

Structural Studies on Proteins from Pathogenic Bacteria

**Structure and Function Investigation of the Lipoprotein
BPSL1204 from the Human Pathogen
*Burkholderia pseudomallei***

A thesis submitted in part fulfilment of the requirements for the
degree of Doctor of Philosophy

By Richard Michael Salmon

BSc (Hons) Biochemistry and Genetics, University of Sheffield



Department of Molecular Biology and Biotechnology
University of Sheffield
September 2012

Abstract

Burkholderia pseudomallei is a Class III bacterial pathogen of humans and the causative agent of melioidosis, a disease endemic to South East Asia and Australia. This work describes X-ray crystallographic and biochemical studies of the BPSL1204 protein from *B. pseudomallei*. The amino acid sequence of BPSL1204 possesses an N-terminal glyceryl-cysteine lipid-anchoring motif for covalent association with a membrane. After genetic truncation of the anchoring sequence, the BPSL1204 truncate was cloned, over-expressed and purified from *E. coli* before being subjected to crystallisation trials.

The 1.05 Å resolution crystal structure of BPSL1204 was solved by Selenomethionine MAD and revealed a two-domain fold similar to that of β-Lactamase Inhibitory Protein I (BLIP-I). Like BLIP-I, the BPSL1204 structure consists of two similar domains. However TEM1 β-lactamase assays showed BPSL1204 to lack the general inhibitory activity of BLIP-I.

Pull-down assays to uncover biological binding partners of the BPSL1204 homologue BCAL2351 from the paraphyletic Class II *B. cenocepacia* were undertaken. Although a methionyl-tRNA synthetase was isolated, no complex was seen to form between pure samples of the proteins. Antibodies raised against pure BPSL1204 were used to locate the BCAL2351 homologue in fractionated *B. cenocepacia*. BCAL2351 was found exclusively in the purified outer membrane by immunoblot analysis.

Some evidence of binding of antibodies to whole *B. cenocepacia* cells was observed by fluorescence microscopy, indicative of external BCAL2351 and BPSL1204. However, other slides containing samples from the same culture showed no such binding, suggestive of internal or absent protein. Surface presence may be cell-cycle dependant. However, no immune response was found when BPSL1204 was challenged with anti-*B. mallei* human sera.

Thus by a cross-disciplinary approach to investigating the structure and function of a protein of unknown function from the pathogenic *B. pseudomallei*, it was concluded that BPSL1204 is a protein with a BLIP-like fold potentially located on either the periplasmic or possibly outer face of the outer membrane of the bacterium.

Acknowledgments

Firstly, I would like to thank my supervisor Professor Pete Artymiuk for giving me the opportunity to take my first steps in the world of scientific research. For years of guidance and patience I am truly grateful. In addition, the assistance and advice on protein purification supplied by Dr Sveta Sedelnikova has been invaluable for this project. Furthermore, crystallographic counselling from Dr Pat Baker and Dr John Rafferty and assistance with microbiological issues from Dr Mark Thomas and Professor Dave Kelly has been instrumental in the experiments undertaken for this thesis. On a professional note, thanks go to the BBSRC for funding.

Many thanks must also go to all members of the X-ray Crystallography group, both past and present, for helping to provide sources of information and a great working environment. In particular, I cannot thank Dr Glyn Hemsworth enough for the various tips and help given to me in the latter stages of my undergraduate studies and the early years of my PhD work. Also, Dr Matt Day has always offered a sympathetic ear and been happy to bounce ideas around, even in the small hours of synchrotron trips. In addition, many thanks go to Fiona Rodgers for her tireless work in keeping the lab stocked and in running order.

Over my years in Sheffield I have made a great many friends, luckily too many to mention here individually. To all, cheers. However, mixing work and pleasure has led to a few also contributing to my research. My friend for seven years and counting, Vicki Kent has helped immensely with microscopy work, and for that I am thankful. Not just for introducing me to hospital breakfasts, I thank Dr Richard Jones as well for allowing me to repeatedly take over both his time and his lab bench for 'just one more' microbiology experiment, especially when his own thesis deadline was looming. I also say cheers to Denny (Dr Peter Drake) who has always supplied me with bits and bobs when my wits were at an end, and with whom having discussions and laughs over a pint or two has kept me relatively sane.

Amongst these friends, I have been lucky enough to have my brother Rob. Also taking up the science mantle in Sheffield, he has helped me immensely, equally through frank opinion on proposed experiments and also by being equipped with our hereditary iron liver, which has been handy in both times of strife and success. Which brings me to the rest of my family: my mum Kate, my dad Paul and my sister Lucy. They have all been there for me, offering support, driving over to visit, giving me a place to relax when needed and giving me a kick up the behind when necessary. Thank you for listening and nodding in the right places, I'm very lucky to belong to such a family.

I also extend my gratitude to Chris Parker, whose kindness led her to give me a roof whilst I finished my thesis, and whose kind words gave me the best environment in which to write. To which, I also add thanks to Tommy Parker, whose camaraderie and sense of humour has been an appreciated avenue of respite from the writing process.

Finally, I am totally indebted to my wonderful girlfriend Siân. She has been by my side during my entire PhD experience, and has supported and encouraged me throughout. Even when faced with her own trials of becoming a teacher, she has always helped me bounce back from the disappointments inherent with research and has celebrated with me when things went well. I apologise for times and events I may have missed due to late nights or weekends in the lab, but I hope that the work represented in this thesis shows that it has paid off for both of us. Thank you so much for being there when I needed you, I am eternally grateful.

Table of Contents

Chapter 1: Introduction	1
1.1 <i>Burkholderia pseudomallei</i>	1
1.1.1 An Introduction to Melioidosis.....	1
1.1.2 <i>Burkholderia pseudomallei</i> is the Causative Agent of Melioidosis	3
1.2 Membrane-Anchored Proteins	9
1.2.1 Characteristics of Lipoproteins and Membrane Association	9
1.2.4 General Functions Associated With Bacterial Lipoproteins	14
1.2.5 Examples of Lipoproteins in <i>B. pseudomallei</i>	14
1.3 Aims of the Research	15
1.3.1 Bioinformatical Discovery of a Novel, Membrane-associated Target	15
1.3.2 Prediction of Unknown Function From Structure	17
1.3.3 Support of Structural Findings with Experimental Data.....	17
1.3.4 Potential Medical Applications.....	19
Chapter 2: X-ray Crystallographic Theory	20
2.1 Introduction	20
2.2 Protein Crystallisation	20
2.2.1 Introduction.....	20
2.2.2 Vapour Diffusion.....	21
2.2.3 Batch Diffusion	25
2.2.4 Dialysis.....	25
2.2.5 Optimisation of Crystallisation.....	26
2.3 Generation of X-rays	26
2.3.1 Rotating Anodes.....	26
2.3.2 Synchrotron Radiation	28
2.3.3 Detection of X-rays.....	29
2.4 Basics of X-ray Diffraction	30
2.4.1 Braggs Law	30
2.4.2 Structure Factors and B-Factors.....	32
2.5 Data Processing	36
2.5.1 Indexing.....	36
2.5.2 Integration.....	37
2.5.3 Scaling and Data Reduction.....	37
2.6 Determination of Phases	39
2.6.1 The Patterson Function	39
2.6.2 SHELX and Direct Methods	40
2.6.3 MAD and Anomalous Scattering	41
2.6.4 Friedel's Law	44
2.6.5 Calculation of Heavy Atom Substructure Phases.....	46
2.6.6 Phasing the Protein Structure.....	47
2.7 Calculations and Modification of Electron Density	49
2.7.1 Calculation of Electron Density	49
2.7.2 Solvent Flattening	50
2.7.3 NCS Averaging	51
2.8 Structure Refinement	52
2.8.1 The Maximum-Likelihood Function	53
2.8.2 Validation of Refinement by R_{work} and R_{free}	55

Chapter 3: Materials and Methods	57
3.1 Materials.....	57
3.2 General Methods.....	57
3.2.1 PCR.....	57
3.2.2 Agarose Gel Electrophoresis.....	58
3.2.3 SDS-PAGE.....	59
3.2.4 Determination of Protein Concentration	60
3.3 Bioinformatical Methods	61
3.4 Cloning of Constructs	61
3.4.1 Oligonucleotide Primers.....	61
3.4.2 Sub-cloning of PCR products	62
3.4.3 <i>E. coli</i> Strains Used For Sub-cloning and Over-expression.....	63
3.5 Protein Expression.....	64
3.5.1 Media and Agar.....	64
3.5.2 Minimal Media for Seleno-methione Incorporation.....	64
3.5.3 <i>E. coli</i> Over-expression – General Overview	66
3.5.4 Glycerol Stocks	67
3.6 Protein Purification.....	67
3.6.1 Preparing Cell-free Extract.....	68
3.6.2 Ion-Exchange Chromatography	68
3.6.3 Ammonium Sulphate Precipitation	69
3.6.4 Gel-Filtration	69
3.6.5 Viva-Spin Concentration.....	70
3.7 Crystallographic methods	70
3.7.1 Initial Screening	70
3.7.2 Siliconisation of Coverslips	71
3.7.3 Condition Optimisation.....	71
3.7.4 Mounting of Crystals onto an X-ray Source.....	72
3.7.5 Cryo-protection of Crystals	73
3.7.6 Metal-Soaking of Crystals	73
3.7.7 Data Collection.....	74
3.7.8 Data Processing.....	74
3.8 Assays	76
3.8.1 β -lactamase Inhibition Assays	76
3.8.2 Affi-Gel15 Pull-Down Assays.....	76
3.9 Localisation and Antigenic Studies.....	78
3.9.1 Fractionation of <i>Burkholderia cenocepacia</i>	78
3.9.2 Buffers used in Immuno-blotting	80
3.9.3 Western Blotting.....	81
3.9.4 Immuno-blotting.....	82
3.9.5 Chemiluminescence Assays.....	82
3.9.6 Preparation of Slides for Fluorescence Microscopy	84
3.9.7 Fluorescence Microscopy Image Collection and Processing	85
Chapter 4: Target-Selection, Over-expression and Purification of BPSL1204	86
4.1 Bioinformatical Selection of BPSL1204	86
4.1.1 Analysis of Hydropathy Plots	86
4.1.2 Selection of Potentially Novel Folds.....	86
4.1.3 Further Exclusion of Targets from the Short-List	88
4.1.4 Final Selection of BPSL1204 and Truncation.....	89
4.2 Cloning of BPSL1204-MA	89
4.2.1 Primer Design and PCR of Truncated BPSL1204	89
4.2.2 Blunt-End Ligation into pETBlue-1	90

4.2.3 Sub-cloning into NovaBlue	93
4.2.4 Identification of Successful Clones.....	93
4.2.5 Cloning of Full-length BPSL1204.....	94
4.3 Over-expression of BPSL1204-MA.....	96
4.3.1 Transformation into Tuner Cells	96
4.3.2 Over-expression and Solubility Tests.....	96
4.3.3 Large-scale Over-expression	97
4.3.4 Over-expression of Full-length BPSL1204	99
4.4 Purification of BPSL1204-MA.....	101
4.4.1 Fractionation of Cell-Paste	101
4.4.2 DEAE Ion-Exchange Chromatography	101
4.4.3 Ammonium Sulphate Precipitation	103
4.4.4 Gel Filtration.....	103
4.4.5 N-terminal Sequencing	105
4.5 Synthesis of a Selenomethionine Derivative.....	105
4.5.1 Large-scale Over-expression and Purification of SeMet BPSL1204-MA	105
4.5.2 Analysis of SeMet Incorporation.....	109
Chapter 5: Crystallisation and Structure Determination of BPSL1204	111
5.1 Protein Sample Preparation.....	111
5.1.1 Buffer-exchanging the Final Sample	111
5.1.2 Concentration of Protein Sample for Crystallography	111
5.2 Automated Screening of Crystallisation Conditions.....	112
5.2.1 Creation of Sitting-drop Robot Trials	112
5.2.2 Incubation of Screens	112
5.2.3 Examination of Growth Conditions	113
5.2.4 Selection of Potential Hits.....	113
5.3 Optimisation to Hanging-drop Trays.....	113
5.3.1 Broad-range Screening Around Potential Hits	113
5.3.2 Focussing onto Optimal Growth Conditions.....	115
5.3.3 Production of High Quality BPSL1204-MA Crystals.....	116
5.4 In-house Diffraction Studies.....	116
5.4.1 Establishing Cryo-buffers	116
5.4.2 Mounting Individual 1204-MA Crystals	118
5.4.3 Detection of Diffraction.....	118
5.4.4 Space-group, Cell-refinement and Collection Strategy Calculation.....	119
5.5 Collection and Analysis of Native Diffraction Data.....	119
5.5.1 Collection of Diffraction Images Using a Synchrotron Light Source	119
5.5.2 Indexing, Integration and Reduction of Native Data.....	121
5.5.3 Scaling of Data, Resolution Limits and Data Quality	121
5.6 Heavy-atom Soaks and Initial Phasing Attempts.....	123
5.6.1 Heavy-atom soaks	123
5.6.2 Overview of Collection of MAD data and Anomalous Signal Detection	124
5.7 Crystallisation of a Selenomethionine Derivative.....	125
5.7.1 Preparation of Protein Samples	125
5.7.2 Attempts to Repeat Native Conditions	125
5.7.3 Original Streak-Seeding Experiments	125
5.7.4 Serial Dilution of Seeding.....	126
5.7.5 Confirmation of Crystallised Seleno-met Protein.....	128
5.8 Collection of MAD Data Using Synchrotron Light	128
5.8.1 Detection of Selenium Fluorescence and Determination of X-ray Wavelengths for MAD Data Collection	128
5.8.2 Data Collection of Peak, Inflection and High-Energy Remote Data	130
5.9 Processing of MAD Data for Phasing.....	132

5.9.1 Indexing, Integrating and Scaling Measured Reflections	132
5.9.2 Detection of Anomalous Signal and Phasing	132
5.9.3 Automatic Model Building.....	137
5.10 Refinement of the BPSL1204 Structure.....	137
5.10.1 Fitting of the Complete Model.....	137
5.10.2 Preventing Bias in the Model.....	139
5.10.3 Fitting of Ligands and Water Molecules	139
5.10.4 Phase Extension to 1.05 Å Resolution	142
5.10.5 Refinement and R-factors	142
5.11 Structure Validation	143
5.11.1 Molprobity	143
5.11.2 PROCHECK.....	147
Chapter 6: Structural Investigation of BPSL1204 Function.....	150
6.1 Structural Features of BPSL1204	150
6.2 DALI analysis of BPSL1204-MA.....	158
6.2.1 The DALI server	158
6.2.2 Results of Dali Search	158
6.3 Comparisons of BPSL1204-MA and BLIP-I.....	161
6.3.1 Sequence Homology	161
6.3.2 Similarities of Individual Domains	161
6.3.3 Differential Arrangement of Tertiary Structure.....	164
6.3.4 Crystal Packing Versus Biological Conformation.	169
6.3.5 Potential Interactions With TEM-1 Target.....	171
6.3.6 Potential Explanations for Structural Similarity of BPSL1204 to BLIP-I	179
6.4 <i>E. coli</i> TEM-1 Inhibition Assay	181
6.4.1 Assay Design and Data Collection	181
6.4.2 Inhibition Assay Results and Observations.....	181
6.4.3 Conclusions of the TEM1 Inhibition Assay.....	183
6.5 Implications of Structural Bioinformatics	183
Chapter 7: Biological Investigation of BPSL1204 Function	184
7.1 Design of a Pull Down Assay	184
7.1.1 <i>Burkholderia cenocepacia</i>	184
7.1.2 BPSL1204 and BCAL2351.....	184
7.1.3 Aims of the Assay.....	185
7.2 N-terminal His-tagged BCAL2351	185
7.2.1 pET28a Cloning and Transformation	185
7.2.2 Insolubility in BL21 (DE3) Cells.....	187
7.2.3 Inclusion Body Preparation and Refolding Assays.....	187
7.3 Affi-Gel15 Pull Down Assay.....	191
7.3.1 Design of a Second Pull-down Method.....	191
7.3.2 Culturing <i>B. cenocepacia</i> and Safe Practice	191
7.3.3 First Pull-Down Assay	192
7.3.4 Analysis of Isolated Fractions By SDS-PAGE	192
7.3.5 Second Pull-Down Assay	193
7.3.6 Identification of Proteins ‘Pulled-Out’ of Cell Extract.....	196
7.3.7 Choice of Target Protein.....	198
7.4 BPSL0998	199
7.4.1 Novel Features of BPSL0998	199
7.4.2 pBAD/hisB Cloning of BPSL0998	202
7.4.3 Insolubility in Top10 Cells.....	202
7.5 Attempts to Purify BCAL2646 by Bulk Pull Down	204
7.5.1 BCAL2646 and BPSL0998 are Homologs.....	204
7.5.2 Log Phase and Overnight Culturing of <i>B. cenocepacia</i>	204

