in Pseudomonas aeruginosa', *Journal of the Royal Society of Medicine*, 95 Suppl 41, 22-6.

Laskowski, R. A. (2001) 'PDBsum: summaries and analyses of PDB structures', *Nucleic Acids Res*, 29(1), 221-2.

References

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L. and Thornton, J. M. (1997) 'PDBsum: a Web-based database of summaries and analyses of all PDB structures', *Trends in Biochemical Sciences*, 22(12), 488-90.

Lautru, S. and Challis, G. L. (2004) 'Substrate recognition by nonribosomal peptide synthetase multi-enzymes', *Microbiology*, 150(Pt 6), 1629-36.

Leblond, P., Redenbach, M. and Cullum, J. (1993) 'Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2)', *Journal of Bacteriology*, 175(11), 3422-9

Lee, D. G., Urbach, J. M., Wu, G., Liberati, N. T., Feinbaum, R. L., Miyata, S., Diggins, L. T., He, J., Saucier, M., Deziel, E., Friedman, L., Li, L., Grills, G., Montgomery, K., Kucherlapati, R., Rahme, L. G. and Ausubel, F. M. (2006) 'Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial', *Genome Biology*, 7(10), R90.

Leslie, A. G. W. and Powell, H. R. (2007) 'Processing Diffraction Data with Mosflm', *Evolving Methods for Macromolecular Crystallography*, 245, 41-51.

Li, Q., Wang, L., Xie, Y., Wang, S., Chen, R. and Hong, B. (2013) 'SsaA, a novel class of transcriptional regulator, controls sansanmycin production in *Streptomyces* sp. SS involving a feedback mechanism', *Journal of Bacteriology*.

Li, X. Z., Livermore, D. M. and Nikaido, H. (1994) 'Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: resistance to tetracycline, chloramphenicol, and norfloxacin', *Antimicrobial Agents and Chemotherapy*, 38(8), 1732-41.

Livermore, D. M. (2002) 'Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: Our worst nightmare?', *Clinical Infectious Diseases*, 34(5), 634-640.

Lyczak, J. B., Cannon, C. L. and Pier, G. B. (2000) 'Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist', *Microbes and Infection*, 2(9), 1051-1060.

References

- Madigan, M. and Martinko, J. (2005) 'Biology of Microorganisms (Eleventh Edition)',
- Marahiel, M. A., Stachelhaus, T. and Mootz, H. D. (1997) 'Modular peptide synthetases involved in nonribosomal peptide synthesis', *Chemical Reviews*, 97(7), 2651-2673.
- Marinelli, F. (2009) 'Chapter 2. From microbial products to novel drugs that target a multitude of disease indications', *Methods in Enzymology*, 458, 29-58.
- McCoy, A. J. (2007a) 'Solving structures of protein complexes by molecular replacement with Phaser', *Acta Crystallographica D Biological Crystallography*, 63(Pt 1), 32-41.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007b) 'Phaser crystallographic software', *Journal of Applied Crystallography*, 40(Pt 4), 658-674.
- McGilvray, D. and Morris J. G. (1969) 'Utilisation of L-threonine by a species of *Arthrobacter*. A novel catalytic role for 'aminoacetone synthase', *Journal of Biochemistry*, 112(5), 657-671
- Mohrig, J.R. (2012) 'Stereochemistry of 1,2-Elimination and Proton-Transfer Reactions: Towards a Unified Understanding', *Accounts of Chemical Research*.
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) 'Refinement of macromolecular structures by the maximum-likelihood method', *Acta Crystallographica D Biological Crystallography*, 53(Pt 3), 240-55.
- Newman, D. J., Cragg, G. M. and Snader, K. M. (2000) 'The influence of natural products upon drug discovery', *Natural Product Reports*, 17(3), 215-234.
- Newman, D. J. and Cragg, G. M. (2012) 'Natural products as sources of new drugs over the 30 years from 1981 to 2010', *Journal of Natural Products*, 75(3), 311-35.

References

Nishimura, H., Katagiri, K., Sato, K., Mayama, M. and Shimaoka, N. (1956) 'Toyocamycin, a new anti-Candida antibiotic', *Journal of Antibiotics*, 9(2), 60-62.

Okamoto, K., Sakagami, M., Feng, F., Takahashi, F., Uotani, K., Togame, H., Takemoto, H., Ichikawa, S. and Matsuda, A. (2012a) 'Synthesis of pacidamycin analogues via an Ugi-multicomponent reaction', *Bioorganic and Medicinal Chemistry Letters*, 22(14), 4810-5.

Okamoto, K., Sakagami, M., Feng, F., Togame, H., Takemoto, H., Ichikawa, S. and Matsuda, A. (2011) 'Total synthesis of pacidamycin D by Cu(I)-catalyzed oxy enamide formation', *Organic Letters*, 13(19), 5240-3.

Okamoto, K., Sakagami, M., Feng, F., Togame, H., Takemoto, H., Ichikawa, S. and Matsuda, A. (2012b) 'Total synthesis and biological evaluation of pacidamycin D and its 3'-hydroxy analogue', *Journal of Organic Chemistry*, 77(3), 1367-77.

Overbye, K. M. and Barrett, J. F. (2005) 'Antibiotics: where did we go wrong', *Drug Discovery Today*, 10(1), 45-52.

Payne, D. J., Gwynn, M. N., Holmes, D. J. and Pompliano, D. L. (2007) 'Drugs for bad bugs: confronting the challenges of antibacterial discovery', *Nature Reviews Drug Discovery*, 6(1), 29-40.

Perrakis, A., Harkiolaki, M., Wilson, K. S. and Lamzin, V. S. (2001) 'ARP/wARP and molecular replacement', *Acta Crystallographica D Biological Crystallography*, 57(Pt 10), 1445-50.

Pollard, T. D. and Earnshaw, W. C. (2008) 'Cell Biology (Second Edition)'.

Price, N. P. and Tsvetanova, B. (2007) 'Biosynthesis of the tunicamycins: a review', *Journal of Antibiotics (Tokyo)*, 60(8), 485-91.

References

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. and Sussman, J. L. (2005) 'FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded', *Bioinformatics*, 21(16), 3435-8.

Rackham, E. J. (2010) *Elucidation of pacidamycin biosynthesis*, thesis submitted to the University of East Anglia.

Rackham, E. J., Grueschow, S., Ragab, A. E., Dickens, S. and Goss, R. J. M. (2010) 'Pacidamycin Biosynthesis: Identification and Heterologous Expression of the First Uridyl Peptide Antibiotic Gene Cluster', *Chembiochem*, 11(12), 1700-1709.

Ragab, A. E., Gruschow, S., Tromans, D. R. and Goss, R. J. M. (2011) 'Biogenesis of the Unique 4',5'-Dehydronucleoside of the Uridyl Peptide Antibiotic Pacidamycin', *Journal of the American Chemical Society*, 133(39), 15288-15291.

Ramachandran, G. N., Ramakrishnan, C. and Saisekharan, V. (1963) 'Stereochemistry of polypeptide chain configurations', *Journal of Molecular Biology*, 7, 95-9.

Saiman, L. (2004) 'Microbiology of early CF lung disease', *Paediatric Respiratory Reviews*, 5(SupA), S367-9.

Sambrook, J., MacCallum, P. and Russell, D. (2001) 'Molecular cloning: a laboratory manual (Third edition).

Sampaleanu, L. M., Coddig, P. W., Lobsanov, Y. D., Tsai, M., Smith, G. D., Horvatin, C. and Howell, P. L. (2004) 'Structural studies of duck delta2 crystallin mutants provide insight into the role of Thr161 and the 280s loop in catalysis', *Biochemical Journal*, 384(Pt 2), 437-47.

Sang, Y. and Blecha, F. (2008) 'Antimicrobial peptides and bacteriocins: alternatives to traditional antibiotics', *Animal Health Research Reviews*, 9(2), 227-35.

References

Schneider, T. and Sahl, H. G. (2010) 'An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway', *International Journal of Medical Microbiology*, 300(2-3), 161-9.

Schwarzer, D., Finking, R. and Marahiel, M. A. (2003) 'Nonribosomal peptides: from genes to products', *Natural Product Reports*, 20(3), 275-287.

Sears, C. L. (2006) 'A dynamic partnership: Celebrating our gut flora (vol 11, pg 247, 2005)', *Anaerobe*, 12(2), 114-114.

Slabinski, L., Jaroszewski, L., Rodrigues, A. P. C., Rychlewski, L., Wilson, I. A., Lesley, S. A. and Godzik, A. (2007a) 'The challenge of protein structure determination - lessons from structural genomics', *Protein Science*, 16(11), 2472-2482.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I. A., Lesley, S. A. and Godzik, A. (2007b) 'XtalPred: a web server for prediction of protein crystallizability', *Bioinformatics*, 23(24), 3403-5.

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', *Journal of Biological Chemistry*, 273(35), 22773-22781.

Stein, N. (2008) 'CHAINSAW: a program for mutating pdb files used as templates in molecular replacement.', *Journal of Applied Crystallography*, 41, 641.

Stein, T., Vater, J., Kruff, V., Otto, A., WittmannLiebold, B., Franke, P., Panico, M., McDowell, R. and Morris, H. R. (1996) 'The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates', *Journal of Biological Chemistry*, 271(26), 15428-15435.

Stone, M. J. and Williams, D. H. (1992) 'On the evolution of functional secondary metabolites (natural products)', *Molecular Microbiology*, 6(1), 29-34.

Stratagene 'QuikChange Site-Directed Mutagenesis Kit'.

References

Struve, W. G., Sinha, R. K. and Neuhaus, F. C. (1966) 'On initial stage in peptidoglycn synthesis- phospho-N-acetylmutamyl-pentapeptides translocase (uridine monophosphate)', *Biochemistry*, 5(1), 82-&.

N. N. I. S. (2004) 'National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004', *American Journal of Infection Control*, 32(8), 470-85.

Takatsuki, A., Kawamura, K., Okina, M., Kodama, Y., Ito, T. and Tamura, G. (1977) 'Structural elucidation of tunicamycin. 2. Structure of tunicamycin', *Agricultural and Biological Chemistry*, 41(11), 2307-2309.

Tripier, R., Chuburu, F., Le Baccon, M. and Handel, H. (2003) 'Phenylglyoxal for polyamines modifications and cyclam synthesis', *Tetrahedron*, 59, 4573.

Vagin, A. and Teplyakov, A. (1997) 'MOLREP: an automated program for molecular replacement', *Journal of Applied Crystallography.*, 30, 1022 - 1025.

Veetil, V. P., Raj, H., Quax, W. J., Janssen, D. B. and Poelarends, G. J. (2009) 'Site-directed mutagenesis, kinetic and inhibition studies of aspartate ammonia lyase from *Bacillus* sp. YM55-1', *FEBS*, 276, 2994-3007.

Waksman, S. A. and Woodruff, H. B. (1941) 'Actinomyces antibioticus, a New Soil Organism Antagonistic to Pathogenic and Non-pathogenic Bacteria', *Journal of Bacteriology*, 42(2), 231-49.

Walsh, C. T. (2003) 'Antibiotics: actions, origins, resistance'.

Walsh, T. R., Toleman, M. A., Poirel, L. and Nordmann, P. (2005) 'Metallo-beta-lactamases: the quiet before the storm?', *Clinical Microbiology Reviews*, 18(2), 306-+.

References

Wang, L., Xie, Y., Li, Q., He, N., Yao, E., Xu, H., Yu, Y., Chen, R. and Hong, B. (2012) 'Draft genome sequence of *Streptomyces* sp. strain SS, which produces a series of uridyl peptide antibiotic sansanmycins', *Journal of Bacteriology*, 194(24), 6988-9.

Wang, M. and Gould, S. J. (1993) 'Biosynthesis of capreomycin. 2. Incorporation of L-serine, L-alanine and L-2,3-diaminopropionic acid', *Journal of Organic Chemistry*, 58(19), 5176-5180.

Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004) 'Prediction and functional analysis of native disorder in proteins from the three kingdoms of life', *Journal of Molecular Biology*, 337(3), 635-45.

Weiner, B., Poelarends, G. J., Janssen, D. B. and Feringa, B. L. (2008) 'Biocatalytic enantioselective synthesis of N-substituted aspartic acids by aspartate ammonia lyase', *Chemistry*, 14(32), 10094-100.

Winn, M., Goss, R. J. M., Kimura, K.-i. and Bugg, T. D. H. (2010) 'Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis', *Natural Product Reports*, 27(2), 279-304.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) 'Overview of the CCP4 suite and current developments', *Acta Crystallographica D Biological Crystallography*, 67(Pt 4), 235-42.

Woodyer, R. D., Shao, Z., Thomas, P. M., Kelleher, N. L., Blodgett, J. A., Metcalf, W. W., van der Donk, W. A. and Zhao, H. (2006) 'Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster', *Chemical Biology*, 13(11), 1171-82.

Wright, G. D. (2005) 'Bacterial resistance to antibiotics: enzymatic degradation and modification', *Advanced Drug Delivery Reviews*, 57(10), 1451-70.

References

Wright, G. D. and Sutherland, A. D. (2007) 'New strategies for combating multidrug-resistant bacteria', *Trends in Molecular Medicine*, 13(6), 260-267.

Zhang, K. Y., Cowtan, K. and Main, P. (1997) 'Combining constraints for electron-density modification', *Methods in Enzymology*, 277, 53-64.

Zhang, W., Ntai, I., Kelleher, N. L. and Walsh, C. T. (2011) 'tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics', *Proceedings of the National Academy of Sciences USA*, 108(30), 12249-53.

Appendix 1

Appendix 1

Amino acid sequences for the each of the proteins discussed in this thesis.

Hypothetical proteins:

Pac1

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MHTKIEADVE	VIVTWRNKTR	HPGGPVAKER	INLKVGDRLN	FGISSYTGFE	QPGATPSIGA
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
FEVSDHLLF	SNFTTSTTFV	IENLEGGTEL	VKARPRQLGM	VIPFEMSRVL	IPSGASISEL
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
TVFTPPRLL	GPDQAATVAE	SAMTKLDPES	KYFAVLVALC	EPRLRGSSMA	AVPSVQEVVE
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
RLKGIKQFRG	ANRSSINYHI	DYLTQKLPV	SQWAKYVDEG	RMHRSKREALV	AFSLRFDLVR
<u>250</u>					
EEHLGLLPDL	KRSRAQPS				

Pac2

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MAIGFTSAIA	DFDQKQFDAL	DTTAGAASAY	SRLRQHEQDA	RWTSRYLGWF	DGDEVRAAIP
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
VYRYRMRSWP	DPSYDPRSWG	LPDGAIEECS	PRASLMVGGC	IDRRTGFHVD	AEARTPRELQ
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
RLLEIAKHA	ADEDMCLTFP	YMYADAQSAL	AAATDDRIVW	AELAREHLF	GLSDAQWESS
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
LSAKIRYRLR	QDQRKIAAVP	MTVGEVSWPE	VDTWASELIS	HHNASKGAHE	HPEFVSFRYS
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
GWQDNPDIDL	MAFTARSAGL	RGVETILLWE	NELEVYEVGM	TGEESDERFA	LYLNLLFHLP
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	
IQYARARGID	HIRLGSKAET	PKALRGAAFE	NLYGGVLSRA	ETKRLACSES	

Pac7

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MAQVLAEATL	QEIKDHLEV	VDSGRGTTFQ	GSESVVRHLA	EPKELRSLIG	QIISDDTALA
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
DIAARSYYHA	NNFLKVVLLA	GDKNPWRLRL	HMWHPQPNAS	GTITEDIHSH	RWDFTTALVV
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
GEYFAQEFKI	GPGTEYYHFK	YLPYGQKTF	SLEAQGKEQL	SSVFEALLPA	GTVYHINHEV
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
LHCISRSAGK	AAASLVLQPP	AVEDFTNVYR	TSPVGEQTKT	EIEVQRPSVA	QLREELEHFL
<u>250</u>					
TWLD					

Pac13

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MTKYKTEVES	ERFNKHGIDL	TVYGQVDPSA	TVVRVSVERG	HFQEFFNVRS	SYTYYVVSQG
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
GVFYLNSEAV	PAGATDLITV	SMEMVLTVAP	AFNEQDERHV	RFISESESPY	

Appendix 1

DABA biosynthesis proteins:

Pac17

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
VRLTGRLLLL	PDDFIRQEFL	EPQFRHEVAH	LLRWYVLIK	ALLREYHRLG	VLSARQVEQL
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
AGALDGLTPE	ELTAGAGAAL	SDIALAMETR	VAKALPEPVP	AWHVDRSRND	LQACAQLLYG
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
REQVSEIAAM	LGELAEVAHE	RAAHDVTSPM	PGYTHLQSAQ	IITPGFYLSA	VVEHALRASS
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
RLLGTYDRIN	FSQLGAGPMA	GQELAWDRDW	LAQAIGCAGP	VPHALAAVAS	REWLLDVAAD
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
IASAGVGFSR	FLTDLMAWSS	SAYGLVELPD	ELAGISAAMP	QKKNYPILLER	LRGRTGHLTA
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
FYVDFATGQR	NTPYSNMVEV	SKEAGLHAST	MFGAIRAVLT	GFTLVVDKQLQ	WRTDRMRVAVC
<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
EEDHFGGFSL	ANELTLRAGV	PWRTAQVIAG	RYVTAVLAEG	DGRAGHSNPA	ALDSAAREAG
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
YPLDDPASYL	TNSVDVDAQL	RAKVSAGSAS	PDSVVALLAE	QRQRDLALTA	EWNARTATVR
<u>490</u>					
DAIAATDAAV	RPTAGGVV				

Pac18

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MTSRQLVQQP	ARDHGSPARA	VAVSSAFGTF	GELLQGALPD	DGPDFLVTLF	IARWATATFE
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
YESAHDRVEV	FPATKTKARR	VAEAVLARHS	GGGSLRLSG	SLPEGKGLAS	SSADLVATAR
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
AVASAVGVDL	PPQGIENLLR	RIEPTDGVMY	PGVVAFEHRN	VALLARCGVL	PPMTIVGIDE
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
GGTVDTVAFN	RIPKNFTAAE	REEYARLLDE	VQTAVRAGDA	AAIGRVATRS	AHLNQLCRK
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
RTLNAMTALS	AEIGGVGVVT	AHSGSTIGLM	LPHDVPGFRS	RLAEAINRCG	QLVHRERVLC
<u>310</u>					
YQTLGFERTG	SGPSNDLQ				