7.5.3 Scaling up of the Assay	204
7.5.4 Limitations of Purifying by Bulk Assay	206
7.5.5 Heparin Column Purification of BPSL2646.....	209
7.5.6 Analysis of BPSL1204: BCAL2646 Complex Formation	210
7.6 Conclusions.....	210
Chapter 8: Cellular Localisation of BPSL1204 and the Discovery of Potential Immunogenic Targets	214
8.1 Preparation For Localisation Study	214
8.1.1 Supposition Behind Study	214
8.1.2 Generation of Polyclonal Antibodies and Rat Hyper-immune Response to BPSL1204	214
8.1.3 Isolation of <i>B. cenocepacia</i> Cellular Fractions	215
8.2 Immunochemiluminescence Assays for Cellular Location of BPSL1204.....	218
8.2.1 Experimental Procedure	218
8.2.2 Results of the Experiment.....	220
8.2.3 Conclusions Drawn From the Experiment	224
8.3 Immunofluorescence Microscopy of the <i>B. cenocepacia</i> Outer Membrane Protein BCAL2351 (BPSL1204)	224
8.3.1 Reasoning for the Fluorescence Microscopy Experiment	224
8.3.2 Experimental Procedure	225
8.3.3 Preliminary Results of Fluorescence Microscopy	226
8.3.4 Conclusions and Limitations of Microscopy Results.....	229
8.4 Investigation into Immunogenic Targets	232
8.4.1 Experimental Theory	232
8.4.2 Components of the Assay	232
8.4.3 Experimental Procedure	233
8.4.4 Results of the Experiment.....	234
8.4.5 Mass Spectrometry of a Potential Antigen or Diagnostic Marker	237
8.4.6 Conclusions of Immunogenic Studies	237
Chapter 9: Conclusions and Suggestions for Future Work	240
References.....	247

List of Figures and Tables

Figure 1.1: The 7 morphotypes of <i>B. pseudomallei</i>	2
Figure 1.2: Global distribution of melioidosis.....	2
Figure 1.3: Manifestations of <i>B. pseudomallei</i> infections.....	4
Figure 1.4: Proteins interactions with a cell membrane.....	10
Figure 1.5: The +2 lipoprotein-sorting rule	12
Figure 1.6a: The chemical structure of a glycerylcysteine anchor.....	13
Figure 1.6b: Processing of prolipoprotein to glyceryl-cysteine lipid- anchored lipoprotein.....	13
Figure 1.7: Hydropathy plots of <i>B. pseudomallei</i> proteins.....	16
Figure 1.8: Growth of the PDB.....	18
Figure 2.1: Crystallisation phase diagram.....	22
Figure 2.2: The set-up of crystallisation experiments	23
Figure 2.3: The Bragg Condition.....	31
Figure 2.4: The Ewald Construction.....	33
Figure 2.5: Argand diagram calculation of structure factors.....	35
Figure 2.6: X-ray absorption plot.....	43
Figure 2.7: Breaking Friedel's Law.....	45
Figure 2.8: Phase solution using anomalous scattering and Harker plots..	48
Figure 4.1: Analysis of hydropathy plots.....	87
Figure 4.2: Amino acid sequence of BPSL1204 and truncation.....	91
Figure 4.3: Pure BPSL1204-MA PCR product.....	92
Figure 4.4: Colony-PCR of Novablue: pETBlue: BPSL1204 colonies.....	95
Figure 4.5: Tuner BPSL1204-MA over-expression.....	98
Figure 4.6: Full-length BPSL1204 over-expression.....	100
Figure 4.7: DEAE ion exchange of BPSL1204-MA.....	102
Figure 4.8: (NH ₄) ₂ SO ₄ -precipitation of BPSL1204-MA.....	104
Figure 4.9: Gel filtration trace of BPSL1204-MA purification.....	106
Figure 4.10: Analysis of the BPSL1204-MA purification protocol.....	107
Figure 4.11: Analysis of Seleno-met BPSL1204-MA over-expression.....	108
Figure 4.12: Mass spectra of Se-Met BPSL1204-MA.....	110
Table 5.1: Potential crystallisation conditions.....	114
Figure 5.1a: Condensed BPS1204 rod-cluster.....	117
Figure 5.1b: Cluster of thick BPSL1204 rods.....	117

Figure 5.2: BPSL1204-MA Diffraction pattern.....	120
Figure 5.3: SCALA output file.....	122
Figure 5.4: Streak-seeded seleno-met BPSL1204 crystals.....	127
Figure 5.5: Crystals of seleno-met BPSL1204-MA.....	129
Figure 5.6: Seleno-met BPSL1204-MA peak (f') and inflection (f'') points.....	131
Figure 5.7a: Output from ShelxC.....	133
Figure 5.7b: Output from ShelxD.....	133
Figure 5.8: ShelXE heavy-atom site density calculations.....	135
Figure 5.9: Investigation of the the .phs outputs from ShelxE	136
Figure 5.10: Electron density prior to automated model building.....	138
Figure 5.11: Correction of ARP/wARP model building.....	140
Figure 5.12: Model bias.....	141
Table 5.2: Data collection, phasing and refinement statistics.....	144
Figure 5.13: Output from the Molprobit analysis.....	146
Figure 5.14: Procheck analyses.....	148
Figure 5.15: PROCHECK Ramachandran analysis.....	149
Figure 6.1: The 1.05 Å resolution structure of BPSL1204.....	151
Figure 6.2: Secondary structure and topology of BPSL1204.....	152
Figure 6.3a: Multiple conformations of methionine residue 33.....	153
Figure 6.3b: Location and flexibility of methionine 33.....	153
Figure 6.4: The interdomain interface of BPSL1204.....	154
Figure 6.5: The major bonds in the BPSL1204 interdomain region.....	156
Figure 6.6: B-factors in the loop region L1.....	157
Figure 6.7: Superimposition of BPSL1204 with OsmE.....	159
Figure 6.8: Alignment of full-length BPSL1204 with full-length BLIP-I.....	162
Figure 6.9: Dali superimposition of BLIP-I on BPSL1204.....	163
Figure 6.10: Superimpositions of BLIP-I and BPSL1204 domains.....	165
Figure 6.11: Self-superimpositions of BPSL1204 and BLIP.....	166
Figure 6.12: The potential for an extra N-terminal α -helix in BPSL1204.....	167
Figure 6.13: Comparison of BPSL1204 and BLIP-I domain arrangement.....	168
Figure 6.14: Superimposition of BPSL1204 on a BLIP-I:TEM1 complex.....	170
Figure 6.15: Analysis of BCAL2351-MA over-expression.....	172
Figure 6.16: Sequence alignment of BPSL1204 with BCAL2351.....	173
Figure 6.17: Inhibition-complex of TEM-1 with the N-terminal domain of BLIP-I superimposed with the N-terminal domain of BPSL1204.....	174

Figure 6.18: Conformational change during BLIP-I binding TEM-1.....	176
Figure 6.19: Comparison of BLIP-I and BPSL1204 inhibitory loop conformations.....	176
Figure 6.20: Alignment of inhibiting loops from multi-species BLIP-Is.....	177
Figure 6.21: Investigation of TEM-1 active site binding by BLIP-I, and corresponding BPSL1204 binding potential.....	178
Figure 6.22: Structural comparison of a β-lactamase with a transpeptidase.....	180
Figure 6.23: The chemical structure of Nitrocefin.....	182
Figure 6.24: Graph of Nitrocefin hydrolysis.....	182
Figure 7.1: Sequence alignment of BPSL1204 and BCAL2351 truncation.....	186
Figure 7.2: N-his BCAL2351-MA PCR.....	188
Figure 7.3: Analysis of N-hisBCAL2351-MA over-expression.....	188
Figure 7.4: Analysis of QuickFold refolding.....	190
Figure 7.5: Gel-filtration trace of N-hisBCAL2351 post-refolding.....	190
Figure 7.6: Analysis of first pull-down.....	194
Figure 7.7: Analysis of second pull-down.....	195
Figure 7.8: MASCOT server analysis of mass spectroscopy.....	197
Figure 7.9: Sequence alignment of BCAL2646 with BPSL0998.....	200
Figure 7.10: Alignment of BPSL0998 with 3H9C MetRS and Trbp111.....	201
Figure 7.11: Analysis of BPSL0998 cloning stages.....	203
Figure 7.12: Analysis of BPSL0998 over-expression.....	205
Table 7.1: Retrieval of BCAL2646 from bulk pull-downs.....	207
Figure 7.13: Analysis of bulk pull-down assays.....	208
Figure 7.14: Heparin column trace of BCAL2646 purification.....	211
Figure 7.15: Analysis of BCAL2646 heparin column purification.....	211
Figure 7.16: Gel filtration trace of BPSL1204:BCAL2646 complex.....	212
Figure 7.17: Analysis of complexing BPSL1204 with BCAL2646.....	212
Figure 8.1: Test ELISA of anti-BPSL1204 immune response in rats.....	216
Figure 8.2: Final ELISA of anti-BPSL1204 immune response in rats.....	217
Figure 8.3: Analysis of <i>B. cenocepacia</i> fractionation.....	219
Figure 8.4: Analysis of Western blot transfer of BPSL1204 and <i>B. cenocepacia</i> fractions.....	221
Figure 8.5: BPSL1204 and <i>B. cenocepacia</i> fractions immuno-blotted with anti-BPSL1204.....	222
Figure 8.6: Secondary antibody control of BPSL1204 localisation.....	223

Figure 8.7: Pre-immune sera control of BPSL1204 localisation.....	223
Figure 8.8: Fluorescence microscopy of FITC-tagged BCAL2351 in the shattered outer membrane.....	227
Figure 8.9: Results of whole-cell fluorescence microscopy.....	228
Figure 8.10: Differences between pure and <i>in vivo</i> outer membrane.....	231
Figure 8.11: Analysis of Western blot transfer of BPSL1204 and <i>B. cenocepacia</i> fractions for immunogenicity studies.....	235
Figure 8.12: Immunoblot analysis of BPSL1204 and <i>B. cenocepacia</i> fraction immunogenicity.....	236
Figure 8.13: Mass spectroscopic investigation of a ~40 kDa OM protein..	238

List of Abbreviations

ADSC – Area Detector Systems Corporation
APS – Ammonium persulphate
ASU – Asymmetric unit
BBSRC – Biotechnology and Biological Sciences Research Council
BLAST – Basic Local Alignment Search Tool
BLIP - β -lactamase inhibitory protein
BPSL – *Burkholderia pseudomallei* Large chromosome
BPSS - *Burkholderia pseudomallei* Small chromosome
B. cenocepacia/mallei/pseudomallei – *Burkholderia* species
BCAL - *Burkholderia cenocepacia* annotation Large chromosome
BCAM - *Burkholderia cenocepacia* annotation Medium chromosome
BSA – Bovine Serum Albumin
CC – Correlation Coefficient
CCD – Charge-Coupled Device
CCP4 – Collaborative Computational Project No. 4
CDC – Centre for Disease Control
CF – Cystic Fibrosis
CFE – Cell-free Extract
DEAE - Diethylaminoethyl
DEAE-FF – DEAE Fast-Flow
DMSO – Dimethyl sulphoxide
DNA – Deoxyribonucleic acid
DPX – Distyrene-plasticizer(tricresyl phosphate)-xylene
DSLS – Diamond Synchrotron Light Source
DTT - Dithiothreitol
ECL – Enhanced Chemiluminescence reagent
E. coli – *Escherichia coli*
EDTA – Ethylenediaminetetraacetic acid
ELISA – Enzyme-linked Immunosorbant Assay
FITC – Fluorescein isothiocyanate
FT – Fourier Transform
FT⁻¹ – Inverse Fourier Transform
GF – Gel Filtration
GFP – Green Fluorescent Protein
GPI - Glycophosphatidylinositol
GUI – Graphical User Interface
HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP – Horse-radish Peroxidase
ID - Identity
IPTG – Isopropyl β -D-1-thiogalactopyranoside
ITC – Isothermal Titration Calorimetry
LB – Lysogeny Broth
LDS – Lithium Dodecyl Sulphate
LSP – Lipoprotein Signal Peptidase
MA – Membrane Anchor
MAD – Multiwavelength Anomalous Dispersion
MALDI-TOF – Matrix-assisted laser desorption/ionisation time-of-flight

MBB – Molecular Biology and Biotechnology
 metRS – methionyl tRNA synthetase
 ML – Maximum Likelihood
 MORU – Mahidol Oxford Research Unit
 MPD – 2-methyl-2,4-pentandiol
 MR – Molecular Replacement
 MRC – Medical Research Council
 MS – Mass Spectrometry
 MSC – Microbiological Safety Cabinet
 MSE – Measuring and Scientific Equipment
 M-T-PBS – Milk-Tween20-PBS
 MW – Molecular Weight
 NCBI – National Center for Biotechnological Information
 NCS – Non-Crystallographic Symmetry
 NEB – New England Biolabs
 NTA – Nitrotriacetic acid
 NTP – Nucleotide triphosphate
 OD – Optical Density
 OM – Outer Membrane
 PBP – Penicillin-binding Protein
 PBS – Phosphate-buffered Saline
 PC – Personal Computer
 PCR – Polymerase Chain Reaction
 PDB – Protein Data Bank
 PEG - Polyethyleneglycol
 POW – Prisoner of War
 RADAR – Rapid Automatic Detection and Alignment of Repeats
 RMSD – Root-mean-square deviation
 RNA – Ribonucleic acid
 RONN – Regional Order Neural Network
 RND – Resistance-nodulation-division
 SAD – Single-wavelength Anomalous Dispersion
 Sarkosyl – Sodium lauroyl sarcosinate
S. clavigulus/exfoliates – *Streptomyces* species
 SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis
 Seleno-met – Selenium-derivative of methionine
S. mutans – *Streptococcus mutans*
 STE – Sucrose-Tris-EDTA buffer
 T4 Ligase – Ligase from bacteriophage T4
 T7F/R – Forward/Reverse DNA sequence for the bacteriophage T7 promoter
 TAE – Tris-Acetic acid-EDTA buffer
 TEMED - Tetramethylethylenediamine
 TLS – Translation/Libration/Screw
 TMB – 3,3',5,5'-Tetramethylbenzidine
 T-PBS – Tween20-PBS
 Tris – Tris(hydroxymethyl)aminomethane
 tRNA – Transfer Ribonucleic Acid
 UK – United Kingdom
 UV - Ultraviolet

List of Units of Measurement

Volume

L – litre (1)

ml – millilitre (1×10^{-3} L)

μ l – microlitre (1×10^{-6} L)

nl – nanolitre (1×10^{-9} L)

Mass

Da – Dalton

kDa – kilo Dalton

bp – base-pairs

kbp – kilo base-pairs

kg – kilogram (1×10^3 g)

g – gram (1)

mg – milligram (1×10^{-3} g)

μ g – microgram (1×10^{-6} g)

ng – nanogram (1×10^{-9} g)

pg – picogram (1×10^{-12} g)

Distance

m – metre (1)

cm – centimetre (1×10^{-2} m)

mm – millimetre (1×10^{-3} m)

μ m – micrometre (1×10^{-6} m)

nm – nanometre (1×10^{-9} m)

Å – Angström (1×10^{-10} m)

Time and Speed

yr – year

hr – hour

min – minute

s – second

rpm – revolutions per minute

krpm – kilo revolutions per minute

Concentration

M – molar (1)

mM – millimolar (1×10^{-3} M)

μ M – micromolar (1×10^{-6} M)

nM – nanomolar (1×10^{-9} M)

[x] – concentration of x

Temperature

°C – degrees Celsius

K – Kelvin

Chapter 1: Introduction

This thesis describes structural studies on a lipoprotein from the pathogen *Burkholderia pseudomallei*. In the first part of this chapter I describe the background to this bacterium. In the second part aspects of the structure and function of lipoproteins are discussed.

1.1 *Burkholderia pseudomallei*

1.1.1 An Introduction to Melioidosis

Melioidosis, from the Greek 'melis' (distemper of asses) and 'eidos' (resemblance), is an infectious disease caused by *Burkholderia pseudomallei*, a Gram-negative, bipolar aerobe that is pathogenic to humans and animals (Aldous, 2005). *B. pseudomallei* was first identified as the causative agent of the disease by the English pathologist Major Alfred Whitmore in opium addicts in Rangoon (Burma/Myanmar)(Cheng & Currie, 2005). It has since been realised that this motile, 2-5µm-long, rod-shaped bacterium exists in up to 7 'morphotypes' (Stone, 2007) (Figure 1.1) and is endemic most notably to South-East Asia and Northern Australia (Figure 1.2) (Dance, 2000). Dubbed 'The Great Mimicker' by some (Wiersinga et al, 2006), this bacterium can reside within its host before exhibiting symptoms anywhere between 2 days to 62 years (Ngauy et al, 2005; Stone, 2007), and even when symptoms occur the disease is often misdiagnosed as TB (Vidyalakshmi et al, 2008) or even cancer (Reechaipichitkul, 2004). This is often due to the formation of abscesses in the spleen and lymph nodes that to an unaware eye may appear cancerous, but are in fact infectious cultures of *B. pseudomallei* (Saravu et al, 2012). The disease can also manifest itself as a septicaemic infection of the blood (Puthucheary et al, 1992), to many the most life-threatening and fast-acting incarnation of the disease of which 9 out of 10 Thai patients die with septic shock (Currie et al, 2000; Stone, 2007), as well as a pneumonic infection often associated with exposure to extreme-weather events or victims of near-drowning (Chaowagul et al, 1989). It has also been noted that

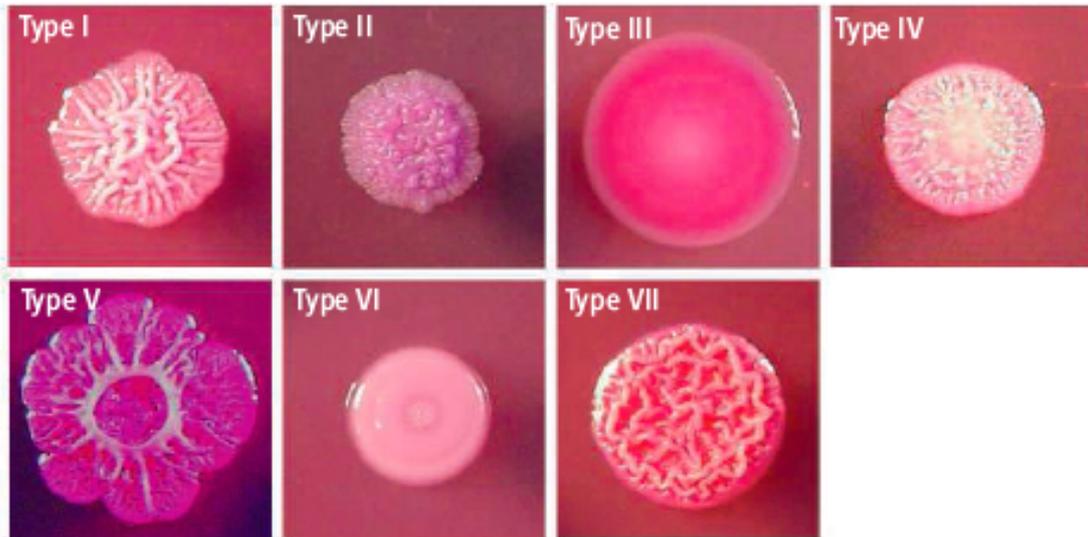


Figure 1.1: The 7 recorded morphotypes of *B. pseudomallei* of which Type I is the most prevalent. A wide range of morphologies further obscures identification of melioidosis upon culturing from infected patients. Adapted from Stone (Stone, 2007).