Pac19

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MIFNDLVDAI	GHTPAVRLRA	APTDTVVAK	LELQNLFAMK	DRVARQVIRE	ARENGVLAPG
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
APIIESSGT	MALGLALAGH	ALGHPVHIVT	DPRIDAITLA	KLKALGCSVH	VVEGMTSNGW
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
QSARLERLAA	LRAEYPGAFW	PRQYSNPQNP	LAYAALAGEL	VADLGRIDVL	VGAVGSGGSL
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
CGTARALRRA	LKAAGHDGPL	RVIGVDAVGS	VLFQPDQPG	RKQSGIGNSL	IPGNLDYAH
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
DEIHWLNDRE	AFVATRDAR	DEGIFAGNSS	GSVYQVLKHL	AATVAPGTRV	VGIFPDRGDR
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>	<u>350</u>	<u>360</u>
YVDSIYDDGY	WQQAGLAELP	LRTEPLEVSY	GTEVSSWARA	EVPTPRKRL	VFVESNTTGT

Appendix 1

<u>370</u>	<u>380</u>	<u>390</u>	<u>400</u>	<u>410</u>	<u>420</u>
GMLALARARD	LGLAPVLLTS	RPSRYPGIGE	ADCEIIRCDT	NSADTLAAAL	RAHFAPDEVA
<u>430</u>	<u>440</u>	<u>450</u>	<u>460</u>	<u>470</u>	<u>480</u>
GITTTSEFYL	EAAASLAASL	GLPGNPPAAV	GACRRKSATR	RLADAGLPQ	PVSMPVTRVA
<u>490</u>	<u>50</u>	<u>510</u>	<u>520</u>	<u>530</u>	<u>540</u>
EVPGAVAATG	LPCVVKPASD	SASTGVLLCT	SVEQAQEHAA	KLLAITHNVR	EQAVPPEVLV
<u>550</u>	<u>560</u>	<u>570</u>	<u>580</u>	<u>590</u>	<u>600</u>
EELVQGPEFS	VETIFVDAAL	HLVGITRKS	SPPPSFVELR	HAFPASLDEA	ESREIEQVVR
<u>610</u>	<u>620</u>	<u>630</u>	<u>640</u>	<u>650</u>	<u>660</u>
AAIQAIGLRH	GACHTELRLT	AAGPTIIEIN	ARLAGGMIPE	LIRLAGGPD	LTTQLRAAAG
<u>670</u>	<u>680</u>	<u>690</u>	<u>700</u>	<u>710</u>	<u>720</u>
MAVELPRGAL	TPTGVAFITS	SAEGRLASD	GVESARQSAG	VAEVQIARRP	GDPVRPATDA
<u>730</u>	<u>740</u>	<u>750</u>			
YDRIGYIIAS	GDSVSQLDQS	LDSALDSITV	RLAND		

Pac20

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MADLIEMRSD	TFTLPTEQMV	SAMTQAVLGD	DVYGEDPTAN	RLEELAAKSV	GKPAACLMP
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
GTMANLAALL	VHVPRGGKVL	VGNESDIYLY	EAGGASVCGG	IVYEPIPTRP	DGTLALDDLA
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
AAFPPDPDDP	QFALPGLICV	ENTHNRMGGR	VLSQAYLAEL	KRFATGHGIP	VHMDGARIFN
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
AAVATGVAAE	QIAAHADSIQ	FCLSKGLSAP	IGSLAGEADF	IEKARRIRKM	LGGGMRQAGV
<u>250</u>	<u>260</u>	<u>270</u>	<u>280</u>	<u>290</u>	<u>300</u>
FAAAGLVALT	SMIDRLAEDH	QRAAQLAAGL	AEVDGIDVDP	SSVTTNIVLF	RVTANGLDD
<u>310</u>	<u>320</u>	<u>330</u>	<u>340</u>		
RFLRAVEQRG	LVMGEFGHGR	IRAVTHRGLS	SADVSAAVAI	VADVUREAS	

Pac22

<u>10</u>	<u>20</u>	<u>30</u>	<u>40</u>	<u>50</u>	<u>60</u>
MSDNDARARL	LRAMEYRQQF	DERLGRIEQE	TPFDSVLDVD	TTESVELWDL	PDADQLWVGR
<u>70</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>110</u>	<u>120</u>
NARYSPTPVR	TVRSALGKCD	VHEEVTFVDV	GCGKGRVLLL	AAELPFRIV	GVEASEALCD
<u>130</u>	<u>140</u>	<u>150</u>	<u>160</u>	<u>170</u>	<u>180</u>
IARSNVEKAS	VARDGCDRIE	VCHADATKFD	IPDDAGLFYF	YEPFSVDVSL	AVLERIEDSV
<u>190</u>	<u>200</u>	<u>210</u>	<u>220</u>	<u>230</u>	<u>240</u>
RRHPRNVVLC	FTGRGQPDGQ	GSDLEKTPVA	ASEMRAHWNL	VEIVPSPDAE	FYDSFLYEVV

ENGAPQA

Publications

Ragab, A.E, Gruschow, S., Tromans, D.R. and Goss, R.J.M. (2011), 'Biogenesis of the unique 4',5'-dehydromucleoside of the uridyl peptide antibiotic pacidamycin', *Journal of the American Chemical Society*, 133, 15288-15291.

Tromans, D. R., Stevenson, C.E.M., Goss, R.J.M. and Lawson, D.M. (2012), 'Crystallisation and preliminary X-ray analysis of Pac17 from the pacidamycin-biosynthesis cluster of *Streptomyces coeruleorubidus*', *Acta Crystallographica F Structural Biology and Crystallization Communication*, F68, 971-974.

Biogenesis of the Unique 4',5'-Dehydronucleoside of the Uridyl Peptide Antibiotic Pacidamycin

Amany E. Ragab,[†] Sabine Grünschow, Daniel R. Tromans, and Rebecca J. M. Goss*

School of Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, U.K.

S Supporting Information

ABSTRACT: The pacidamycins belong to a class of antimicrobial nucleoside antibiotics that act by inhibiting the clinically unexploited target translocase I, a key enzyme in peptidoglycan assembly. As with other nucleoside antibiotics, the pacidamycin 4',5'-dehydronucleoside portion is an essential pharmacophore. Here we show that the biosynthesis of the pacidamycin nucleoside in *Streptomyces coeruleorubidus* proceeds through three steps from uridine. The transformations involve oxidation of the 5'-alcohol by Pac11, transamination of the resulting aldehyde by Pac5, and dehydration by the Cupin-domain protein Pac13.

Nucleoside antibiotics exhibit versatile biological properties such as antibacterial, antifungal, antiviral, and antitumor activity.^{1,2} The wide scope of such activities stems from the structural similarity of these compounds to nucleosides and nucleotides, which play key roles as energy donors, metabolite carriers, and enzyme cofactors. Puromycin is probably the best known example because of its widespread use in the study of the ribosome and in protein evolution through mRNA display.^{3,4} Here we report the first insights into the biogenesis of the pacidamycin aminonucleoside **1** and demonstrate that the unusual dehydroaminonucleoside can be formed in vitro from uridine by three enzymes. The enzymes involved in the generation of this rare aminonucleoside demonstrate substrate flexibility and have the potential to be utilized for biotransformations that will enable access to series of new bioactive compounds.

Figure 1 shows the chemical structures of aminonucleoside moieties found in peptidyl nucleoside antibiotics. The peptide portions of these compounds share no similarities. Despite all the differences, however, the biosyntheses of all the aminonucleosides characterized to date begin with their corresponding nucleotides. Puromycin and the related antibiotic A201A contain a 3'-amino-3'-deoxyadenosine moiety **4** that has been shown to be derived from adenosine-5'-triphosphate (ATP).⁵ The biosyntheses of the 5'-deoxynucleoside **3** (found in nikkomycins) and the bisaminonucleoside **2** (found in liposidomycins) begins with uridine-5'-monophosphate (UMP).^{6–8}

The biosynthesis of the aminonucleoside moiety **1** of the pacidamycin group of uridyl peptide antibiotics has not been investigated until now. The uridyl peptide antibiotics, such as pacidamycin, napsamycin, and mureidomycin, are translocase I inhibitors. This mode of action is shared with the liposidomycins. The aminonucleoside motif **1** of the pacidamycin-like antibiotics

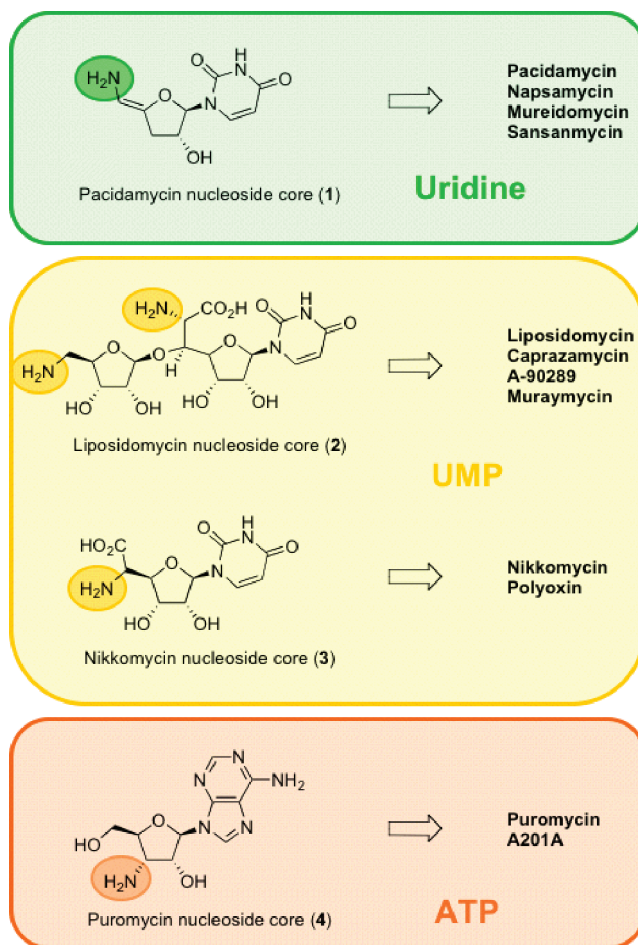


Figure 1. Aminonucleoside portions of peptidyl nucleoside antibiotics.

differs from other aminonucleosides in that it lacks the 3'-hydroxyl group and has an exocyclic double bond. The unique features of **1** imply a novel biosynthetic pathway.

On the basis of sequence analysis of the pacidamycin biosynthetic cluster, we had suggested the involvement of the flavin-dependent oxidoreductase Pac11 and the aminotransferase Pac5 in the biosynthesis of the aminonucleoside.⁹ These two enzymes together could potentially provide 5'-amino-5'-deoxyuridine **7** from uridine or UMP. We postulated that at least one additional enzyme was needed to account for the double bond in the

Received: July 3, 2011

Published: August 23, 2011

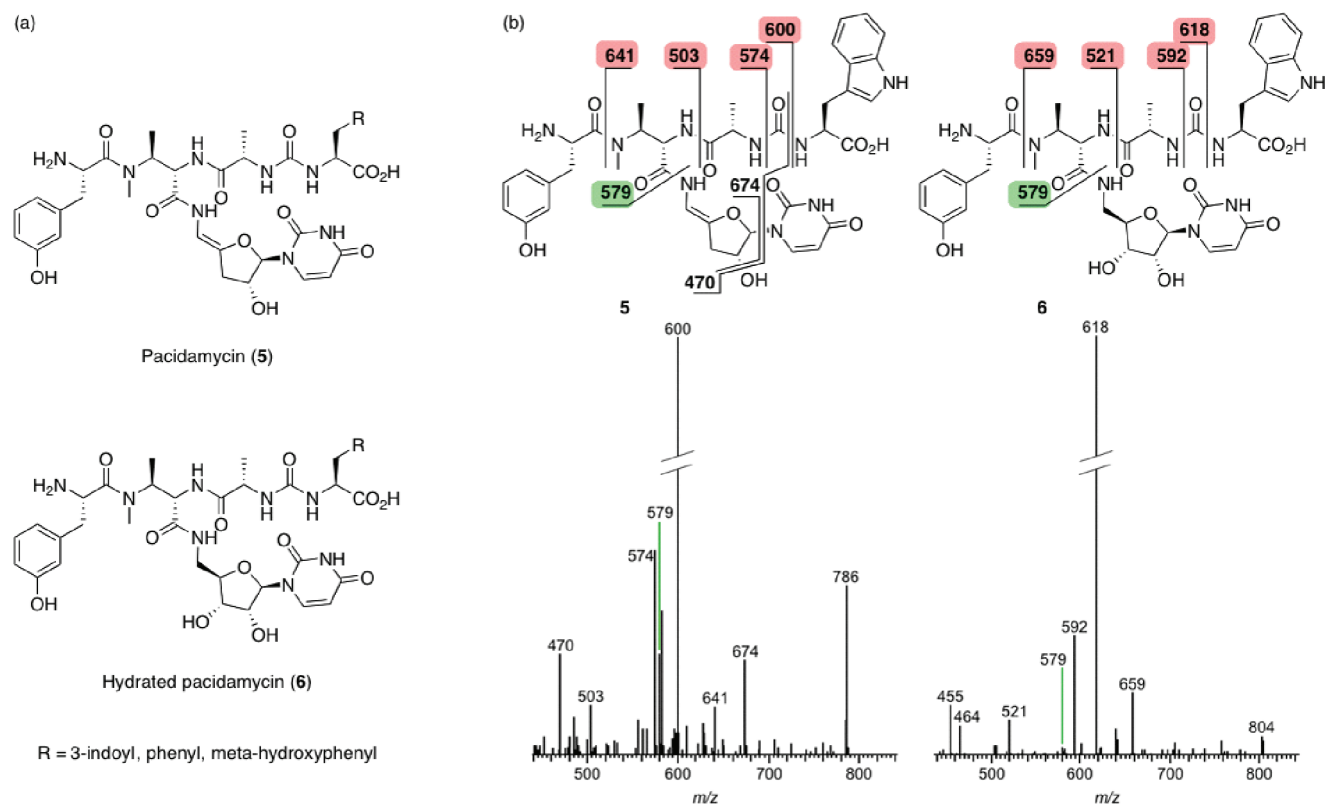


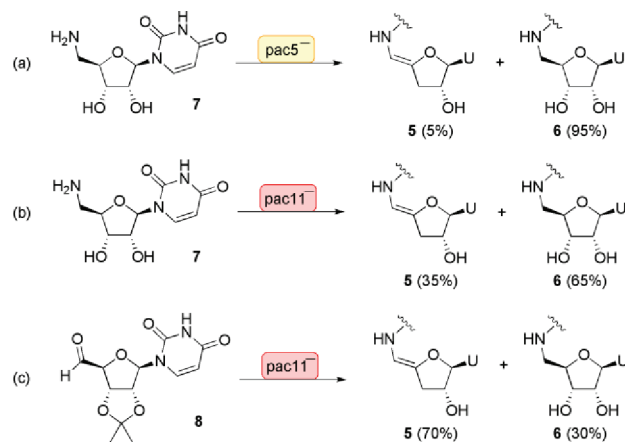
Figure 2. (a) Structures of pacidamycins found in wild-type *S. coeruleorubidus* (5) and the *pac13* mutant (6). (b) MS/MS analyses of 5 and 6.

pacidamycin aminonucleoside 1. Of the 22 genes in the pacidamycin core cluster, only *pac6*, *pac7*, and *pac13* had not been assigned a function. Pac6 shows sequence homology to non-heme iron dioxygenases belonging to the TauD family of enzymes¹⁰ and was therefore an alternative candidate for formation of the uridyl aldehyde. However, none of the strictly conserved iron-binding residues are found in Pac6 [see the Supporting Information (SI)], indicating that the enzyme may not be active as a dioxygenase. Pac7 has no significant homologies to characterized proteins, and Pac13 belongs to the Cupin family of proteins. The Cupin family is characterized by a metal-binding motif, and members of this protein family have a diverse set of functions.¹¹

In order to determine which of these genes were involved in the biosynthesis of the aminonucleoside, knockout mutants for each of the five genes were constructed. Analysis of the culture extracts by liquid chromatography–tandem mass spectrometry (LC–MS/MS) showed that none of the mutant strains were able to generate the typical pacidamycin suite of compounds. Upon closer inspection of the metabolite profile of the *pac13* mutant, hydrated pacidamycins 6 were detected (Figure 2). These pacidamycins were not detected in any of the other mutant strains or in the wild-type organism. The presence of hydrated pacidamycins strongly suggests that Pac13 acts as the dehydratase. Furthermore, this finding indicates that the aminonucleoside 7 is a chemically competent substrate for attachment to the pacidamycin peptide portion in vivo, as had been previously demonstrated in an in vitro study with the pacidamycin non-ribosomal peptide synthetases.¹²

The origins of the aminonucleoside portion in primary pyrimidine metabolism were established by synthesizing and

Scheme 1. Feeding of Precursors to Pacidamycin Mutants



feeding deuterated ribose, uracil, and uridine to the wild-type pacidamycin producer. All three compounds resulted in the incorporation of the deuterium label into the nucleoside portion of 5, as established through MS/MS (Figures S7–S9 in the SI). Next, we sought to re-establish pacidamycin biosynthesis in the mutant strains through chemical complementation. To this end, putative biosynthetic intermediates for the aminonucleoside portion were synthesized (see the SI) and fed to cultures of the *S. coeruleorubidus* mutants.