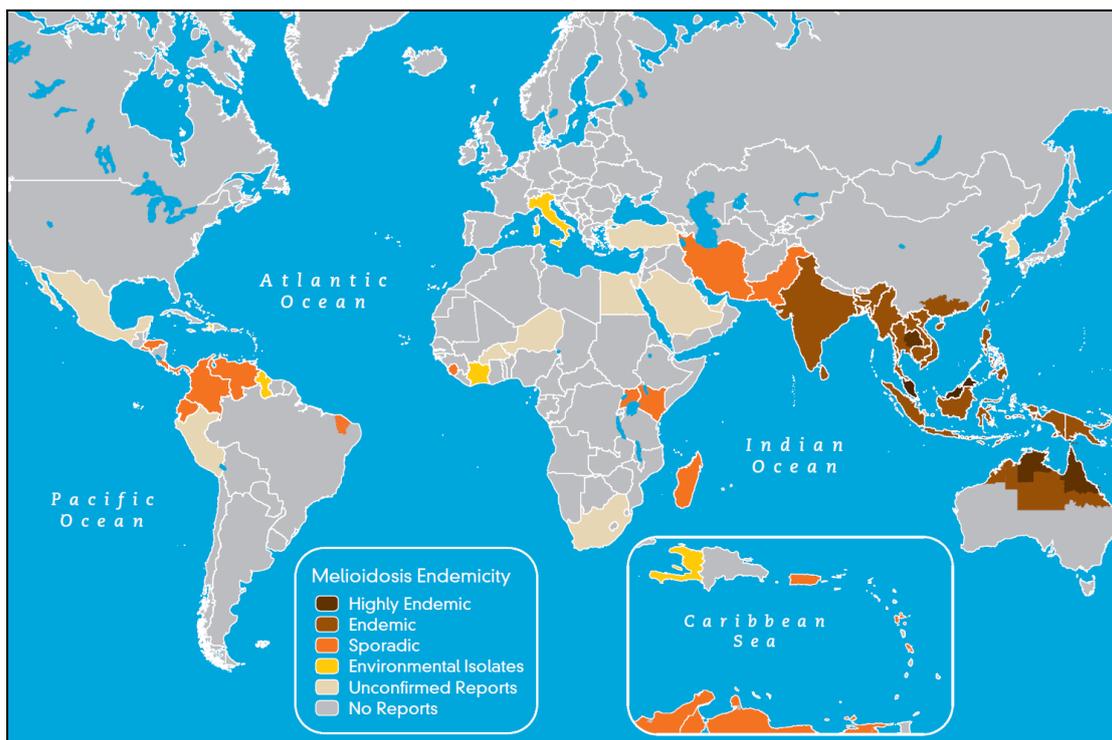


Figure 1.2: Global distribution of the occurrence of melioidosis, linked to regions of *B. pseudomallei* endemicity. Map supplied by the CDC, adapted from Cheng (Cheng & Currie, 2005).

patients that suffer from an onset of melioidosis have often a chronic co-illness such as diabetes (Wiersinga et al, 2006), kidney disease (Stone, 2007) and thalassemia (Cheng & Currie, 2005), the significance of which is still under investigation. Added to this, the organism has been seen to infect both phagocytic and non-phagocytic cells, whilst avoiding the immune-response, leading to the formation of multi-nucleated giant cells (Suparak et al, 2005) created using an actin-based membrane protrusion system (Kespichayawattana et al, 2000).

It is widely suspected (Ashdown, 1979) that infection is initiated by entering the body/blood-stream via cuts and abrasions in the feet and hands (such as with Thai paddy farmers) (Stone, 2007), via ingestion through contaminated foodstuffs (Galyov et al, 2010) or inhalation (such as US Vietnam-veterans breathing in helicopter up-draft) (Aldhous, 2005; Cheng & Currie, 2005) (Figure 1.3). Due to these characteristics, *B. pseudomallei* is considered a potential bio-terror threat, with its high infectivity based upon inhalation, and thus it carries potential as an aerosolized bio-weapon (Holden et al, 2004; Stone, 2007).

1.1.2 *Burkholderia pseudomallei* is the Causative Agent of Melioidosis

Burkholderia pseudomallei (also known as *Pseudomonas pseudomallei*) (Dance et al, 1989; Yabuuchi et al, 1992) is a bacterium pathogenic to humans, and causes the multi-symptom disease known as melioidosis (White, 2003). With low peptidoglycan content in its cell wall, *B. pseudomallei* is a gram-negative, soil-dwelling saprophyte that respire aerobically, and is commonly found to inhabit rice-paddy fields (Vongphayloth et al, 2012). *B. pseudomallei* can be differentiated from its paraphyletic cousin *Burkholderia mallei* (the causative agent of the cattle disease Glanders) (Lever et al, 2003; Tomaso et al, 2005) cytologically by *B. pseudomallei*'s motility on hanging-drop microscopy as *B. mallei* is a non-motile bacillus (Neubauer et al, 2005), by 24-48 hours visual growth on Ashdown's medium (*B. mallei* grows slower, taking about 72 hours) (Peacock et al, 2005) or physiologically by occurrence of the Strauss reaction in inoculated Guinea pigs, in which only *B. mallei* causes swelling of

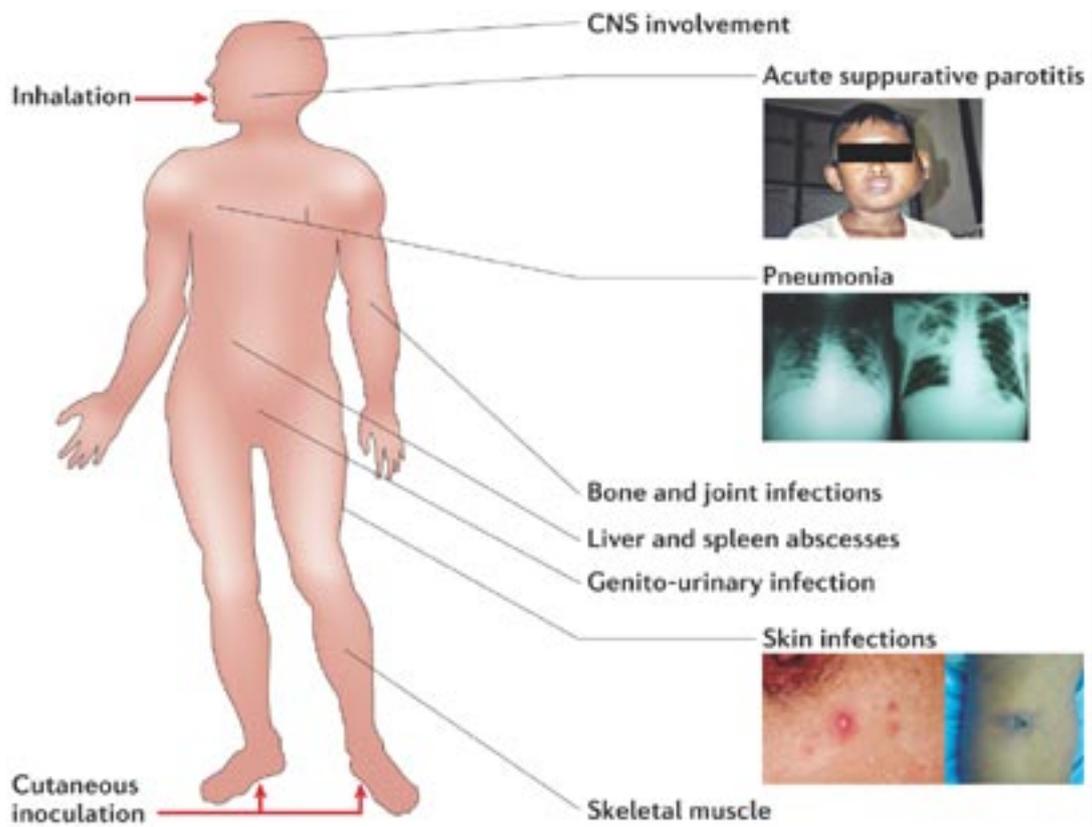


Figure 1.3: Manifestations of *B. pseudomallei* infections, highlighting physical presentation of melioidosis symptoms and proposed modes of infection. Image taken from Nature Reviews: Microbiology (Wiersinga et al, 2006).

the Guinea pig scrotum (Cheng & Currie, 2005; Scholz et al, 2006).

The 7.25 mega-base (Mb) genome of *B. pseudomallei* is composed of 2 chromosomes (Holden et al, 2004), of which BPSL (*Burkholderia pseudomallei* Large) is 4.07 Mb and BPSS (*Burkholderia pseudomallei* Small) is 3.17 Mb. In comparison, *B. pseudomallei*'s closest cousin *B. mallei* has a large chromosome of 3.5 Mb and a small chromosome of 2.3 Mb (Nierman et al, 2004). These two significantly larger chromosomes in *B. pseudomallei* also show some degree of partitioning in terms of the functional genes they contain, with the larger BPSL controlling metabolism and growth in majority and BPSS genes functioning more towards roles of adaptation and survival (Holden et al, 2004). A further point of interest in the genomic structuring of *B. pseudomallei* is the presence of 16 'genomic islands' that contribute up to 6.1% of the genome, yet are totally absent in the paraphyletic *B. mallei* (Holden et al, 2004). It has been seen that these GIs are highly variable between strains of *B. pseudomallei*, and thus actively transmissible, which may play a key role in the varying pathogenicity of strains to humans and animals as well as perhaps defining specificity for hosts compared to the relatively non human-pathogenic *B. mallei* and *B. thailandensis* or the plant pathogen *B. cenocepacia* (Holden et al, 2004).

Consistent with the locations of reported cases of melioidosis (Cheng & Currie, 2005), *B. pseudomallei* is endemic to quite specific areas of South-East Asia and the Northern Territories of Australia (Dance, 2000)(see Figure 1.2). A 'hotspot' for *B. pseudomallei* is Thailand (Aldhous, 2005), where the first case of melioidosis wasn't reported until 1955, but since 1986 the MORU lab (Mahidol-Oxford Tropical Medicine Research Unit) in Bangkok has documented more than 3000 cases (Stone, 2007) and Sappasitprasong Hospital in Ubon Ratchathani in north-east Thailand admits around 200 patients annually that are positive for *B. pseudomallei* infection, of which nearly half die (Aldhous, 2005). These figures may however be a mere shadow of a much larger number of fatalities, due to lower accessibility to appropriate healthcare in more peripheral areas where *B. pseudomallei* is still

endemic (Dance, 1991). Many cases admitted correlate with the suspected modes of acquisition, as rice farmers work in paddy fields bare-foot and become infected through cuts to the feet thus inoculating patients in an environment teeming with *B. pseudomallei*. This thinking is supported by the fact that when boots were made available to farmers in 2004 to reduce water-borne leptospirosis, the incidence rate of melioidosis dropped in areas such as Ubon Ratchathani where boots were supplied, but when farmers began to discard footwear in favour of traditional methods, the cases of melioidosis began to appear more rapidly (Stone, 2007).

Outside of areas of high endemicity in South-East Asia (Malaysia, Singapore and Thailand) and Northern Australia, there are relatively few reported cases of melioidosis (Cheng & Currie, 2005), although over time there appears to be a gradually emerging spread of the disease's occurrence in humans (Dance, 2000). Methods of measuring the extent of the disease are limited to published case-reports and sero-positivity sampling for antibodies against *B. pseudomallei* (Dance, 2000), even though the bacterium appears to be able to evade the immune response in most cases (Cheng & Currie, 2005). Sero-positivity studies have shown that in Vellore, India, 10.7% of villagers were sero-positive (Kang et al, 1996), whilst the southern Chinese provinces of Guangdong and Guangxi have exhibited sero-positivity of 3.8-15.2% (Li et al, 1994). Both these areas of India and China share sub-tropical status with the regions of established endemicity in areas such as north-east Thailand, but fall considerably short of the ~70% sero-positivity there (Stone, 2007). Any cases of *B. pseudomallei* infection reported farther afield exist as individual reports and appear to have originated from areas of higher occurrence such as the Indian or South-East Asian sub-continent. For example, of 15 UK cases since 1988 5 originated in Bangladesh and 2 from India and Pakistan (Dance et al, 1999), suggestive of 'imported' melioidosis rather than any claim to an endemic spread of environmental *B. pseudomallei* (Dance, 2000). Other examples of imported melioidosis include 'L'affaire du Jardin des Plantes' in the mid-1970s, in which an outbreak of *B. pseudomallei* infections occurred in a zoo in Paris killing many mammals, including several zoo keepers, which

was suspected to have occurred when a panda infected with the bacteria was donated from China (White, 2003).

While the threat of sustainable endemic spread of *B. pseudomallei* seems to be limited by specific environments (even within the Thai 'hot-spot' *B. pseudomallei* isn't prevalent in central Thailand where its relatively harmless cousin *B. thailandensis* presides, presumably due to soil conditions) (Aldhous, 2005), and in areas where the bacterium can be found to flourish it does so with great effect, and thrives in many conditions otherwise intolerable to other species. An example of exactly how hardy it is lies in its ability to live in double-distilled water, devoid of any sustenance, for 18 years and still form viable plate cultures (Moore et al, 1999). This has been compared to the survival rate of various other microbial species in nothing but distilled water (some closely related to *B. pseudomallei*, some not), of which only *B. pseudomallei* and *P. aeruginosa* survived beyond 30 days (Moore et al, 1999). This of course may lead to horrifying consequences when water supplies become contaminated, a mode of delivery through which several Australian cases of melioidosis have already been connected (Inglis et al, 2000). In the other extreme, *B. pseudomallei* can also 'hide-out' in its host for a great length of time before some form of trigger causes melioidosis eruption, with records of US Vietnamese war veterans being diagnosed 26 years after serving, and the farthest case is of an 82-year-old American diagnosed 62 years after being a Japanese POW (Ngauy et al, 2005). In one unlucky case, a 61-year-old Englishman died of melioidosis over 2 weeks after visiting Thailand, not through infection whilst there but instead when gardening wearing his holiday shoes and a thorn punctured his *B. pseudomallei*-contaminated shoe's sole and inoculated him (Torrens et al, 1999).

This hardiness is not just limited to environment, but also is reflected in the application of antibiotics. The aminoglycoside gentamicin and the polymyxin colistin are usually very effective against Gram-negative species, by interrupting the bacterial 30S ribosome subunit (in the case of gentamicin) (Yoshizawa et al, 1998) or by permeating the bacterial cytoplasmic

membrane (as is the mode of action for colistin) (Falagas & Kasiakou, 2005). However, these antibiotics have no discernable effect on *B. pseudomallei* due to an efficient multi-drug efflux system known as AmrOB-OprA (Moore et al, 1999). Until recently, a 'cocktail' of 4 more conventional antibiotics was the most effective treatment for melioidosis (deoxycycline, chloramphenicol, cotrimoxazole and kanamycin) (Leelarasamee, 1998), but even with these the mortality rate of patients with septacemic melioidosis was 82-87% (Leelarasamee, 1998). In more recent times a newer, more effective treatment has emerged and using the cephalosporin ceftazidime the mortality rate has been halved to 35-40% (White, 2003). Ceftazidime functions by binding and inhibiting penicillin-binding proteins (PBPs) such as PBP-3 (Hayes & Orr, 1983), thus increasing susceptibility to lysis by incorrect peptidoglycan synthesis (Horii et al, 1998). However due to the highest rates of melioidosis occurring in relatively poor, in some cases '3rd world' countries, due to the endemicity of *B. pseudomallei*, the expensive treatments can often deny patients a fair chance of survival (Aldhous, 2005). A class of antibiotics thought to be emerging as more effective is the carbapenems that bind a different class of PBPs (PBP-1/-2) (Horii et al, 1998) but due to the high cost of \$100 per patient, per day the currently inconclusive trials are difficult to pursue further due to the large overhead (Aldhous, 2005).