Production of wild-type pacidamycins 5 was restored in the *pac5* and *pac11* mutants through addition of aminonucleoside 7 (Scheme 1a,b). The hydrated pacidamycins 6 were detected

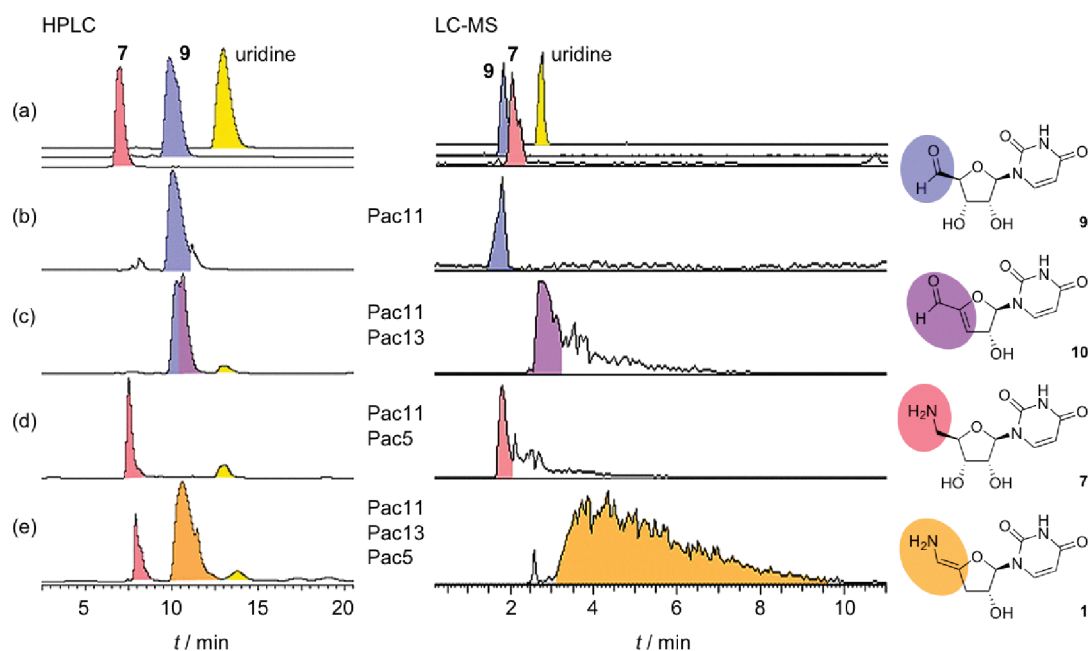


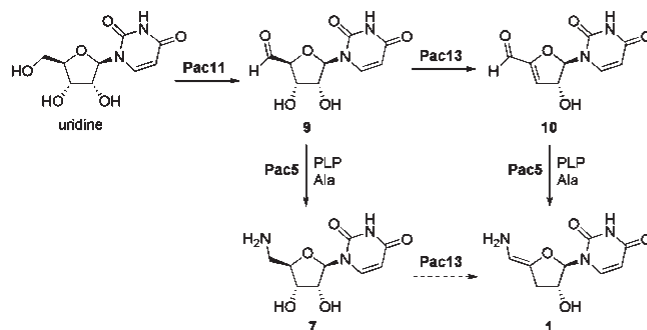
Figure 3. HPLC and LC–MS analyses of (a) standards and (b–e) in vitro assays. The structure of the main product for each enzyme combination is shown. In the LC–MS chromatograms, the extracted ion trace for the main product is shown.

alongside the enamide-containing pacidamycins **5** in both of these extracts. In contrast, no change in metabolite profile was observed in culture extracts of the *pac6* or *pac7* mutants (Figure S10). Further chemical complementation experiments were planned with the protected 5'-uridylaldehyde **8**. We decided to use the protected aldehyde because of stability issues with unprotected 5'-uridylaldehyde **9**, which slowly dehydrates under mildly acidic conditions. We reasoned that deprotection of **8** would occur during culturing, either spontaneously or enzymatically. To test the chemical competence of **8** as a pathway intermediate, we administered its deuterated form [$5\text{-}^2\text{H}$]-**8** to wild-type *S. coereborubidus*. This resulted in the detection of deuterated pacidamycins. Rather surprisingly, however, the deuterated form of the hydrated pacidamycin **6** was also detected (Figure S18). When aldehyde **8** was administered to cultures of the *pac11* mutant, pacidamycin production was restored (Figure S10 and Scheme 1c). Again, hydrated pacidamycins **6** were detected alongside **5** in the *pac11* mutant. Supplementing cultures of the *pac6* and *pac7* mutant strains with **8** did not result in the production of any pacidamycins (Figure S10).

Taken together, the data suggest that *pac5*, *pac11*, and *pac13* are involved in the biosynthesis of pacidamycin aminonucleoside **1**. The involvement of *pac6* or *pac7* in the generation of the aminonucleoside is not supported by our findings. Our experiments clearly indicate that the *pac13*-mediated dehydration is not required for pacidamycin biosynthesis to go to completion. The results obtained from the chemical complementation experiments are consistent with initial action of Pac11 to generate uridyl aldehyde **9** followed by Pac5-catalyzed transamination to give the 5'-aminouridine derivative.

In order to test whether Pac5, Pac11, and Pac13 were sufficient to generate aminonucleoside **1** in vitro, we heterologously expressed the three enzymes in *Escherichia coli* BL21(DE3) as His₆-fusion proteins. The enzymes were purified by affinity chromatography. The putative uridine dehydrogenase Pac11 is homologous to flavin-dependent oxidoreductases of the vanillyl

Scheme 2. Biosynthesis of Pacidamycin Nucleoside 1



alcohol oxidase family. Many enzymes of this family covalently bind their flavin cofactor.¹³ Spectroscopic and mass spectrometric analysis of purified Pac11 confirmed the presence of a covalently attached flavin cofactor (Figures S4 and S5). For the putative aminotransferase Pac5, spectral analysis indicated that pyridoxal-5'-phosphate (PLP) did not copurify with the enzyme (Figure S3). Addition of Pac5 to PLP, however, did result in the shifts characteristic for Schiff base formation between the Pac5 active-site lysine (Lys270) and PLP.¹⁴

Enzyme reactions were allowed to incubate at 28 °C for 18 h prior to LC–MS or HPLC analysis of the products. The identities of **7**, **9**, and **10** were established through comparison of their MS and MS/MS data to those obtained from synthetic standards (Figures S22–S24) and HPLC coinjections with these synthetic standards (Figure S21). Pac11 was capable of oxidizing uridine to uridyl aldehyde **9** (Figure 3). Traces of dehydrated aldehyde **10** were also detected (Figure S22). Although rare, molecular oxygen can act as a terminal electron sink for flavin, thus bypassing the need for nicotinamide-mediated cofactor recycling. No conversion was observed when UMP was used as the substrate instead of uridine (Figure S21). This is the first firm

experimental evidence that the biosynthesis of pacidamycin nucleoside **1** starts from a different precursor than for liposidomycin-type nucleoside **2** or nikkomycin-related nucleoside **3**. The reaction of Pac13 with uridine did not yield any product (Figure S21), which strongly suggests that Pac11 is the first enzyme to act in aminonucleoside biosynthesis. The reaction of Pac11 in combination with Pac13 generated **10** as the major product from uridine, and uridyl aldehyde **9** was detected as a minor product (Figure 3 and Figure S22). Enzyme assays using Pac5 were performed in the presence of PLP and alanine (see section 4.3 in the SI). As expected, aminonucleoside **7** was obtained when uridine was incubated with Pac11 and Pac5.

When all three enzymes were incubated together, a new peak appeared as the major product. Though under the conditions of the HPLC its retention time was similar to those of **9** and **10**, in LC–MS it differed (Figure 3 and Figure S24). Its molecular ion and MS/MS fragmentation pattern were consistent with pacidamycin nucleoside **1** (see the SI). This product was detected only when Pac5, Pac11, and Pac13 were present. The new compound exhibited extreme tailing under the acidic conditions employed for LC–MS, possibly because of the presence of interconverting isomers. Acetylated nucleoside **1** model compounds were shown to be surprisingly stable toward double-bond isomerization, but the free amine **1** was never tested in those studies.¹⁵ Interestingly, in the three-enzyme mixture, aminonucleoside **7** was also present alongside residual amounts of aldehydes **9** and **10** (Figure S24). Pac13 showed the ability to convert **7** to **1** (Figure S21). Pac13 was also incubated with hydrated pacidamycins purified from a *pac13* mutant, but the enzyme was unable to mediate the dehydration of these species (Figure S25).

In summary, we have identified the genes that are required for formation of pacidamycin nucleoside **1**. The biosynthesis of the nucleosidic portion follows a pathway that is different from previously characterized nucleoside antibiotic pathways. The biogenesis of the nucleoside has been demonstrated to start from uridine, which is converted to its aldehyde **9** by the flavin-dependent dehydrogenase Pac11. This is in contrast to the first step in the biosynthesis of liposidomycin nucleoside **2**, where UMP is oxidized to aldehyde **9** through the action of LipL, a non-heme iron α -ketoglutarate-dependent enzyme.⁸ To the best of our knowledge, this is the first example wherein the biosynthesis of a nucleoside antibiotic starts from the nucleoside and not the nucleotide. Furthermore, we have demonstrated that the dehydration is mediated by the Cupin family enzyme Pac13 and that the transamination is catalyzed by Pac5. Both enzymes are relatively flexible in their substrate requirements, allowing the biosynthesis of **1** to follow a randomized order, as shown in Scheme 2.

The generation of the unnatural pacidamycin **6** that occurs upon feeding of aldehyde **8** and amine **7**, hints at the inherent substrate flexibility within the pathway.^{16,17} While the observed metabolic plasticity leads to problems in pinpointing the timing of the dehydration step, it will facilitate the generation of pacidamycin analogues with altered nucleoside portions. We are carrying out further structural and biochemical assessment of the enzymes on this pathway and exploring the use of these enzymes in the generation of nucleoside analogues for antiviral therapies.

■ ASSOCIATED CONTENT

● **Supporting Information.** Full experimental details, LC–MS chromatograms, MS/MS spectra, and protein sequence

alignments. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

r.goss@uea.ac.uk

Present Addresses

[†]Ph.D. sabbatical from Tanta University, Egypt.