Given its effectiveness at manifesting into such a life-threatening disease-state as melioidosis, *B. pseudomallei* is gaining increasingly more potential as a prospective bio-weapon, with most concern focused on the threat posed should such a weapon fall into the hands of bioterrorists (Aldhous, 2005). The Center for Disease Control and Prevention (CDC) categorises *B. pseudomallei* as a Category B biothreat, alongside Typhus fever and the Ricin toxin (DeShazer, 2004). This is especially of importance as its cousin *B. mallei* is one of the few bacteria to have ever been deployed in biological warfare, used by German agents to cause Glanders in Allied horses during World War I (Stone, 2007). Indeed it is regrettably not too hard to imagine the deadliness of an aerosolized weapon that when inhaled can kill within 42 hours and is resistant to the most readily available, large-scale produced anti-biotics. Even those who escape

immediate acute septicaemia may also relapse or form life-threatening abscesses later in life (Cheng & Currie, 2005).

1.2 Membrane-Anchored Proteins

1.2.1 Characteristics of Lipoproteins and Membrane Association

In order to maintain the smooth operation of such a complex biochemical environment as a cell, a wide variety of molecular machinery is required to fulfill all the required processes that ensure cellular survival. In the vast majority of cases these processes are mediated by the functional abilities endowed upon proteins in their own very exact niches. The amino-acid compositions of a protein and the resulting tertiary structures they form upon folding strictly define their role in a cell or organism.

Lipoproteins are a subset of the proteins that are associated with the inner or outer membrane of bacteria. Proteins with a long stretch of hydrophobic residues at their carboxyl or amino terminus can potentially insert into the hydrophobic interior of a cell membrane and act as membrane-anchored proteins (Eisenberg et al, 1984), assuming that the hydrophobic 'tail' or loop is not in fact an integral component of a protein's hydrophobic core (Munoz et al, 1995). This form of membrane anchor is not the sole type of membrane-anchored protein. Aside from the hydrophobic sequence-insertion, proteins can also become membrane-anchored via covalent bonding to a hydrophobic, membrane-affiliated compound such as glycosylphosphatidylinositol (GPI) (Fujita & Kinoshita, 2012), by electrostatic binding to phospholipid head-groups (Mukhopadhyay & Cho, 1996) and also by membrane-insertion of amphipathic helices into the plane of the phospholipid bilayer, known as IPM anchors (in-plane membrane anchors) (Sapay et al, 2006). In addition, a hydrophobic N-terminal sequence may act as a signal for export (Lee & Hannink, 2001) or for lipid anchoring (Zimmermann et al, 2010) (Figure 1.4).

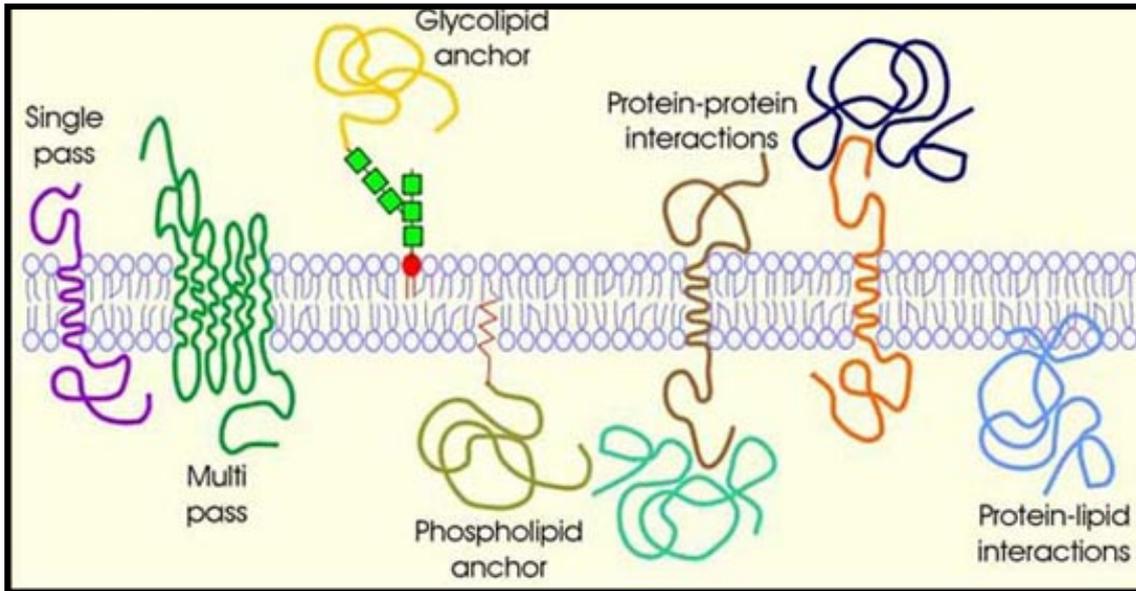


Figure 1.4: Diagram to illustrate the variety of methods through which proteins can interact with a cell membrane. The forms of protein membrane-associations that were potentially under investigation in this project were the GPI-glycolipid, phospholipid and IPM protein-lipid interactions. Figure adapted from online material made publically available by the University of Toronto.

1.2.2 Glycerylcysteine Lipid Anchoring

Another method of lipoprotein membrane anchoring is the focus of the main target in this body of work, namely membrane-association via a covalent glycerylcysteine moiety (Nielsen & Lampen, 1982). Following a highly hydrophobic amino-terminal signaling sequence a lipoprotein cleavage (/) motif of LLAG/CX can be found, where X acts as a determinant of membrane localization (Hayashi & Wu, 1990). In *E. coli*, an Asp residues at this +2 position localises a lipoprotein to the periplasmic face of the inner cell membrane, whereas a greater variety of residues, most commonly Ser+2, leads to localisation to the periplasmic face of the outer cell membrane or less commonly the outer face of the outer cell membrane (Yamaguchi et al, 1988) (Figure 1.5).

In this mode of membrane-association, upon translocation across the cytoplasmic membrane into the periplasmic space, a prolipoprotein +1 Cys residue is acted upon by a glyceryl transferase recognizing the LLAG/C motif, to create a glycerylcysteine moiety from a donating phosphatidylglycerol (Hayashi & Wu, 1990). This glyceryl-prolipoprotein is then recognised by a glyceryl-prolipoprotein transacylase to form a +1 Cys diglyceride via an ester-linkage between the glyceryls (Hayashi & Wu, 1990). Diglyceride-prolipoprotein is then cleaved at the LAG/C-diglyceride site by a prolipoprotein signal-peptidase to release the hydrophobic signal peptide (to be degraded to composite amino acids by peptidases) (von Heijne & Abrahmsen, 1989) from the lipoprotein, with the exposed N-terminal -NH group of the lipoprotein becoming further modified by acetylation by the action of a transacylase (Hayashi & Wu, 1990). This leaves the N-acyl-S-diacylglyceryl-cysteine lipoprotein associated to a membrane by covalently attached diglyceride and acyl-moieties (Figure 1.6a and b).

Once tethered to a membrane by this or any type of anchor, a membrane-anchored protein is now localised to fulfill its specialised role, such as acting as effectors (Liscum & Faust, 1987) or adaptors linking extra-cellular signals to internal down-stream signalling cascades (Kovacs-Simon et al) or aiding the delivery of bio-molecules via membrane trafficking (Gandhi et al, 2010).

Protein Construct	Construct Sequence
Lipoprotein-28	MKLTTHHLRTGAALLLAGILLAG ▼ CDQSSSDAKHIKVGVI----
Major lipoprotein	MKATKLVLGAVILGSTLLAG CSSNAKIDQLSS-----
pJG 311	MKATKLVLGAVILGSTLLAG CSSNAKIDQGEDQ- β -lactamase
pKY 201	MKLTTHHLRTGAALLLAGILLAG CDQSSSDAKHIKDLRIPG- β -lactamase
pKY 316	MNRTKLVLGAVILGSTLLAG CDQSSSDAKHIKDLRIPG- β -lactamase
pKY 313	MNRTKLVLGAVILGSTLLAG CDQSSSDAKHIKVGVIN--
pKY 623	MKLTTHHLRTGAALLLAGILLAG CSSNAKIDQLSS-----
pKY 702	MKATKLVLGAVILGSTLLAG C SNAKIDQGEDQ- β -lactamase
pKY 841	MKLTTHHLRTGAALLLAGILLAG C SQSSSDAKHIKDLRIPG- β -lactamase
pKY 831	MNRTKLVLGAVILGSTLLAG C SQSSSDAKHIKDLRIPG- β -lactamase

Figure 1.5: Investigation of the +2 lipoprotein-sorting rule in *E. coli*. Lipoprotein-28 (+2D) is located to the inner membrane, whilst the major lipoprotein (+2S) is an outer membrane protein. ▼ represents the site of LLAG/C cleavage. The pJG and pKY constructs represent variations in exchange of N-terminal signal sequence and replacements of +2D with +2S and vice-versa. β -lactamase constructs provided an identification tool for location of the protein when membranes were purified.

In all cases, when lipoprotein-28 +2 was changed to serine, it was detected in the outer membrane instead of the inner membrane, regardless of alterations to the N-terminal sequence. Conversely, the alteration to +2D in the major lipoprotein sent it to the inner membrane. This experiment illustrated that localisation of lipoproteins in *E. coli* is directed by the residue found in the +2 position after LLAG/C cleavage. Figure adapted from the findings of Yamaguchi (Yamaguchi et al, 1988).

1.2.4 General Functions Associated With Bacterial Lipoproteins

Bacterial lipoproteins play a fundamental role in many of the key processes that ensure the survival of all the members of the bacterial phylogenetic domain. They can perform extracellular roles such as mediating cell adhesion (Otto & Silhavy, 2002) and reproduction through conjugation (Seubert et al, 2003), to carrying out survival functions such as aiding in the process of transporting proteins, nutrients and toxins across cell membranes, either as a mode of uptake or release (Hayashi & Wu, 1990). Bacterial lipoproteins are commonly involved in sensing the environment and transmitting signals across membranes and through the cell to affect core regulatory processes such as gene transcription and protein synthesis, often acting as adaptors of the primary detection-and-signalling systems found associated with the outer and inner membranes of a bacterial cell (Leverrier et al, 2011). In spore-forming bacteria, lipoproteins can act as a trigger for switching between a spore-forming or germination event (Banse et al, 2011). All of these processes link to the manipulation of cell growth and division that exists as a fine balance to ensure the viability of each bacterium of all bacterial species.

1.2.5 Examples of Lipoproteins in *B. pseudomallei*

Several lipoproteins with putative or assigned function have been identified in *B. pseudomallei*, although due to the high variation in biological functions from a small number of lipoproteins discovered it was proposed that many more functions for many more different lipoproteins are yet to be discovered in the organism. In regards to glyceryl-cysteine lipid-anchored lipoproteins, the discovery of a putative Lipoprotein Signal Peptidase II (LSP II) responsible for the cleavage of a diacylglycerol-prolipoprotein prior to membrane insertion illustrates the cyclical nature of biology, as a lipoprotein that potentially acts upon a different lipoprotein of unrelated function (Holden et al, 2004). An example of such a *B. pseudomallei* protein that may be processed by lipoproteins such as LSP II is the LAGC-motif-containing OprM-homolog, homologous to a protein thought responsible for multi-drug resistance in *Pseudomonas aeruginosa* (Bianco et al, 1997). This resistance-nodulation-division (RND)

protein forms part of an outer membrane drug-efflux system, which expels toxins from the bacterial cell, such as antibiotics (Bianco et al, 1997).

Exhibiting a different biological role, OppA is a periplasmic lipoprotein anchored to the inner membrane and acts as an oligopeptide binding protein of the Opp system, which is responsible for bacterial nutrient uptake and signalling. It has also been shown to perform a role in virulence of species such as *Staphylococcus aureus* as well as *B. pseudomallei* (Harland et al, 2007). Amongst other examples, a protein such as the small, 45 residue, 4.3 kDa Entericidin B is a lipoprotein putatively assigned to be part of the *B. pseudomallei* response to externally sensed toxins from the growth environment, representing a cell-signalling role for lipoproteins within the organism (DeShazer et al, 2006). The lipoproteins briefly described in this section illustrate the essential and varied roles such membrane-associated proteins play in the growth, survival and virulence of the *B. pseudomallei* bacterium.

1.3 Aims of the Research

1.3.1 Bioinformatical Discovery of a Novel, Membrane-associated Target

It was proposed that by studying the Kyte-Doolittle hydrophathy plots (Kyte & Doolittle, 1982)(Figure 1.7) created for the entire catalogue of hypothetical proteins translated from all the genes of *B. pseudomallei*, it may be possible to identify membrane-associated proteins by the presence of potentially membrane-interacting hydrophobic termini. Once a short-list of targets was created, database analysis would be employed to identify a target to which no known function was assigned, and after the selection of several targets sequence analysis would be undertaken to identify the final selection that would be likely to be most suitable for crystallographic experiments.

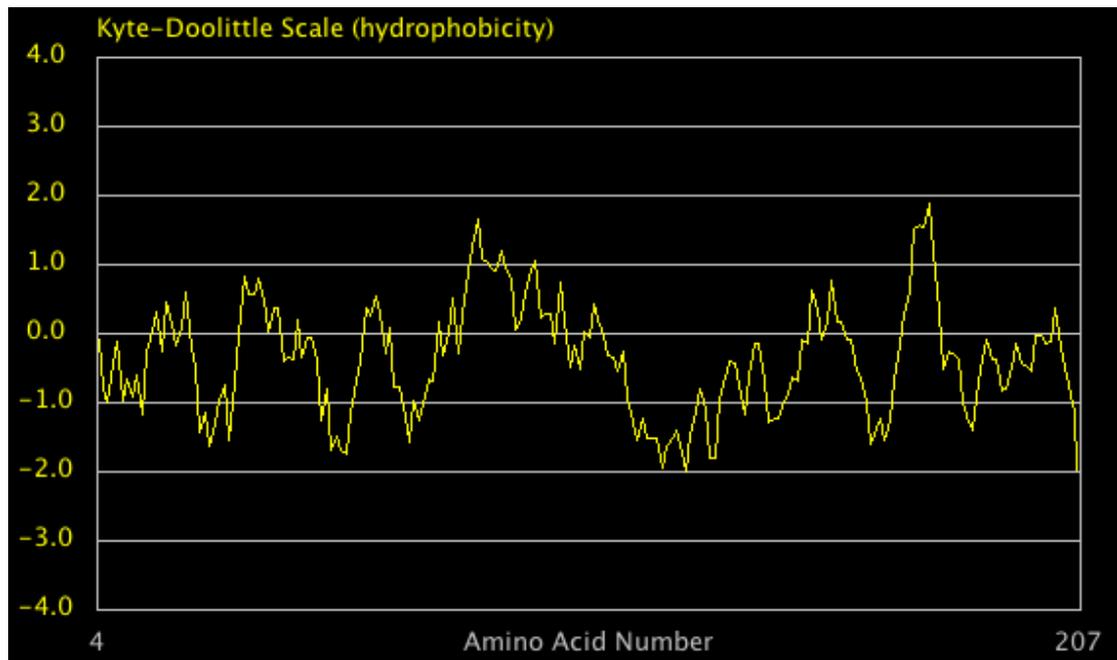
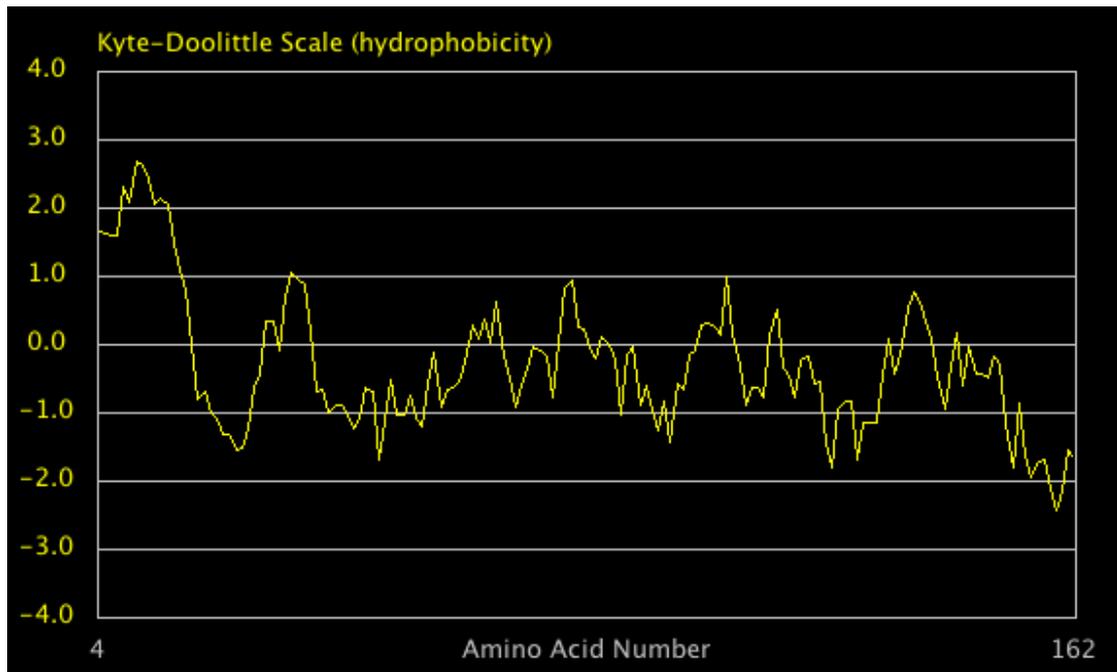


Figure 1.7: Hydropathy plots of *B. pseudomallei* proteins using the Kyte-Doolittle scale of hydrophobicity (Kyte & Doolittle, 1982), created using the Molecular Toolkit server from Colorado State University (www.vivo.colostate.edu/molkit/).

Top: The hydropathy plot of BPSL1204, a lipoprotein of unknown function. As can be seen, a highly hydrophobic (>1.0) N-terminal region is seen, which forms a signalling peptide for insertion though a cell membrane.

Bottom: The hydropathy plot of BPSL1549, a known secreted toxin (Cruz-Migoni et al, 2011). No N- or C-terminal hydrophobic sequences are seen, as this protein is secreted and does not associate with a membrane.

1.3.2 Prediction of Unknown Function From Structure

The most common application of structural biology is to take a protein of known function, and through structural analysis discover the biomechanics and chemistry of the molecules' biological functions at the atomic level. In recent years, as the Protein Database (PDB) (www.pdb.org) (Berman et al, 2000) has compiled an increasing amount of structural coordinates (Figure 1.8), the process of discovering structures of known function has often become experimentally simpler due to the international accessibility to a database of homologous structures that facilitate a more pipeline-based method of solving structures by Molecular Replacement (Blow and Rossmann, 1962) (McCoy et al, 2007). Therefore, the structural solution of a protein of totally unknown function with no known homologous structures provides both a more experimentally challenging aim to a project, but also holds within the potential to discover a novel biological process in the organism under investigation.

With an X-ray crystal structure solved, structural coordinates of the protein of unknown function can be compared to databases of known structural coordinates to search for structural similarity, even if sequence homology does not exist. Many proteins carry out similar processes to each other, but totally lack any sequence homology. For example, the FtsZ bacterial septum protein (Bramhill & Thompson, 1994) has no sequence homology to any eukaryotic tubulin proteins (Howard & Hyman, 2003), yet structurally they are very similar and play analogous roles in contraction and cell division (Carballido-Lopez & Errington, 2003). Therefore a protein that has no function assigned to its sequence may be investigated structurally to discover a possible function in a biological process using comparative bioinformatics techniques.

1.3.3 Support of Structural Findings with Experimental Data

Upon structural assessment and the assignment of a putative biological role to the protein under investigation, a series of experiments are required to give biological proof of the proposed function. These take the form of activity assays,

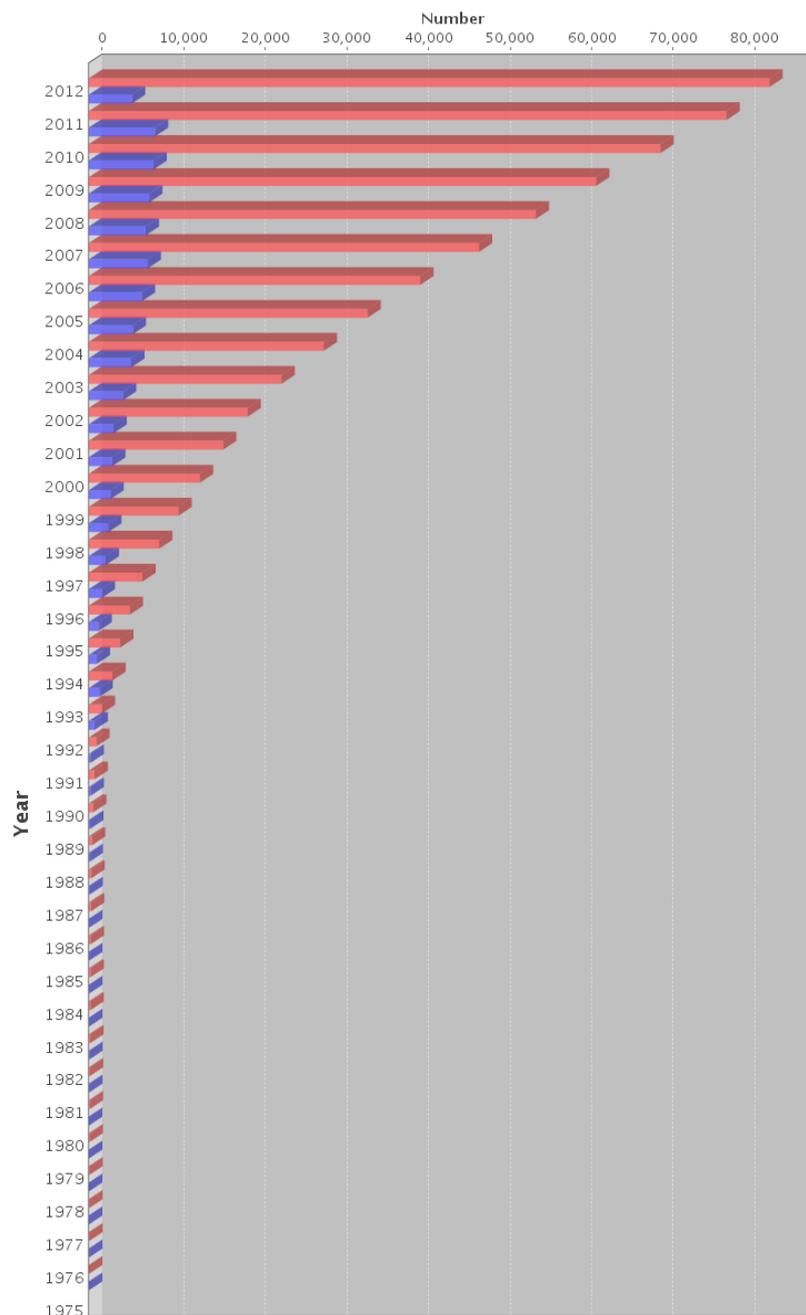


Figure 1.8: A bar chart plotting the growth of the PDB (**Total** and **Yearly** depositions) (www.pdb.org) (Berman et al, 2000). Since its conception in 1971 with a total of 7 structures, the PDB has slowly grown over time, and began to develop rapidly in the late 1990's with the dawn of modern structural biology techniques and the World Wide Web.

As of 1st September 2012, there were 84,381 structures deposited, of which 74131 were X-ray crystallographic. Of these, nearly 70% were solved by Molecular Replacement methods.

interaction assays, structural investigation of complexes and potentially microbiological investigation in the form of knockout and localisation studies. With results that corroborate with the structural assignment of function, it can be declared that the function of a previously undiscovered protein has been deduced through the application of structural biology. This process will also indicate the likely function of proteins with high sequence homology.

1.3.4 Potential Medical Applications

Very little is known regarding the biochemical background to melioidosis, and even less is known about the role of membrane-associated proteins in *B. pseudomallei*. Therefore the investigation of proteins of unknown function and the discovery of their cellular roles holds the potential to unlock new understanding about the pathogenic nature of *B. pseudomallei*, the disease state of melioidosis and hopefully give insight into the treatment of such a dangerous disease.

As discussed previously, membrane-associated proteins are often involved with processes such as defence against antibiotic attack, survival by nutrient uptake and toxin expulsion and are often instrumental in the processes that allow a bacterial cell to adhere to, culture inside of and attack a host organism (Hayashi & Wu, 1990). As such, especially in the latter example, membrane-associated proteins such as lipoproteins may also act as surface antigenic agents that can be exploited as a means of detecting an infection with increased rapidity. By understanding the biological defence, survival and attack roles that an unknown membrane-protein may play in *B. pseudomallei* infections, therapeutic applications could be developed to aid in the treatment of melioidosis and thus decrease the mortality rate of a life-threatening disease that affects so many people in regions of endemicity.

Chapter 2: X-ray Crystallographic Theory

2.1 Introduction

The science of determining the 3D structure of a biomacromolecule using the diffraction of X-rays from a crystalline protein sample has been evolved over a long period of time. Wilhelm Conrad Röntgen discovered X-rays in Würzburg in 1895 with the first crystallographic diffraction experiments carried out in Cambridge by William and Henry Bragg in 1912. Max Perutz and John Kendrew eventually adapted these to provide the first protein crystal structure of myoglobin in 1958 (Kendrew & Perutz, 1957). The introduction of multiple-wavelength synchrotron radiation for phasing experiments around 1994 has led to the modern age of protein structure determination, in which the variation and number of known protein structures has rapidly increased, reflected by the contents of the PDB (Berman et al, 2000) (Figure 1.8 in Section 1.3.2).

This section aims to briefly describe the basic practices and principles that allow for the determination of a novel protein structure using X-ray methods upon crystalline protein. For more in depth historical accounts and greater details on experimentally and theoretically derived evidence, the author directs you to the texts of Rhodes, G (2000), Blow, D (2002) and Rupp, B (2009).

2.2 Protein Crystallisation

2.2.1 Introduction

In order to immobilise protein in a form that can be manipulated, and to amplify the signal generated from the weak X-ray diffraction from a protein's electrons, the crystallisation of a protein is an essential first step in structure determination by X-ray methods. The ultimate aim of a crystallisation experiment is to increase the saturation of a protein in solution as to favour the organised removal of protein from solution into a solid state consisting of a regularly arranged crystal lattice, with the lattice held together by weak intermolecular forces. As such, protein crystals are very fragile and easily

damaged or destroyed. Crystallisation is done in the presence of various precipitants, buffers and additives that create an environment with the correct pH for appropriate surface-charge distributions, conformational homogeneity and availability of intermolecular contacts to allow a periodic array of macromolecules to form. The phase diagram in Figure 2.1 represents the overall aim of a protein crystallisation experiment (Rupp, 2009).

2.2.2 Vapour Diffusion

The most commonly employed method for crystallisation involves the process of water vapour diffusion from a protein-containing solution leading to an increase in protein concentration that eventually leads to super-saturation, resulting in crystallisation success or material precipitation (Benvenuti & Mangani, 2007). The following sub-sections describe the key steps and aims in variations of vapour-diffusion experiments:

2.2.2a: Hanging-drop Vapour Diffusion: a small drop (commonly 2 μ l) of protein solution is placed in the centre of a sterile, siliconised cover slip, into which is mixed an equal volume of the crystallisation solution. This leads to the two-fold dilution of the crystallisation solution in the drop. The cover slip is then inverted over a reservoir-well containing the original crystallisation solution in much greater volume (500-1000 μ l), with the well-lip greased to allow the sealing of a enclosed system by the inverted cover-slip.

Over time, water vapour diffuses from the drop into the reservoir in order to equilibrate the concentration of chemicals in the reservoir with those in the drop, and in doing so slowly raises the concentration of protein and precipitant chemicals in the drop (Mikol et al, 1990). At some point, protein becomes saturated and slowly begins to drop out of solution. If the conditions this occurs in are appropriate, then periodic protein nucleation begins, leading to crystal growth (Chayen, 1998). The set-up for a hanging-drop experiment is illustrated in Figure 2.2. This method has the advantage of often producing larger crystals,

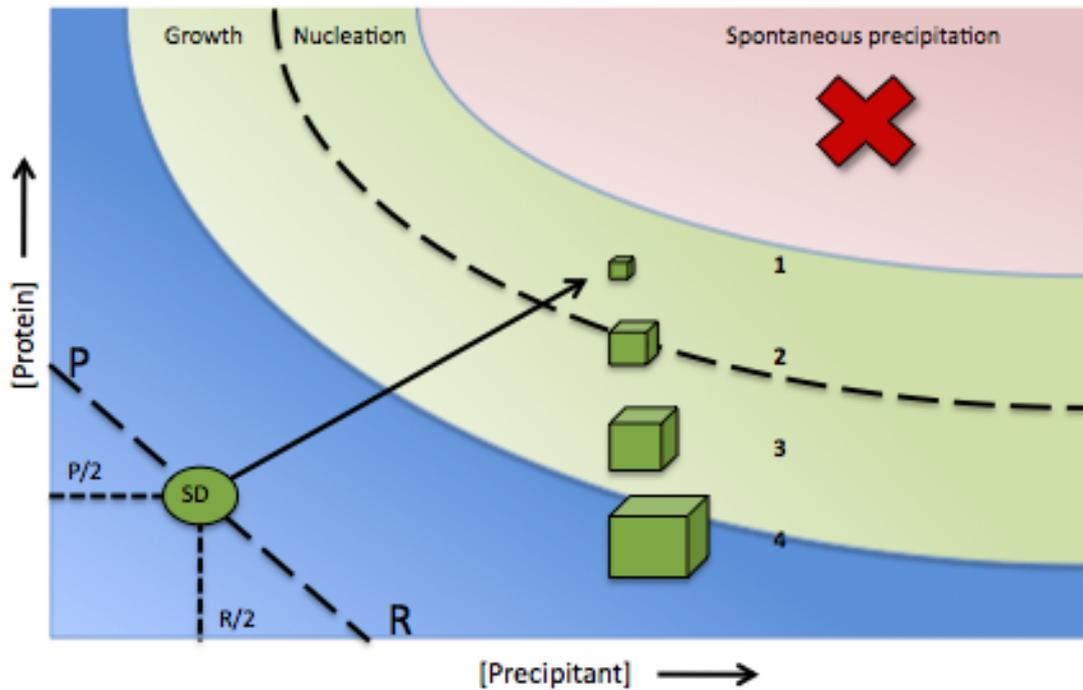


Figure 2.1: Crystallisation phase diagram. The starting drop (SD) consists of 1:1 protein solution (P): precipitant reservoir (R). As vapour diffusion into the reservoir occurs, the concentration of protein and precipitant in the drop increases, until:

- 1) Solution becomes saturated enough for spontaneous nucleation
- 2) Crystals may start to grow
- 3) While the crystals grow, the [protein] depletes in the solution, and a few initial crystals grow larger in the growth region
- 4) Once the larger crystals are in equilibrium with the saturated protein solution, growth slows and stops. Crystals have reached final size.

If the protein solution supersaturates too quickly, or the precipitant reservoir does not stabilise protein folding, spontaneous precipitation occurs. Figure adapted from Rupp, 2009

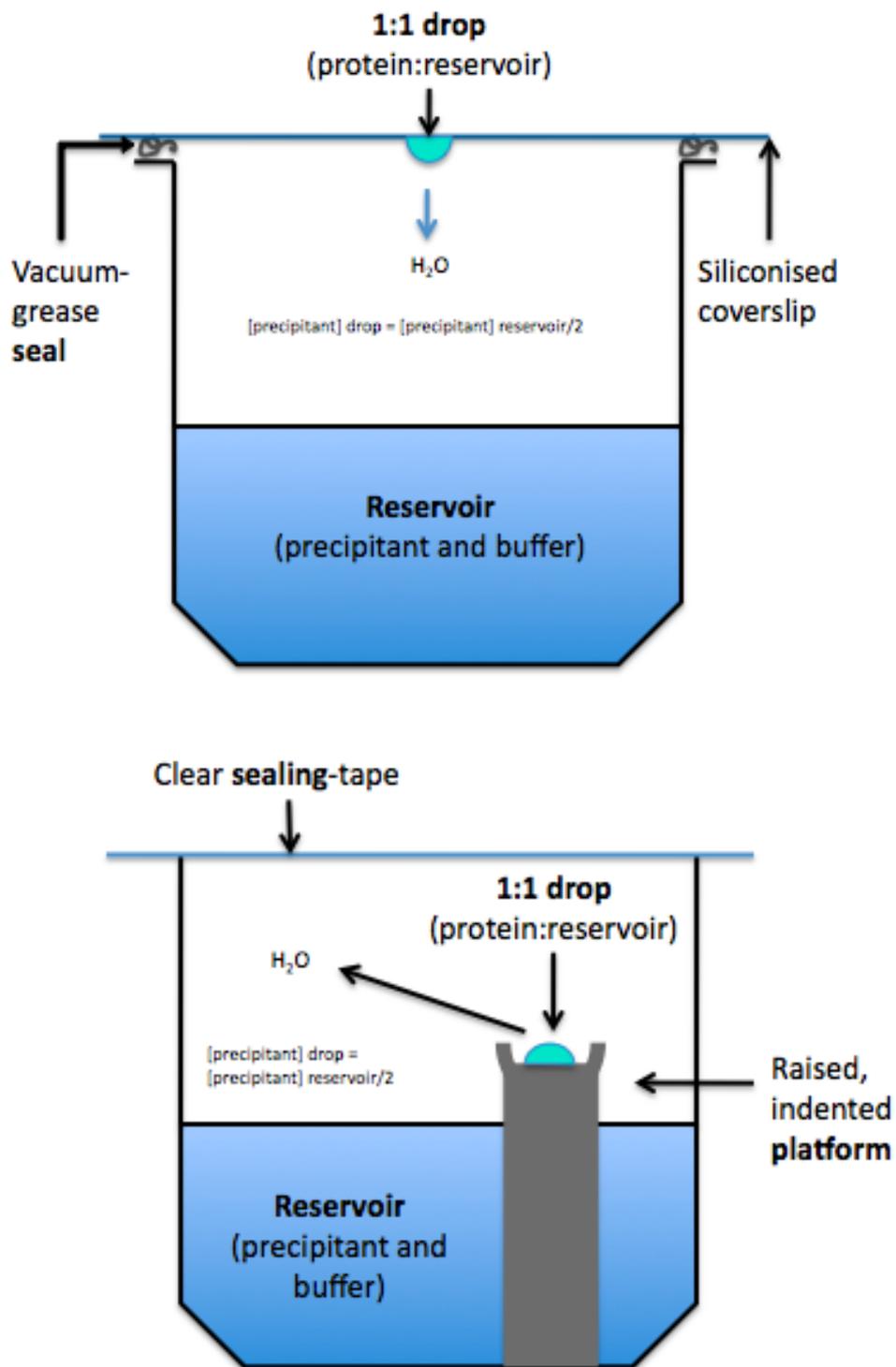


Figure 2.2: Illustrations of the set-up of crystallisation experiments. *Top:* Hanging-drop vapour diffusion. *Bottom:* Sitting-drop vapour diffusion.

with the volume of drops and reservoirs used easily varied. In addition, the manipulation of crystals from the flat-bottomed (when cover slips are re-inverted) larger drops is a relatively easier process than in other methods, as crystals rarely grow in contact with the cover slip and are accessible from multiple angles. A major disadvantage of this technique is that it cannot be automated, and thus the setting-up of experiments can be laborious, time-consuming and expensive.