■ ACKNOWLEDGMENT

We thank Prof. Julia Butt (UEA) and Dr. Miles Cheeseman (UEA) for useful discussions, Prof. M. J. Bibb (John Innes Centre) for provision of *E. coli* BW25113/pIJ790 and plasmid pIJ773, and Dr. L. M. Hill (JIC) for assistance with LC–MS/MS. Funding by the Egyptian Ministry of Higher Education and the Leverhulme Trust (F/00204/AF) is acknowledged.

■ REFERENCES

- (1) Winn, M.; Goss, R. J. M.; Kimura, K.; Bugg, T. D. H. *Nat. Prod. Rep.* **2010**, *27*, 279.
- (2) Isono, K. *J. Antibiot.* **1988**, *41*, 1711.
- (3) Beringer, M.; Rodnina, M. V. *Mol. Cell* **2007**, *26*, 311.
- (4) Lipovsek, D.; Pluckthun, A. *J. Immunol. Methods* **2004**, *290*, 51.
- (5) Rubio, M. A.; Espinosa, J. C.; Tercero, J. A.; Jimenez, A. *FEBS Lett.* **1998**, *437*, 197.
- (6) Gijnj, C.; Ruegger, H.; Amrhein, N.; Macheroux, P. *ChemBioChem* **2005**, *6*, 1974.
- (7) Chen, W. Q.; Huang, T. T.; He, X. Y.; Meng, Q. Q.; You, D. L.; Bai, L. Q.; Li, J. L.; Wu, M. X.; Li, R.; Xie, Z. J.; Zhou, H. C.; Zhou, X. F.; Tan, H. R.; Deng, Z. X. *J. Biol. Chem.* **2009**, *284*, 10627.
- (8) Yang, Z. Y.; Chi, X. L.; Funabashi, M.; Baba, S.; Nonaka, K.; Pahari, P.; Unrine, J.; Jacobsen, J. M.; Elliott, G. I.; Rohr, J.; Van Lanen, S. G. *J. Biol. Chem.* **2011**, *286*, 7885.
- (9) Rackham, E. J.; Grünschow, S.; Ragab, A. E.; Dickens, S.; Goss, R. J. M. *ChemBioChem* **2010**, *11*, 1700.
- (10) Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21.
- (11) Dunwell, J. M.; Culham, A.; Carter, C. E.; Sosa-Aguirre, C. R.; Goodenough, P. W. *Trends Biochem. Sci.* **2001**, *26*, 740.
- (12) Zhang, W.; Ntai, I.; Bolla, M. L.; Malcomson, S. J.; Kahne, D.; Kelleher, N. L.; Walsh, C. T. *J. Am. Chem. Soc.* **2011**, *133*, 5240.
- (13) Heuts, D.; Scrutton, N. S.; McIntire, W. S.; Fraaije, M. W. *FEBS J.* **2009**, *276*, 3405.
- (14) Vedavathi, M.; Girish, K. S.; Kumar, M. K. *Biochemistry (Moscow)* **2006**, *71*, S105.
- (15) Gentle, C. A.; Bugg, T. D. H. *J. Chem. Soc., Perkin Trans. 1* **1999**, 1279.
- (16) Grünschow, S.; Rackham, E. J.; Elkins, B.; Newill, P. L. A.; Hill, L. M.; Goss, R. J. M. *ChemBioChem* **2009**, *10*, 355.
- (17) Ragab, A. E.; Grünschow, S.; Rackham, E. J.; Goss, R. J. M. *Org. Biomol. Chem.* **2010**, *8*, 3128.

Daniel R. Tromans,^{a,b} Clare E. M. Stevenson,^a Rebecca J. M. Goss^b and David M. Lawson^{a*}

^aDepartment of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, England, and ^bSchool of Chemistry, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, England

Correspondence e-mail: david.lawson@jic.ac.uk

Received 22 May 2012

Accepted 26 June 2012

Crystallization and preliminary X-ray analysis of Pac17 from the pacidamycin-biosynthetic cluster of *Streptomyces coeruleorubidus*

Pac17 is an uncharacterized protein from the pacidamycin gene cluster of the soil bacterium *Streptomyces coeruleorubidus*. It is implicated in the biosynthesis of the core diaminobutyric acid residue of the antibiotic, although its precise role is uncertain at present. Given that pacidamycins inhibit translocase I of *Pseudomonas aeruginosa*, a clinically unexploited antibiotic target, they offer new hope in the search for antibacterial agents directed against this important pathogen. Crystals of Pac17 were grown by vapour diffusion and X-ray data were collected at a synchrotron to a resolution of 1.9 Å from a single crystal. The crystal belonged to space group *C2*, with unit-cell parameters $a = 214.12$, $b = 70.88$, $c = 142.22$ Å, $\beta = 92.96^\circ$. Preliminary analysis of these data suggests that the asymmetric unit consists of one Pac17 homotetramer, with an estimated solvent content of 49.0%.

1. Introduction

The pacidamycins (Fig. 1) are a suite of secondary metabolites produced by the Gram-positive bacterium *Streptomyces coeruleorubidus*. These compounds show an exquisitely narrow range of activity against the pathogen *Pseudomonas aeruginosa* by inhibiting the action of translocase I. Although peptidoglycan biosynthesis is a validated target for antimicrobial agents such as penicillin and vancomycin, to date compounds that specifically inhibit translocase I have not seen clinical application (Winn *et al.*, 2010). Thus, the pacidamycins offer scope for the development of new antibiotics in the fight against multidrug-resistant bacteria.

At the core of the structure of the pacidamycins is a (2*S*,3*S*)-diaminobutyric acid (DABA) moiety (Fig. 1), a nonproteinogenic amino acid that is found in a number of other natural products including the related uridyl peptide antibiotics, namely the mureidomycins (Isono & Inukai, 1991) and napsamycins (Chatterjee *et al.*, 1994), and the lipopeptide antibiotic friulimicin (Vértesy *et al.*, 2000). The pacidamycin-biosynthetic gene cluster has recently been identified in the antibiotic producer and the functions of a number of gene products have been inferred (Rackham *et al.*, 2010; Zhang *et al.*, 2010). Amongst these are Pac18 and Pac19, which are implicated in DABA biosynthesis, specifically catalysing the ATP-dependent phosphorylation of L-threonine and the conversion of the resultant phosphothreonine to DABA by a pyridoxal-phosphate-dependent β -replacement, respectively. The precise biological function of Pac17

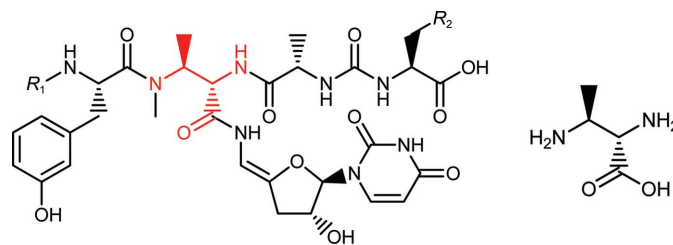
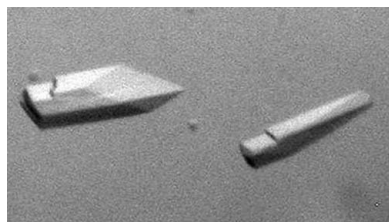


Figure 1
 The pacidamycins have a common core that is decorated by groups R_1 and R_2 , where R_1 is alanine, glycine or hydrogen and R_2 is indolyl, phenyl or 3-hydroxyphenyl. The central diaminobutyric acid (DABA) moiety is highlighted in red and shown separately on the right. This figure was created using *ChemDraw*.

in pacidamycin biosynthesis is uncertain. It shares significant amino-acid sequence homology to argininosuccinate lyases and the putative catalytic His and Ser residues are conserved in Pac17, whilst all of the substrate-binding residues are not (Sampaleanu *et al.*, 2002). The observation that *pac17* is translationally coupled to *pac18* and *pac19* suggests that it also plays a role in the synthesis of DABA (Rackham *et al.*, 2010; Zhang *et al.*, 2010). Homologues of Pac17 exist in the genomes of several other actinomycetes, and in at least two cases these occur in gene clusters for uridyl peptide antibiotic biosynthesis (Müller *et al.*, 2007; Kaysser *et al.*, 2011), but none of these proteins have been characterized to date.

The closest known structural homologue of Pac17, with an amino-acid sequence identity of 33% over 74% of the sequence, is argininosuccinate lyase from *Thermus thermophilus* HB8. This was used as a template to solve the Pac17 structure by molecular replacement. Here, we report the crystallization and preliminary X-ray analysis of Pac17, which represents the first of the pacidamycin-biosynthetic enzymes to be crystallized.