2.2.2b: Sitting Drop Vapour Diffusion: The principles involved in this technique are similar to those of hanging-drop experiments. The main difference is the set-up of the experiment. The crystallisation drop is mixed on a raised, indented pedestal in the reservoir well, with the system often sealed with clear tape (Chayen, 1998). This set-up is illustrated in Figure 2.2. This allows for the creation of small-scale crystallisation plates, often with 96 wells, which can be readily automated for speed and consistency.

The smaller-scale (often 200-400 nl drops and 50-100 μ l reservoirs) and automated speed allows for more screens to be performed per volume of pure protein solution available. This can lead to more efficient investigation on the effects of temperature or protein concentration on crystal growth, with repeats of the same screen left to incubate in a variety of temperatures and protein concentrations. As such, modern crystallography laboratories regularly use automated sitting-drop plates as a first screening step in the crystallisation process. This technique does have disadvantages though, as frequently only small crystals grow in the smaller drops and they have the potential to settle on and stick to the bottom of the raised well. The fragile protein crystals are also much harder to manipulate for diffraction studies due to fewer angles of accessibility. In addition the dispensing of 96 or more conditions before the sealing of plates can lead to conditions that differ slightly from those intended, making re-growth more challenging, especially if solutions contain high concentrations of volatile chemicals are exposed to the environment for too long.

2.2.3 Batch Diffusion

In this method, high-concentration drops of protein solution and crystallisation condition are mixed and sealed under oil, or are applied to an oil covered micro-titer plate and allowed to sink through the oil to mix. This keeps the supersaturated drop sealed away from the environment. Water-permeable oil can be used to allow for the very slow diffusion of water vapour from the drop into the environment to further increase concentrations in the drop over time (Chayen, 1998). This method lends itself to automation due to sealing by the application of small volumes of oil, and has a main advantage in that conditions are sealed immediately and not exposed to the environment during set-up, reducing variation from intended conditions. However, this process is the most difficult to manipulate crystals from, with access to crystals blocked by the depth of the vessel used for the experiment and the need to loop a weak protein crystal through a layer of immiscible oil.

2.2.4 Dialysis

The well-established principle of dialysis has been adapted to the growth of protein crystals. Dialysis 'buttons' made of acrylic with a dialysis-membrane-sealable opening allow the free transfer of small molecules such as buffers and precipitants across the membrane, but retain the larger protein molecules and high molecular weight precipitants such as poly-ethylene glycols (PEGs). This facilitates the transfer of the crystallisation drop from one crystallisation solution to another, often carried out in a screen-designed sequence (Thomas et al, 1989). This can allow for the supersaturation of the solution, and thus protein crystallisation, by increasing the concentration and/or type of precipitant in the drop. This method is particularly useful for the growth of larger crystals, and allows for screening of many conditions using just one dialysis button. However, this process cannot be miniaturised and cannot be automated.

2.2.5 Optimisation of Crystallisation

After the establishing of a solution that generates at least poor quality crystals, a screen is set up to optimise this condition to produce the best quality crystals possible for diffraction studies. Often this is achieved by a screen set up to vary all applicable parameters (pH, [precipitant], [salt], [protein], temperature and drop volume and ratio) around the original hit to probe for the best condition possible. However, this may also involve the use of a different technique than the one originally employed, for example slowing down the rate of rapidly nucleating hanging-drop conditions by batch-diffusion methods or by replicating sitting-drop conditions as hanging drops to improve crystal size, number of crystal per drop or accessibility to particularly weak crystals. In addition to method variation, occasionally crystals can be encouraged to grow larger by seeding drops with nucleation-points derived from the fragmentation of smaller pre-formed crystals. This induces crystal growth by reducing the saturation required before crystal can grow, as a nucleation origin is already supplied.

If screening by the methods described above yields no crystals, which is possible as no protein can be guaranteed to crystallise, then modification of the protein sample may be required (Dale et al, 2003). Common methods included removal of purification tags, directed or random proteolysis, cloning of individual domains, co-crystallisation with a binding-partner or substrate and modifications such as glycosylation or deglycosylation (Dale et al, 2003). Modified protein samples are then re-screened and optimised. Once high quality crystals are achieved, then X-ray diffraction studies can begin.

2.3 Generation of X-rays

2.3.1 Rotating Anodes

X-rays are required to be used for diffraction experiments, as they exist in the electromagnetic spectrum as photons with wavelengths that closely match the lengths of chemical bonds and therefore allow atomic structures to be resolved. Moderately hard X-rays are used, due to their ability to penetrate matter. Most

in-house X-ray production is carried out using a copper anode to emit X-rays as $\text{CuK}\alpha$ -wavelength photons. This results in monochromatic X-rays of 1.54 Å wavelength at an energy of 8040 eV. X-rays are produced when high-energy electrons decelerate rapidly in an electron-dense target material; therefore a high-energy electron beam is required. This is produced by a metal filament cathode (commonly tungsten) opposite a copper anode in a vacuum-sealed tube. When high voltage is applied to the cathode high-energy electrons leave the cathode and travel to the anode. Upon impacting the anode, occasionally an electron is removed from the inner-shell of a copper atom, leaving a 'hole' that is filled by an outer-shell electron dropping down energy levels to complete the inner-shell and relax the high-energy state.

The energy released by the transfer of the electron down the electronic energy shells is released as heat and photons in the X-ray range. It is important to note that approximately 99% of energy released transfers into heat. In order to prevent the melting of the anode during high-energy X-ray generation (which historically occurred in sealed cathode ray tubes) the target anode is made to rotate at high speed (~10 krpm) through a water-cooled housing. This allows the dissipation of heat energy and spreads the impingement of the electron beam over the entire surface of the anode instead of on one fixed point, allowing for greater energies used in X-ray production and increasing the longevity of the anode material.

Emitted X-rays are then directed and collimated by allowing a small fraction through an entrance slit, which is made from a material that absorbs X-rays above the $\text{CuK}\alpha$ range, such as nickel, or by use of a graphite or germanium crystal as a monochromator. This is then collimated into a less divergent beam by a secondary slit, before the X-rays pass through the small holes of a collimator to leave a near parallel beam for focussing onto a crystal. This beam, though dramatically reduced in divergence, can still have a relatively large minimum area of exposure of 500 μm .

2.3.2 Synchrotron Radiation

In order to achieve energies that allow for high-resolution diffraction from a protein crystal to contain information bordering on a sub-atomic level, and to utilise the unique absorptive properties of heavy atoms in calculating phase solutions (see Section 2.6.2), tuneable, high-energy particle accelerators called synchrotrons are used in modern crystallography (Hendrickson, 2000). When a beam of electrons is forced by application of electrical and magnetic fields to move in a circular orbit, acceleration occurs due to centripetal force and radiation is emitted. When supplied with sufficient energy, electrons can maintain a stable circular orbit. Radiation is emitted tangentially from and restricted to the plane of the electron orbit.

Synchrotrons (so named as they keep electrons in a synchronous path) are initiated by a electron gun fired into a linear injector for pre-acceleration, from which high-energy electrons are fed into a main storage ring in bunches after passing through a smaller booster ring. When injected into the storage ring, in order to maintain a stable orbit, energy lost by radiation is replaced on each pass of the ring using radio-frequency generators. Electrons are then accelerated to near the speed of light by deflection into a ring by bending magnets, and more powerful insertion devices such as undulators and wigglers that 'wobble' the electrons through a high, alternating magnetic field. As the magnetic contribution to an electromagnetic wave contributes significantly less than the electronic component, the considerable size of these magnets required to generate a strong enough magnetic field is one of the reasons synchrotrons are so large (Rupp, 2009).

Beamlines are built from the main ring at a tangent to the forward direction of the electron orbit to collect the tangentially emitted radiation that includes photons in the X-ray range. X-rays are then selected from the radiation beam by an array of optics many metres long, such as collimating mirrors that fine-focus a large fraction of polychromatic light onto double-crystal monochromators. The relative positions of these devices allow for the selection of specific wavelengths

of radiated light with high-intensity and low-divergence, resulting in brilliant, focussed beams (down to micro-beams of $10\ \mu\text{m}^2$) of a tuneable energy.

2.3.3 Detection of X-rays

2.3.3a: Image Plate Detectors: This method of X-ray detection utilises the photoluminescence induced in BaFBr crystals, treated with lanthanide ions such as Eu^{2+} , when exposed to X-ray radiation (Miyahari et al, 1986). The non-diffracted beam is blocked by the diffractometer backstop, as the sample protein crystal does not diffract the majority of the X-ray beam. The BaFBr-lanthanide crystals in the detector phosphoresce at the points of exposure to diffracted beams, and are scanned by He-Ne lasers to induce energy release from the plate-crystals as photons, which are read by a photomultiplier and saved as a digital image. High intensity white light is used to erase the plate for use with the next image collection. The slow readout time generally makes these detectors only compatible with in-house sources (such as $\text{CuK}\alpha$ rotating anode sources).

2.3.3b: Charged-Coupled Device (CCD) detectors: CCD detectors deliver a digital image of diffraction via high-pixel CCD semi-conductor chips. These are connected to a thin fluorescent screen (commonly made of $\text{Gd}_2\text{O}_2\text{S}$ treated with Tb) via optical glass-fibre tapers (Pflugrath, 1999). Photons emitted from X-rays absorbed by the fluorescent screen pass via the optical glass-fibre, and upon contact with the semi-conductor chips in the CCD core generate free electrons proportional to the number of photons emitted from the screen to create a raw electronic image (Kraft et al, 2009). Due to limitations of how many chips and optical tapers can be associated with one screen, an array of detectors (3 x 3 for example) are usually combined to record each image to the highest resolution. Due to very high sensitivity and near-instantaneous image collection, with no resetting of the detector required, exposure time to high-intensity synchrotron radiation sources is kept low. Entire datasets can be collected in a matter of minutes using CCD detectors and synchrotron radiation (Hammersley et al, 1997), compared to the long exposures required (5 minutes per image and

several minutes to record each = 21 hours for a standard 180° dataset in space-group C2) using image-plates and CuK α radiation sources.

2.4 Basics of X-ray Diffraction

2.4.1 Braggs Law

One of the most important concepts in X-ray crystallography, which is crucial to the interpretation of diffraction experiments, is the satisfaction of the Bragg condition for coherent, constructive scattering (Ilari & Savino, 2008). For the derivation of this law, groups of atoms are considered as being arranged in sets of adjacent planes. In order for the scattering of X-rays from the electrons of an atom to be constructive (that is to add together) as opposed to destructive (cancel each other out) the distance between sets of adjacent atomic planes must keep the beam of incidence and the beam of reflection in phase. The Bragg equation is:

$$n\lambda = 2d_{hkl}\sin\theta$$

where θ is the scattering angle and d_{hkl} is the interplanar distance between sets of reflecting planes hkl (in Miller indices) and n is an integer (Perkins, 1988). In order for constructive diffraction to occur, n is required to be an integer.

If the diffracted beams from two planes emerge out of phase, then they cancel each other out. However if they emerge in-phase, then they become additive and generate a stronger, cumulative reflection. Figure 2.3 illustrates the conditions required in order to satisfy the Bragg condition for constructive diffraction.

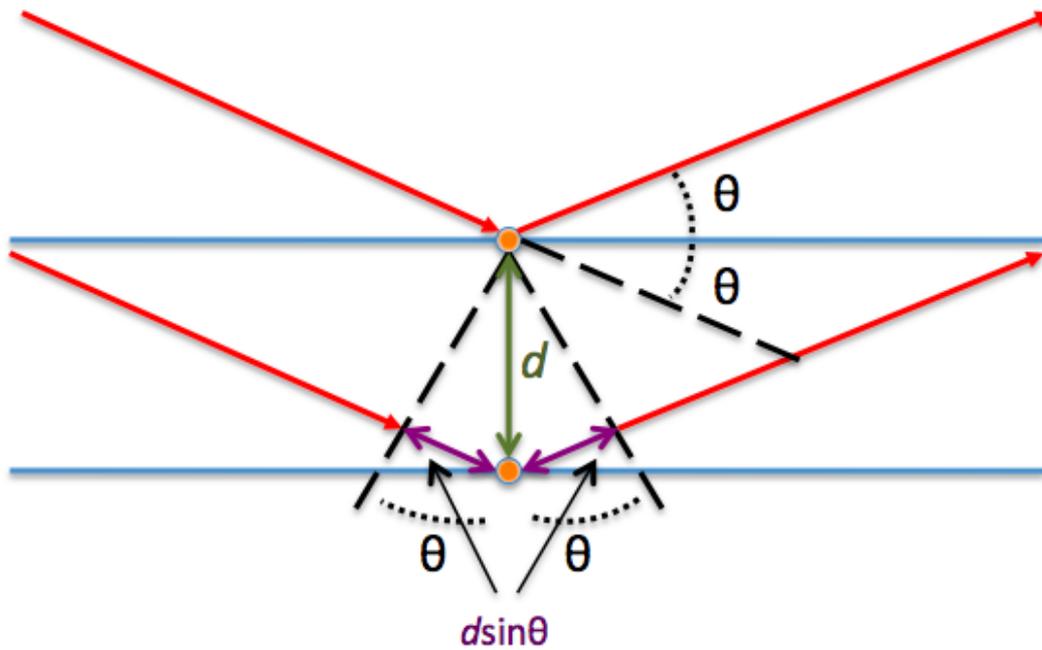


Figure 2.3: Illustration to explain the Bragg Equation. The difference in path of the two partial waves between atoms in parallel planes is equal to $2d\sin\theta$. In order to achieve constructive interference, in which both waves emerge in phase, the path difference must equal an integer multiple of $n\lambda$.

The application of Bragg's law, whereby X-ray diffraction is treated as reflection from discrete planes of atoms, allows for the interpretation of diffraction patterns, as the angle of diffraction becomes linked to the spacing within the lattice. This can allow the rearrangement of the commonly quoted Bragg equation to give:

$$n\lambda = 2d_{hkl}\sin\theta \quad \text{becoming} \quad 1/d_{hkl} = 2\sin\theta/n\lambda$$

where $1/d_{hkl}$ is the distance in the reciprocal lattice, or in other words the distance between reflections measured in the crystal diffraction pattern.

A change in oscillation angle during data collection is employed to ensure that all planes meet the Bragg's condition. The positioning of a Bragg-plane within a crystal into a geometric construction known as the Ewald Sphere allows for the prediction of 2-D reciprocal lattice points that can be potentially measured from 3-D diffraction over planes of distance $1/d_{hkl}$ (Figure 2.4) (Ewald, 1969). The sphere (shown as a 2-D circle in Figure 2.4) with $1/\lambda$ radius is combined with a representation of all the reciprocal lattice diffraction possibilities of the given crystal lattice placed over the point of origin on the Ewald Sphere. The reflections that carry the potential to be detected upon constructive diffraction of X-rays from the Bragg-planes in the crystal lie on the sphere's circumference. When rotation of the crystal occurs, the correlating rotation of the reciprocal lattice on the sphere's circumference allows the calculation of positions of reciprocal lattice reflections that fulfil the Bragg-diffraction condition upon the change of diffraction angle (Leslie, 2006).