2. Materials and methods

2.1. Protein expression, purification and crystallization

The *pac17* gene of *S. coeruleorubidus* (UniProtKB entry E2EKP9; synonym *pacQ*) was amplified by PCR from the cosmid 2H-5 (Rackham *et al.*, 2010), which contained the minimal pacidamycin gene cluster, using a forward primer containing an *NdeI* restriction site (5'-GGACGACATATGGTGAGACTGACCGGTCGACTT-3') and a reverse primer containing a *BamHI* site (5'-TCCGGATCC-TCGTCAGACTCCCCCGG-3'). The amplified DNA was *NdeI/BamHI*-digested and subsequently ligated into *NdeI/BamHI*-digested expression vector pET28a(+) to produce the expression construct pET28a(+)Pac17, which encodes the native 498-residue Pac17 protein preceded by a thrombin-cleavable N-terminal hexahistidine tag. The inclusion of this tag appended an additional 21 amino acids onto the N-terminus of the protein with the sequence MGSS-HHHHHHSSGLVPRGSHM, giving a total molecular weight of 55 859.1 Da. The pET28a(+)Pac17 expression vector was introduced into *Escherichia coli* BL21 (DE3) cells by transformation. For protein production, 10 ml of an overnight culture of these cells was used to inoculate 1 l autoinduction medium broth containing 50 µg ml⁻¹ kanamycin. The culture was grown at 310 K for 4 h and for a further 16 h at 289 K. The cells were harvested by centrifugation using a Sorvall Evolution centrifuge (15 min, 5000 rev min⁻¹, 277 K, SLC-4000 rotor) and stored at 253 K prior to purification.

All purification steps were performed at 277 K. The cell pellet was resuspended in buffer A (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 40 mM imidazole) containing a Complete EDTA-free protease-

inhibitor cocktail (Roche) and lysed by sonication. The supernatant and pellet were separated by centrifugation in a Sorvall Evolution centrifuge (45 min, 18 000 rev min⁻¹, 277 K, SS34 rotor). Pac17 was purified from the supernatant using a two-step procedure performed in series using an ÄKTAexpress FPLC (GE Healthcare). The sample was applied onto a 5 ml Ni²⁺-charged His-Trap Chelating HP column (GE Healthcare), washed with 20 column volumes (CV) of buffer A and then eluted with 5 CV buffer A containing 500 mM imidazole at a flow rate of 4.0 ml min⁻¹. The major protein peak (based on an absorbance of >100 mAU at 280 nm) was automatically applied onto a Superdex 200 HiLoad HP gel-filtration column (GE Healthcare) in buffer B (20 mM HEPES pH 7.5, 150 mM NaCl) and eluted over 1.3 CV at a flow rate of 3.2 ml min⁻¹. Fractions containing the Pac17 protein (as confirmed by SDS-PAGE) were pooled and concentrated to approximately 11 mg ml⁻¹ (as measured using the Bradford assay) in buffer B using an Amicon Ultra-15 30 kDa cutoff centrifugal concentrator (Millipore) for crystallization. The N-terminal His tag was not cleaved from the purified protein. Approximately three quarters of the protein sample was flash-frozen in liquid nitrogen as 50 µl aliquots in PCR tubes and stored at 193 K for subsequent use. The remainder was used immediately in crystallization trials.

Crystallization trials of His-tagged Pac17 were set up using an OryxNano robot (Douglas Instruments Ltd) in sitting-drop vapour-diffusion format with 96-well MRC plates (Molecular Dimensions) using a variety of commercially available screens (Hampton Research and Molecular Dimensions) at a constant temperature of 293 K. Drops consisted of 0.3 µl protein solution mixed with 0.3 µl precipitant solution and the reservoir volume was 50 µl. A number of conditions produced crystals, which were then optimized in a 24-well hanging-drop vapour-diffusion format using VDX plates (Hampton Research) with a reservoir volume of 1 ml and drops consisting of 1 µl protein solution and 1 µl precipitant solution. For each optimization, a fresh aliquot of frozen protein was used. In preparation for cryogenic data collection at the synchrotron, crystals were grown from precipitant solution supplemented with 15%(v/v) glycerol.

2.2. X-ray data collection and analysis

Crystals were mounted for X-ray data collection using LithoLoops (Molecular Dimensions), flash-cooled by plunging them into liquid nitrogen and stored in Unipuck cassettes (MiTeGen) prior to transport to the synchrotron. Crystals were subsequently transferred robotically to the goniostat on station I02 at the Diamond Light Source (Oxfordshire, England) and maintained at 100 K with a Cryojet cryocooler (Oxford Instruments). Diffraction data were recorded using an ADSC Quantum 315 CCD detector with the wavelength set to 0.9795 Å and were integrated using *XDS* (Kabsch, 2010) and scaled using *SCALA* (Evans, 2006). Further data analysis was performed using the *CCP4* program suite (Winn *et al.*, 2011).

3. Results and discussion

The His-tagged Pac17 construct was overproduced and purified to greater than 95% purity as determined by SDS-PAGE analysis, with a final yield of approximately 24 mg per litre of culture. The gel-filtration column had previously been calibrated using molecular-weight gel-filtration standards (GE Healthcare). Pac17 eluted at a volume that corresponded to a molecular weight of ~150 kDa, being indicative of a multimeric species and closest to the expected value for a homotrimer (~168 kDa).

Crystals grew within 24 h at 293 K from a number of crystallization conditions. These conditions were optimized to improve crystal size,

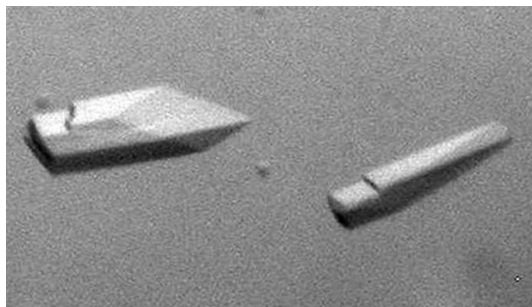


Figure 2 Crystals of Pac17 with approximate dimensions of 150 × 50 × 30 µm.

with the largest crystals appearing in a precipitant solution consisting of 15% (w/v) PEG 3350, 0.2 M potassium sodium tartrate, 0.1 M bis-tris propane pH 7.5, 15% (v/v) glycerol. The largest crystals formed were approximately $150 \times 50 \times 30 \mu\text{m}$ in size (Fig. 2).

Native X-ray diffraction data were collected from a single Pac17 crystal: $1000 \times 0.2^\circ$ oscillation images were recorded in a single sweep to a maximum resolution of 1.9 Å. Indexing of the data was consistent with a *C*-centred monoclinic lattice with unit-cell parameters $a = 214.12$, $b = 70.88$, $c = 142.22$ Å, $\beta = 92.96^\circ$. The resultant reduced data set was 99.3% complete to a resolution of 1.9 Å. Data statistics are given in Table 1.

Solvent-content analysis suggested that the asymmetric unit was most likely to contain three or four His-tagged Pac17 monomers, giving estimated solvent contents of 61.8 and 49.0%, respectively (Matthews, 1968). Inspection of a self-rotation function calculated using *MOLREP* (Vagin & Teplyakov, 2010) revealed a noncrystallographic twofold axis perpendicular to *b* in the *ac* plane which, when combined with the crystallographic twofold, generates apparent 222 symmetry. This would be consistent with an asymmetric unit comprised of a 222-symmetric homotetramer. Further analysis of the data with *SFCHECK* (Vaguine *et al.*, 1999) revealed a pseudo-

translation vector of 0.386, 0.000, 0.491 (fractional coordinates) at 28% of the origin peak (Fig. 3c).

Interrogation of the Protein Data Bank (<http://www.rcsb.org/pdb>) using a protein *BLAST* search revealed that the closest structural homologue was argininosuccinate lyase from *T. thermophilus* HB8 (PDB entry 2e9f; M. Goto, unpublished work), which shows 74% sequence coverage and 33% sequence identity to Pac17. The biological unit (and asymmetric unit) of the former is a homotetramer with 222 symmetry. Both monomer and tetramer polyalanine molecular-replacement templates were created from this structure using *CHAINSAW* (Stein, 2008). These were used as inputs to *Phaser* v2.3.0 (McCoy *et al.*, 2007), which was run at 4.5 Å resolution using the default cutoffs for peak selection. With the monomer template *Phaser* reported only two solutions, each with four molecules per asymmetric unit (TFZ-equivalent scores of 36.0 and 33.4, respectively). Inspection of these solutions using *Coot* (Emsley & Cowtan, 2004) revealed them to be very similar, differing only in the placement of the subunits relative to the origin. In both cases the asymmetric unit was comprised of two equivalent 'dimers' arranged in a back-to-back fashion; after the application of crystallographic symmetry one of these dimers generated a homotetramer