2.4.2 Structure Factors and B-Factors.

Structure factors are expressions of the amplitude and phase-angle of a wave after diffracting from a plane in a crystal lattice described by hkl (or **h**) Miller indices. Structure factors (F_{hkl}) are summations (Σ) of the scattering of X-rays of every atom on a unit cell, and are calculated by:

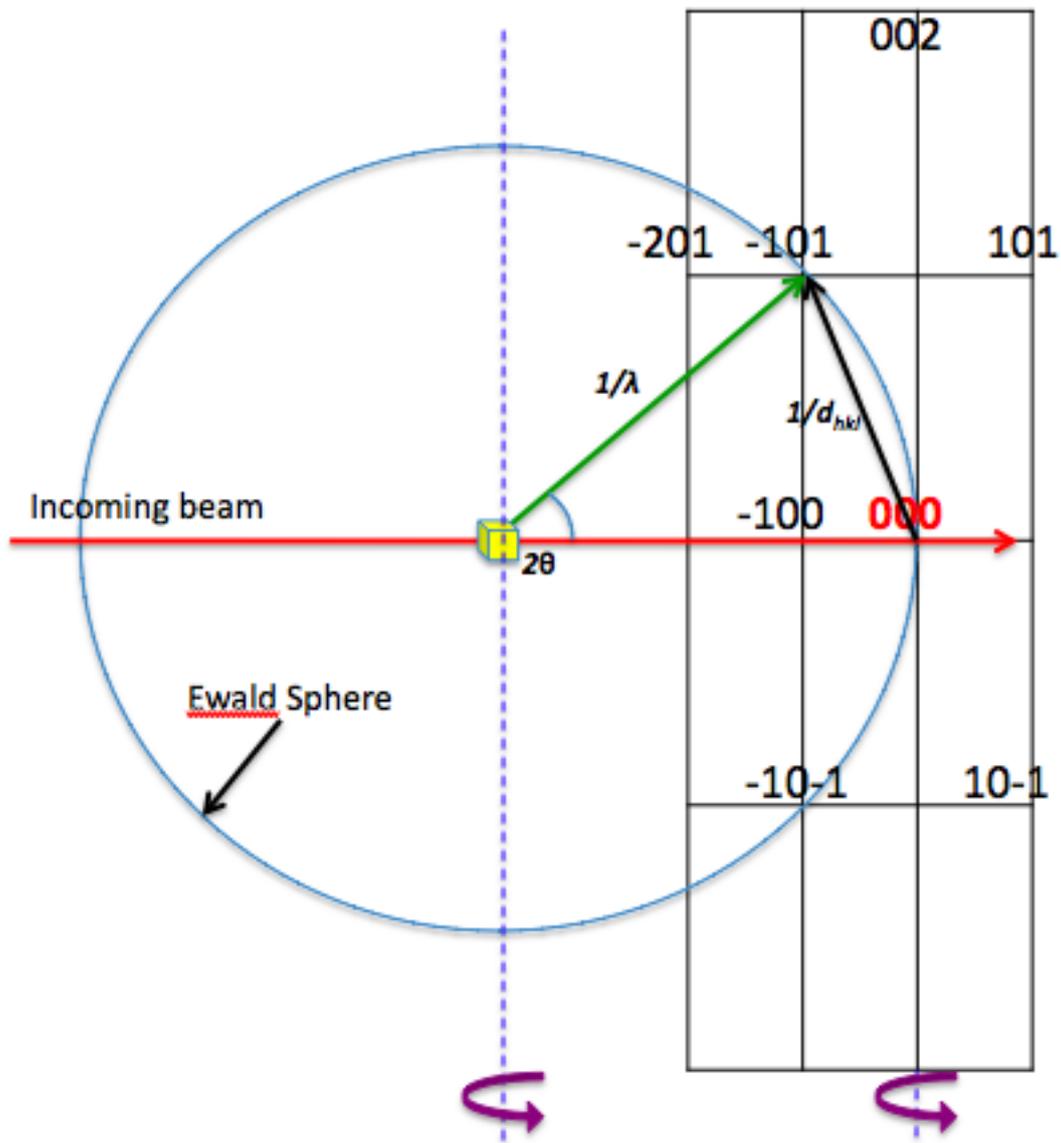


Figure 2.4: The Ewald Construction for a 2-D reciprocal lattice with Miller indices $h0l$. Reciprocal lattice points that fall upon the Ewald Sphere fulfil the Bragg-condition for diffraction, as the distance from the 000 origin (reflections -101 and $-10-1$ above) equals $1/d_{hkl} = 2\sin\theta/\lambda$.

Rotation of the crystal above rotates the reciprocal lattice by the same increment, bringing other lattice points such a -201 into the Ewald Sphere, creating a newly positioned diffraction spot that fulfils the Bragg condition from a different crystal plane.

The construction above demonstrates the relationship between reciprocal lattice reflections and real lattice diffraction in 2-D for ease of illustration. In actuality these processes occur in 3-D, on the surface of a sphere and with a reciprocal lattice of variable hkl .

$$\mathbf{F}_{\mathbf{hkl}} = \sum_{\text{atom}} f_s^0 \cdot \exp[-\mathbf{B}_j(\sin\theta/\lambda)^2] \exp(2\pi i \mathbf{h} \cdot \mathbf{x}_j)$$

where \mathbf{h} is the direction of atomic scattering (f_s), whilst \mathbf{x}_j is the position of the j th atom measured in fractional coordinates.

\mathbf{B}_j represents an isotropic displacement parameter, or in other words a **B-factor** to account for the motion of atoms within a molecule where $B_{\text{iso}} = 8\pi^2 \langle u^2_{\text{iso}} \rangle$ (where $\langle u^2_{\text{iso}} \rangle$ is the mean square isotropic displacement of an atom from its mean position) (Einstein, 1926). Also known as a Debye-Waller factor (Debye, 1913; Waller, 1923), B-factors describe displacement of an atom from a mean position, as a product of thermal or vibration movement, and disorder within a lattice. These effects attenuate atomic scattering, and as such have a negative effect on diffraction intensity and thus amplitude. The B-factor of each atom scattering is related to the scattering vector calculated using the diffraction angle, represented by the $(\sin\theta/\lambda)^2$ term. As the summed amplitudes of diffraction are subject to atomic scattering positions in a defined direction, attenuated by B-factors (Rupp, 2009), these terms can be condensed and structure factors can be re-written as:

$$\mathbf{F}_{\mathbf{h}} = |\mathbf{F}_{\mathbf{h}}| \exp [2\pi i \varphi(\mathbf{h})]$$

Where $\varphi(\mathbf{h})$ represents the interpretation of $\mathbf{h} \cdot \mathbf{x}_j$ as a summed phase-angle applied to summed amplitudes $|\mathbf{F}_{\mathbf{hkl}}|$. In a Bragg's diffraction such as in Figure 2.3, the addition of an atom in a new, parallel plane that still fulfils the same Bragg condition creates a phase difference between the scattering from the atoms in the original and new planes. Upon the addition of all atoms in all planes that satisfy the Bragg condition in a given crystalline molecule, a net phase-difference is created. These summations can be shown in an Argand diagram (Figure 2.5) in which cumulative plotting of atomic scattering factors (proportional to amplitudes) of all atoms with individual phase differences leads to the final structure factor with a net-phase angle (Taylor, 2010).

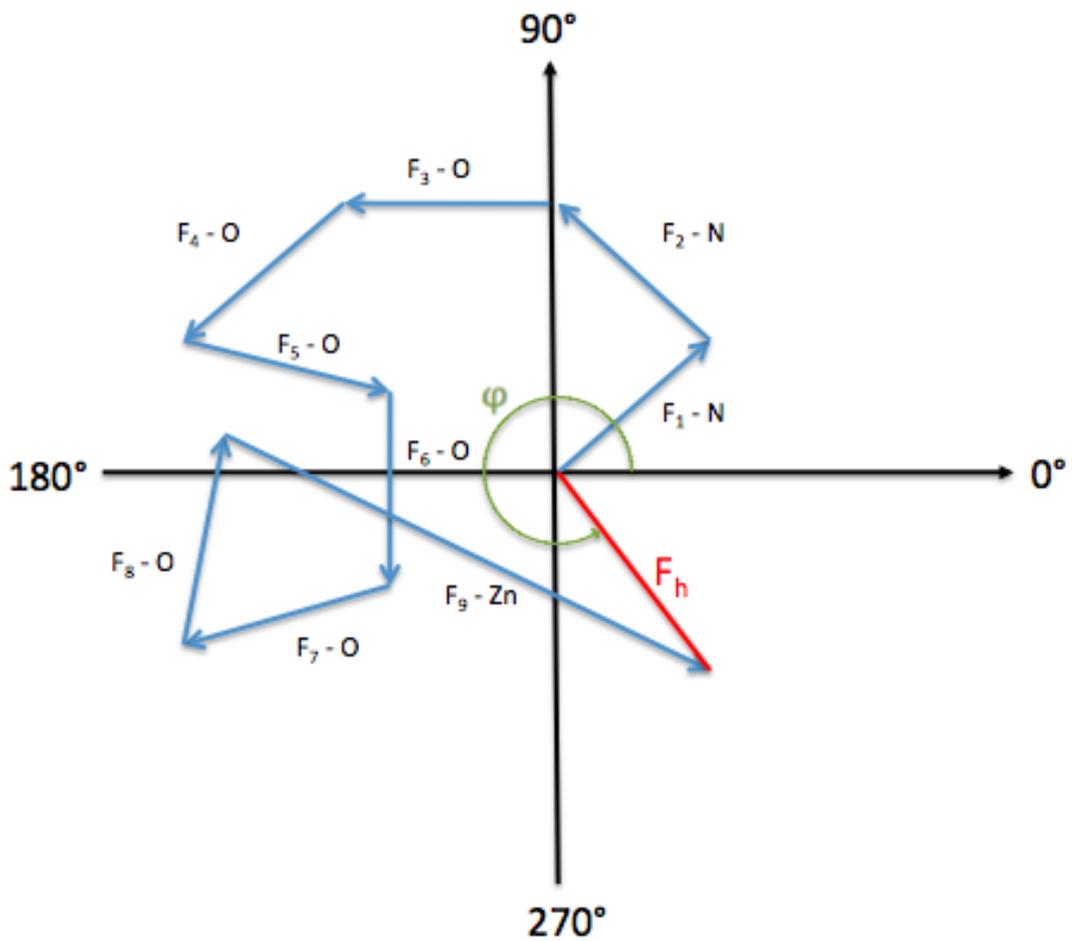


Figure 2.5: Argand diagram to illustrate the calculation of structure factors as a summation of individual scattering vectors. Plotted are the vectors of $\text{Zn}(\text{NO}_3)_2$, containing 8 light atoms and one heavy atom, for one reflection.

The lengths determined by scattering power (proportional to the number of scattering electrons) and related by differences in phase shows the creation of a structure factor (F_h) with overall phase angle (φ).

However, in a diffraction experiment only the intensities (the square of structure factor amplitudes) of scattering at $1/d_{hkl}$ are measured and the phase information is lost. Therefore experiments (Section 2.6) are required to calculate a protein crystal's phases. Once phases have been determined, electron density can be calculated and a model built into it, and structure factors and electron density can be interconverted without loss of information using Fourier transforms, therefore allowing the assessment of the refined model built into the density (Section 2.8).

2.5 Data Processing

2.5.1 Indexing

Indexing is a process in which the diffraction spots (as representatives of the reciprocal lattice) are assigned a consistent set of three reciprocal lattice vectors designated **h**, **k** and **l** (a^* , b^* and c^*). These relate to **a**, **b** and **c** real lattice vectors (separated by angles α , β and γ) that define the unit cell parameters. Factors such as minimum and maximum spot size and background pixel density are defined to discern individual reflections. Two or more diffraction images are taken in order to account for any anisotropy present and to retrieve enough strong reflections from more than one angle to assign a unit cell without ambiguity. Transformed lattice vectors are then created, to which maximum symmetry criteria can be applied for the unit cell calculated.

There are 7 crystal systems, from which a total of 230 space groups originate. Of these, only 65 allow for the chirality seen in proteins (Rupp, 2009) (Wukovitz & Yeates, 1995) and therefore impose limitations on the symmetry operations possible in biomolecular crystallography. Indexing algorithms in programmes such as iMOSFLM (Battye et al, 2011) then present the possible space-groups and unit cells from the measured reciprocal vectors with a penalty function. The lowest penalty (excluding triclinic P1, which can always be fitted with no penalties) represents the most likely space-group unit-cell parameters for the reflections supplied. This is not always totally accurate, as internal symmetry within the unit cell may override cell symmetry, but if this appears to be the case

a lower symmetry lattice for strategy determination is selected. This allows for the collection of redundant data that can be incorporated later, as opposed to collecting data for an incorrect, complicated symmetry group that leads to uncollected data post-strategy determination.

In general, data-collection strategies focus on the collection of datasets with the minimum exposure to radiation that allow for the optimal recording of the maximum number of reflections from the minimum number of angles for the cell-symmetry (Laue symmetry) calculated.

2.5.2 Integration

Once images are collected and indexed, multiple images are used to refine cell parameters and measure the mosaicity of the crystal. Indexed, raw pixel intensities of reflection spots then require correction against several factors, such as detector-specific spatial calibrations, beam geometry, crystal orientation and beam strength over time are corrected for, alongside the potential monochromatic polarisation of light by the crystal and the time each Bragg plane has spent in a diffracting position (the Lorentz-polarisation correction) (Phillips et al, 1976). Once corrected, the raw pixel intensities over all images are integrated, with any partial reflections caught in a Bragg reflection plane spread over several images combined into single reflections. Spots are identified by 2-D box-profiles of pixel intensity, with locations predicted based upon unit-cell and symmetry projections. Pixel counts outside of the photon-counting profile provide background. This allows the calculation and listing of intensity and error values for each reflection (Leslie, 2006).

2.5.3 Scaling and Data Reduction

The unmerged, integrated reflections are then scaled against each other to reduce the data to a list of a single, merged intensity values for each unique reflection measured over several images by a process of intensity averaging. A high redundancy, or multiplicity, of the collected data improves the reliability of this averaging. On the occasion that scaling weak intensities gives negative

values, skewing factors are applied to the averaging of weak intensities to generate positive values (Evans, 2006). Data is further improved by reduction by applying of resolution limits to measured reflections, performed as re-scaling with imposed limits on the resolution of data used.

High-resolution data is weaker than the more reliably and regularly measured low-resolution reflections, therefore in order to achieve the best signal:noise ratio ($I/\sigma I$) with the highest resolution, scaled data is analysed for several factors. Completeness of the dataset is assessed, as randomly missing data will create noise in electron density maps, whilst regularly missing reflections such as those that do not enter the Ewald Sphere discussed in Section 2.4.3 may lead to total phasing failure. Redundancy of data is calculated and represented in the terms R_{merge} and R_{pim} . R_{merge} represents how well unique reflections merge together over multiple images in the dataset, and as such normally is higher at high-resolution as fewer well-measured reflections may vary (due to physical equipment/crystal slippage or radiation damage of the unit cell) and are thus poorly merged over the dataset. R_{pim} is a more precise indication of merging, by incorporation of variance of distribution into the R_{merge} statistic, therefore representing well-related intensities as it decreases with increasing redundancy (Evans, 2006).

Once an acceptable resolution limit is found that gives high-redundancy, reliable data with a good signal:noise ratio, data are re-scaled and can also be used to recalculate cell symmetry for confirmation that processed data has been prepared with the correct unit cell parameters applied. Scaled and merged intensities (accompanied by a list of Miller indices) can then be used to determine their associated amplitudes by calculation of the square root of each intensity. This gives a final scaled and reduced dataset listing Miller indices with associated amplitudes and their associated error estimations (Evans, 2006).

2.6 Determination of Phases

2.6.1 The Patterson Function

A key utility in crystallographic structure solution is the Patterson function. In essence it is a Fourier transform of the intensities (or the square of structure factor amplitudes) of measured reflections, with phases set to zero (Grosse-Kunstleve & Adams, 2003). This is important, as it does not require phase information (yet to be discovered), and can be used despite the 'phase problem' (see Section 2.6.6). The Patterson function allows the calculation of a vector map by effectively placing every atom at the origin in turn, and plotting the protein's transformed intensities as interatomic vectors. When this is repeated for each atom of the protein, with all positions of the peaks recorded, it theoretically allows for the location of all the atoms in the substructure, with the Fourier transform of intensities proportional to the electron density of the scattering atoms (Grosse-Kunstleve & Adams, 2003).

However, due to the very large number of light atoms with weak scattering power that make up proteins it is impossible to solve a protein structure using this method alone. If the structure has N atoms, then a plotted Patterson map has $N \times N$ peaks, of which N are superimposed upon the origin, therefore giving $N^2 - N$ of non-origin Patterson peaks. This creates a noisy Patterson map filled with many indistinguishable peaks. Using direct methods (Section 2.6.2), small molecules have the potential to be solved this way (Sheldrick, 2008). In addition, the rotation and translation of Patterson maps of target protein and phased-density homologues used in Molecular Replacement (not discussed in this thesis) allow for the correct alignment of target and model data for phase-imposition (DeLano & Brunger, 1995). However, Patterson maps can be used in experimental-phasing in protein crystallography to locate and position a small number of heavy atoms used in solving the phase problem (Section 2.6.5).

2.6.2 SHELX and Direct Methods

Direct methods utilise the implicit relations between structure factor amplitudes to solve the phases of a structure, or of a heavy atom substructure in proteins. This method originates from the structure solution methods used for small molecules but due to the complexity and size of proteins, often with relatively low resolution data, direct methods are usually only applicable to solving the phases of heavy atom substructures as a first step to solving protein phases (Sheldrick et al, 1993). Successful phase solution requires the fulfilment of Sheldrick's Rule, in which data giving atomic separation $>1.2 \text{ \AA}$ are required (Morris & Bricogne, 2003). Due to stereochemical and repulsion limitations, anomalously scattering atoms in a protein heavy atom substructure are rarely $<3.5 \text{ \AA}$ apart.