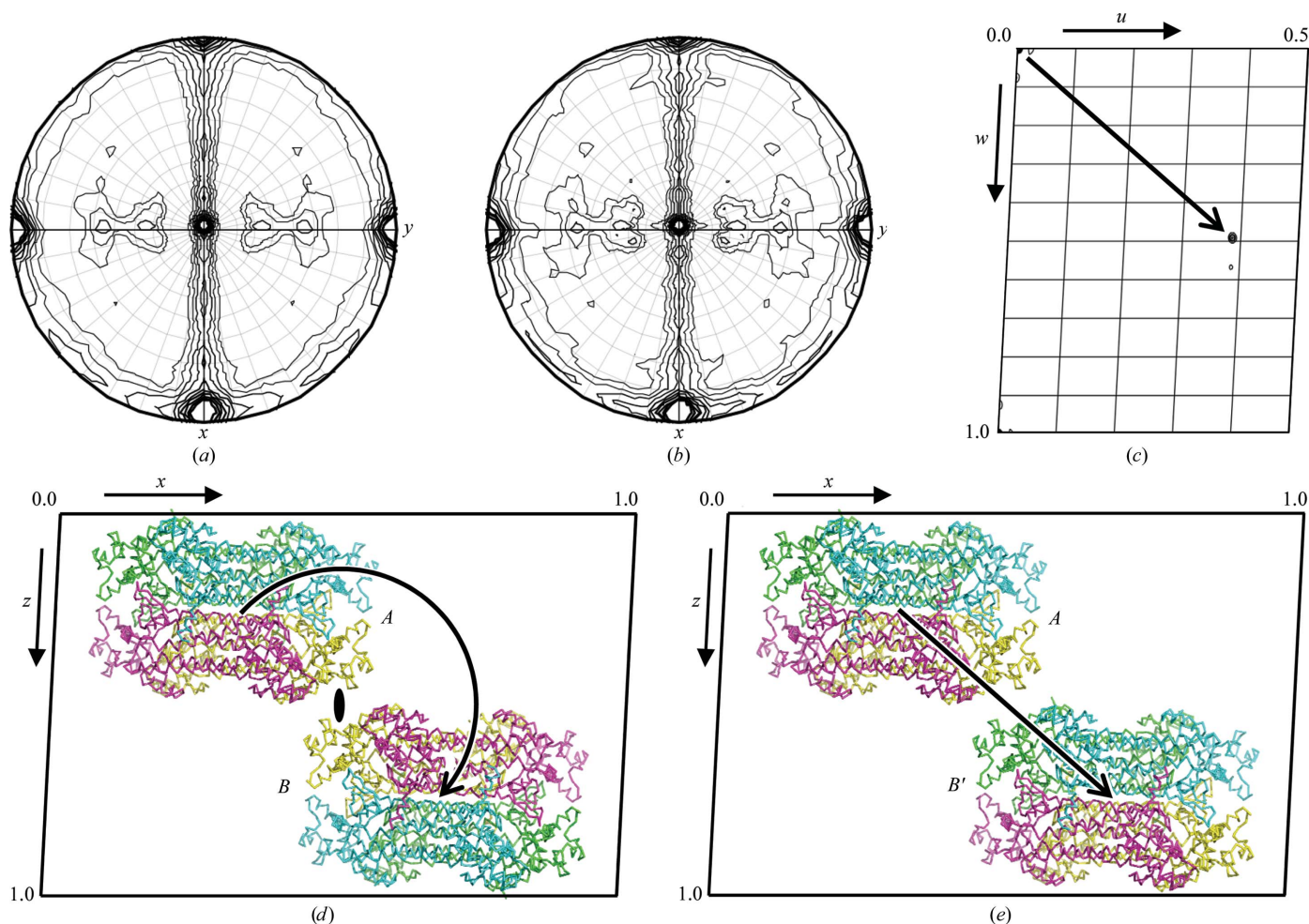


Figure 3

Preliminary crystallographic analysis of Pac17. Self-rotation functions calculated to 4.5 Å resolution from (a) the experimental data and (b) the molecular-replacement solution, showing the same noncrystallographic twofold axes ($\chi = 180^\circ$ section). Note the alignment of one of these with the crystallographic twofold axis (*y* axis). (c) Self-Patterson function (section $v = 0$) calculated to 4.5 Å resolution from the experimental data, revealing a clear pseudotranslation vector of 0.386, 0.000, 0.491 (fractional coordinates). The alignment of crystallographic and noncrystallographic twofold axes gives rise to similarly oriented tetramers in the unit cell. Specifically, (d) the application of twofold crystallographic symmetry (operator: $1 - x, y, 1 - z$) generates tetramer *B* from tetramer *A*, which is essentially equivalent to (e) the translation of molecule *A* by the pseudotranslation vector to give molecule *B'*. For clarity, only two of the four copies of the tetramer in the unit cell are shown in (d) and (e). (d) and (e) were created using *PyMOL* (DeLano, 2002).

Table 1

Summary of X-ray data for Pac17.

Values in parentheses are for the outer resolution shell.

No. of crystals	1
Beamline	I02, Diamond Light Source
Wavelength (Å)	0.9795
Detector	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	290.7
Rotation range per image (°)	0.2
Exposure time per image (s)	0.25
Beam transmission (%)	27.2
Total rotation range (°)	200.0
Resolution range (Å)	67.28–1.90 (2.00–1.90)
Space group	C2
Unit-cell parameters (Å, °)	$a = 214.12, b = 70.88,$ $c = 142.22, \beta = 92.96$
Estimated mosaicity (°)	0.2
Total No. of measured intensities	672568 (74768)
Unique reflections	166584 (23088)
Multiplicity	4.0 (3.2)
Mean $I/\sigma(I)$	8.5 (2.0)
Completeness (%)	99.3 (95.1)
$R_{\text{merge}}^{\dagger}$	0.127 (0.583)
$R_{\text{meas}}^{\ddagger}$	0.147 (0.704)
$CC_{1/2}^{\S}$	0.994 (0.706)
Wilson B value (Å ²)	15.6

$\dagger R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$. $\ddagger R_{\text{meas}} = \frac{\sum_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl . $\S CC_{1/2}$ is the correlation coefficient between intensities from random halves of the data set.

corresponding to the biological unit of the template structure, whilst the other dimer did not generate a homotetramer and in fact partially overlapped with crystallographically related dimers (PAK scores of 73 and 68 for the two solutions, respectively). With the tetramer template, only a single solution was reported (TFZ-equivalent score of 49.4). This was judged to be a correct solution because (i) the packing looked reasonable, *i.e.* there were no clashes (PAK score = 0) and no large gaps in the lattice, (ii) a self-rotation function calculated from the model structure factors (Fig. 3*b*) was consistent with that calculated from the experimental structure factors (Fig. 3*a*), indicating that the twofold axes of the model were correctly oriented, and (iii) since one of the model twofold axes was parallel to the crystallographic twofold axis (Fig. 3*a*), the application of twofold crystallographic symmetry resulted in similarly oriented tetramers within the same unit cell (Fig. 3*d*), which were therefore also related by translational symmetry alone (Fig. 3*e*). The corresponding translational vector agreed with that reported by *SFCHECK* (Fig. 3*c*). Rigid-body refinement of this model at 4.5 Å resolution using *REFMAC5* (Murshudov *et al.*, 2011) gave an R_{free} value of 0.547, a free correlation coefficient of 0.525 and a figure of merit of 0.393.

Subsequent analysis of the *Phaser* solutions generated from the monomer template indicated that in both cases all four subunits were correctly oriented and the two dimers corresponded to halves of the biological unit. It is not clear why *Phaser* failed to correctly combine the two dimers in the translation function, but it may have been

influenced by the pseudo-symmetry, or perhaps in some way ‘confused’ by the densely packed core of the homotetramer, which is comprised of 20 long roughly parallel α -helices (five per subunit). Indeed, this densely packed core, and the distinctly nonspherical overall shape of the assembly, could account for the anomalously low estimate of the multimeric state from the gel-filtration column.

Rebuilding and refinement of the preliminary Pac17 model are under way. A full description of this process and analysis of the resultant structure will be reported elsewhere. This work represents the first step towards a full structural and functional characterization of Pac17.

We thank the Norwich Research Park for a studentship (awarded to DRT) and the Biotechnology and Biological Sciences Research Council for support through grants BB/I022910/1 (to RJMG) and BB/J004561/1 (to the John Innes Centre) and from the John Innes Foundation. We are grateful for the support of the Diamond Light Source and for the assistance of the beamline scientists at station I02. S. Grüşchow is acknowledged for helpful advice and technical assistance.

References

- Chatterjee, S., Nadkarni, S. R., Vijayakumar, E. K., Patel, M. V., Ganguli, B. N., Fehlhaber, H. W. & Vertesy, L. (1994). *J. Antibiot.* **47**, 595–598.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst. D* **60**, 2126–2132.
- Evans, P. (2006). *Acta Cryst. D* **62**, 72–82.
- Isono, F. & Inukai, M. (1991). *Antimicrob. Agents Chemother.* **35**, 234–236.
- Kabsch, W. (2010). *Acta Cryst. D* **66**, 125–132.
- Kayser, L., Tang, X., Wemakor, E., Sedding, K., Hennig, S., Siebenberg, S. & Gust, B. (2011). *Chembiochem*, **12**, 477–487.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Müller, C., Nolden, S., Gebhardt, P., Heinzelmann, E., Lange, C., Puk, O., Welzel, K., Wohlleben, W. & Schwartz, D. (2007). *Antimicrob. Agents Chemother.* **51**, 1028–1037.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst. D* **67**, 355–367.
- Rackham, E. J., Grüşchow, S., Ragab, A. E., Dickens, S. & Goss, R. J. M. (2010). *Chembiochem*, **11**, 1700–1709.
- Sampaleanu, L. M., Yu, B. & Howell, P. L. (2002). *J. Biol. Chem.* **277**, 4166–4175.
- Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst. D* **55**, 191–205.
- Vértsey, L., Ehlers, E., Kogler, H., Kurz, M., Meiwes, J., Seibert, G., Vogel, M. & Hammann, P. (2000). *J. Antibiot.* **53**, 816–827.
- Winn, M., Goss, R. J. M., Kimura, K. & Bugg, T. D. (2010). *Nat. Prod. Rep.* **27**, 279–304.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.
- Zhang, W., Ostash, B. & Walsh, C. T. (2010). *Proc. Natl Acad. Sci. USA*, **107**, 16828–16833.