Heavy atom data, with anomalous intensities and thus amplitudes identified and related to scattering factors (Section 2.4.2), are prepared by the solution of phasing equations described in Section 2.6.6. Computation of information about the phase-shift between heavy atom phase and protein phase can now occur, but this is not yet representative of the protein's phase (Rupp, 2009). Reflections representing heavy atoms in the lattice planes can now have starting-phase relationships derived. Multiple trial phase sets are produced for the strongest structure factors, however remaining weaker phases are also required. A process of tangent refinement follows, in which squared density peaks are recycled through the Hauptman-Karle equation to determine potential phase values (Sheldrick et al, 1993).

The macromolecular phasing program SHELXD employs dual-space direct methods (Schneider & Sheldrick, 2002). Observed heavy atom structure factor vectors are selected from a Patterson analysis, and used to identify the strongest translational superimpositions by assigning a score, with the best solutions extended to include the expected number of heavy atoms by further superimpositions. In the case of Seleno-met MAD phasing, the number of atoms is determined by the number of methionines per copy of protein in each unit cell.

The best correlation score of each set of heavy atoms between the observed values and those calculated by superimposition are kept to generate a trial structure. This 'seeding' by Patterson allows for the rapid, automated computation of in-phase peaks that are then cycled through the Hauptman-Karle tangent formula to refine test-phases against structure factors. The more sharp peaks generated in the trial electron density maps, the greater the success of the refinement (Schneider & Sheldrick, 2002).

SHELXD also calculates the occupancy of each heavy atom in the structure, allowing for identification of partial sites (Schneider & Sheldrick, 2002). This allows for analysis of substructure solution, as a sharp drop in occupancy indicates a correct solution.

2.6.3 MAD and Anomalous Scattering

Multi-wavelength Anomalous Dispersion (MAD) is a technique that utilises unique X-ray scattering properties of electron-dense atoms to provide information for calculating phases and thus the solution of *de novo* protein structures (Ealick, 2000). All materials absorb X-rays, with the heavier the atom the greater the amount of absorbance. In experimental terms, this can be a major source of radiation damage when heavy atoms are incorporated (even in low abundance) into a protein crystal (Murray et al, 2005). However, X-rays absorbed by heavy atoms causes useful properties in the atom.

As electrons in the electric field of the atom oscillate, certain electromagnetic frequencies induce energy transfer in the electron levels in the core shells, not just the valence shells seen in optical transitions, leading to energy release by emission of X-ray radiation equal to the energy of the original. The higher the X-ray energy used, the greater the energy of the electron that can be dislodged. The terms K-, L- and M-edge describe the energies required for heavy-atom emission from electrons shells of decreasing energy, with the K-edge being the highest energy needed to resonate a high-energy inner-shell electron (Hendrickson,

1991). Figure 2.6 describes the X-ray absorption spectrum of the heavy atoms Selenium and Gold compared to that of the light atom Oxygen.

The absorption and emission of photons from heavy atoms induces anomalous dispersion components into atomic scattering factors, represented as:

$$f_s = f_s^0 + f' + i \cdot f''$$

The real and imaginary anomalous scattering contributors, f' and $i \cdot f''$ respectively, are scattering angle-independent and are additive to the normal scattering-angle-dependant atomic scattering f_s^0 , seen in the non-anomalous Structure Factor calculation in Section 2.4.2. The Kramers-Kronig equation allows precise calculation of f' from the measured $i \cdot f''$ measured at the absorption edge peak, where $i \cdot f''$ is proportional to an atomic absorption coefficient at the wavelength used. The negative f' peak coincides with the steepest slope of the $i \cdot f''$ inflection (Hodeau et al, 2001). In practice, data collected nearest to the peak wavelength gives the best estimation of the imaginary component $i \cdot f''$, with data collected at the wavelength of the inflection point approximating the real f' scattering component.

At an absorption edge, dispersive differences arise that result in greater intensity of measured intensity of the same reflection at the peak absorption wavelength than a less absorptive wavelength, giving a measurable change in scattering due to absorption. This property is taken advantage of in MAD, as the same reflection can be measured from the same crystal at different X-ray energies (peak/imaginary $i \cdot f''$ component, inflection/ real f' component and a non-anomalous remote energy) to create a comparable difference in diffraction intensity (and thus amplitude by square root). Compared to non-absorptive scattering of light atoms, absorption of X-ray energy by heavy atoms also creates a retardant effect that affects f'' by inducing a phase change (90°). This anomalous difference is utilised by techniques such as MAD to allow for the calculation of phases by the breaking of internal centrosymmetry within diffraction patterns, described in the next sections (Hendrickson, 1991).

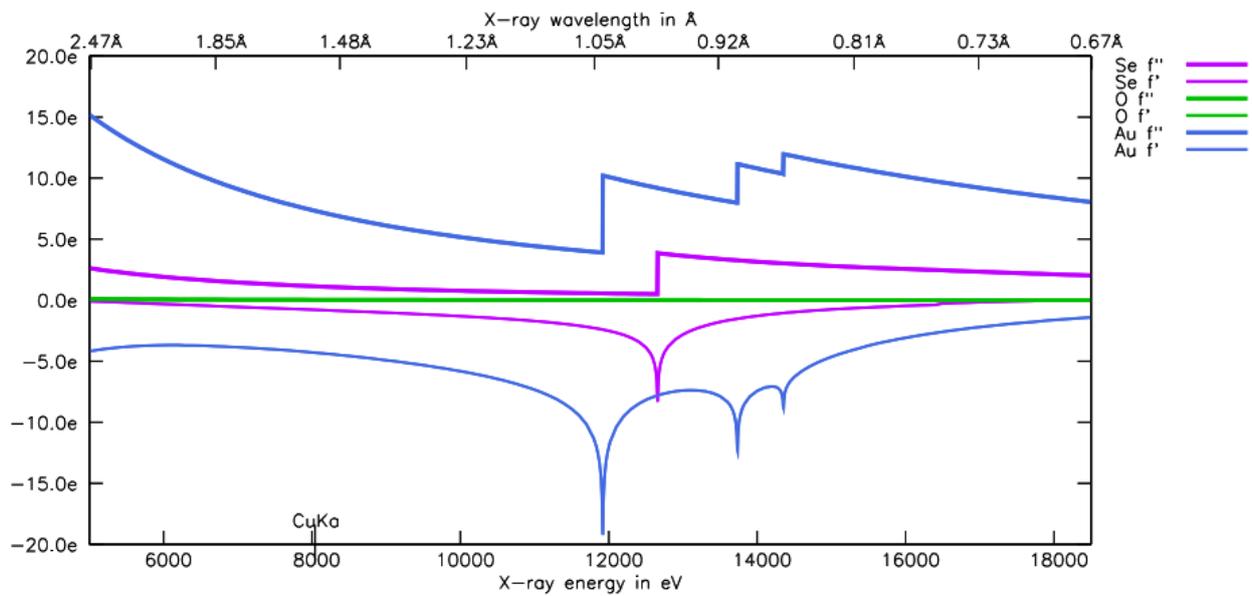


Figure 2.6: X-ray absorption plot (f' and f'') for the light element oxygen, the ‘medium’ electron-density selenium and the very electron-dense gold in an energy range achievable by synchrotron light sources (5-18.5 KeV). A key is shown top right.

The standard in-house copper-anode energy ($\text{CuK}\alpha$) is marked, to illustrate the lack of anomalous signal peaks achievable and thus to highlight the importance of synchrotron radiation use.

From left to right, the L-III, L-II and L-I edges are achievable for Au, with the greatest anomalous difference seen at L-III. Only the K-I edge is available for use in Se experiments. Oxygen (the heaviest of the most common, light protein atoms, H, C, N and O) gives no anomalous signal.

This plot was produced using Ethan Merritt’s online anomalous scattering tutorial (<http://skuld.bmsc.washington.edu/scatter/>).

2.6.4 Friedel's Law

Centric reflections are phase-reflected reflections caused by diffraction from centrosymmetrically related atoms, defined by the symmetry of the crystal structure itself. They are found in centric positions in the diffraction pattern (reciprocal space) and structure factors for these reflections have exactly the same magnitude with conjugate phase-angles. These are known as Friedel Pairs, and follow Friedel's Law:

$$F_h = F_{-h} \quad \text{with} \quad \varphi_h = -(\varphi_{-h})$$

As structure factors are summations of amplitudes, and amplitudes are proportional to the square root of measured diffraction intensities, this leads to a connection between phases and intensity. Anomalous scattering (measured as intensity changes) allows for the breakdown of Friedel's Law, and as such provides a means to calculate the phase angle associated with an anomalously dispersive heavy atom (Zwart, 2005).

Figure 2.7 shows an Argand diagram visualisation of how anomalous dispersion breaks Friedel's Law. When the sum of normally scattering atoms (F_P) is plotted against an anomalously dispersive atom (F_A) in an inappropriate energy for anomalous absorption, structure factor amplitudes (F_{PA}) remain mirrored about the real axis as Friedel's law remains true. However, upon anomalous diffraction conditions, the real (f') component of anomalous scattering (with the same phase as the F_P as the scattering atom has not changed position in the two scenarios) also mirrors about the real axis, with Friedel's Law holding. However, the f'' -phase-lagging-component induces a $+90^\circ$ (or -270°) phase shift, resulting in $|F_{PA}| \neq |F_{-PA}|$ and thus breaking Friedel's Law (Taylor, 2003).

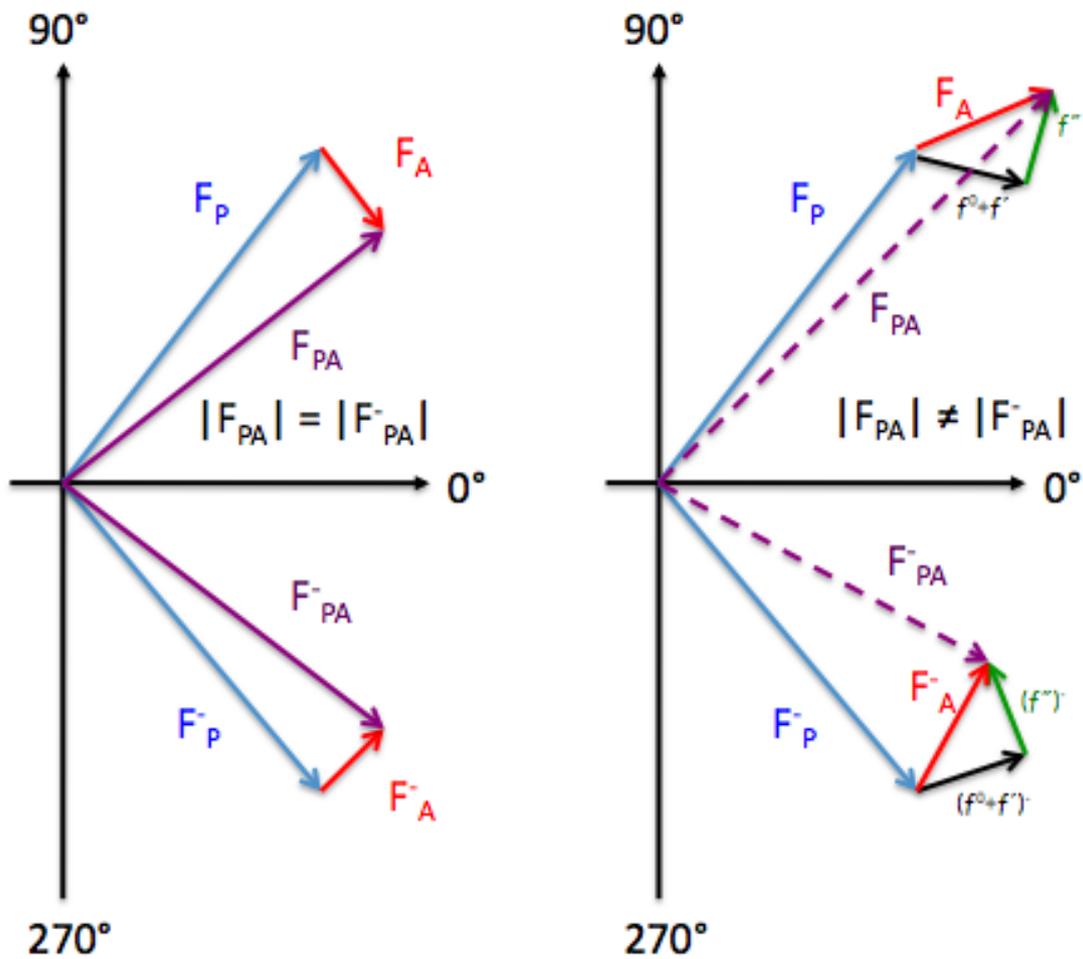


Figure 2.7: Breaking Friedel's Law. In a non-anomalous scattering crystal, the vector F_P represents the partial sum of non-anomalous atomic contributions, with F_A representing the contribution of the anomalously scattering atoms at an inappropriate wavelength for anomalous absorption. Friedel's Law holds and $F_{PA} = F^-_{PA}$, mirrored about the real axis.

Upon generation of anomalous signal, the real (f^0 and f') and (f^0 and f') contribution to F_A and F^-_A remain symmetrical about the real axis, but the imaginary anomalous component f''' (with the greatest anomalous difference between Bijvoet pairs) induces a 90° phase-retardation. This alters the F_A vector orientation with respect to F^-_A , leading to $F_{PA} \neq F^-_{PA}$ and the breakdown of Friedel's Law.

The difference in measured amplitude between broken Friedel mates is called the Bijvoet Difference:

$$\Delta F = |F^+| - |F^-|$$

Bijvoet pairs are true symmetry equivalents of Friedel pairs, which do not have identical amplitudes in the presence of anomalous scattering (Blow, 2003). Friedel pairs exist as $F(hkl)$ and $F(-h-k-l)$, whilst Bijvoet pairs are other pairs that can take the form $F(hkl)$ and $F(-hkl)$, dependant on space group. Therefore one Friedel mate forms a set of true symmetry equivalents with the same amplitude ($|F^+|$), while the other Friedel mate forms a set of equivalents that share a different amplitude ($|F^-|$) (Hendrickson, 1991). This aids the location of anomalously scattering heavy-atoms for MAD phase-calculation, as a Patterson map of ΔF^2 shows only the interatomic distance vectors of anomalous-scattering heavy-atoms.

2.6.5 Calculation of Heavy Atom Substructure Phases

The Difference Patterson described at the end of Section 2.6.3 shows $N(N-1)$ strong interatomic distances, where N is the number of peaks of anomalously scattering atoms. As in Section 2.6.1, this plot depends on the intensities (squared amplitudes) and is phase-independent. The location of heavy atoms is manually facilitated by the application of general position generators (tabulated in the International Tables of Crystallography) of the proposed space group of the substructure. Harker sections, that represent space-group dependant slices of the Patterson map, are used to assign atomic coordinates to the locations of heavy atoms in the substructure (Terwilliger & Berendzen, 1999).

In automated methods, the SHELX suite of programs is designed to facilitate the calculation of heavy atom substructure solution by direct methods, as described in Section 2.6.2. Through a process of locating heavy atom density peaks, calculated from reflection-originated structure factors that share phases from

similar Bragg lattices, phases are refined for the solution of a heavy atom substructure. This is similar to the solution of entire small molecule structures. Once phases for the heavy-atom substructure are known, their contribution to the overall phases of the heavy atom protein derivative can be used to calculate phase-angles for the protein structure factors (Sheldrick, 2008).

2.6.6 Phasing the Protein Structure.

As a useful aid to visualise phase-ambiguity, Harker diagrams can be drawn, utilising the structure factor vector relationship $\mathbf{F}_{PA} = \mathbf{F}_P + \mathbf{F}_A$ (Figure 2.8a). A structure factor (F_P) of known magnitude but unknown phase can lie anywhere on a circle of radius F_P centred on the origin of F_P . A second circle can be drawn with radius F_{PA} centred on a point determined by $-\mathbf{F}_A$ and the phase-angle (φ_A) determined by the solution of the marker atom substructure (Figure 2.8b). The potential phases of both of these circles lies anywhere upon their circumference. Where these two circles cross creates two potential phase-angle solutions (φ_{P1} or φ_{P2}) when the amplitude vectors for F_P are compared to the real axis (or X-axis = 0°). This creates a 2-fold phase-ambiguity (Taylor, 2003).

With the MAD data, peak f'' data contain anomalous phases of centric reflections offset from other data such as f' inflection by 90° as a product of heavy atom absorption (the source of breaking Friedel's Law). Therefore circles of radii F_{PA+} and F_{PA-} respectively are centred at origins of the F''_{A+} and (F''_{A-})-vectors, 90° from the origin of the $-\mathbf{F}_A$ vector. When applied to the Harker construct in Figure 2.8b, the diagram in Figure 2.8c reveals the true protein phase value, thus solving the phase ambiguity (Taylor, 2003). Such phasing calculations were made into equations by Karle and Hendrickson (Hendrickson et al, 1985) to allow for more linearised, and more importantly, computationally processed solutions.

A process of probability distribution analysis then allows for the refinement of original phases, accounting for factors such as circle-closure errors in Harker-construct-based calculations, positional errors in the heavy-atom substructure and the agreement of calculated phases from all datasets. The calculation of the

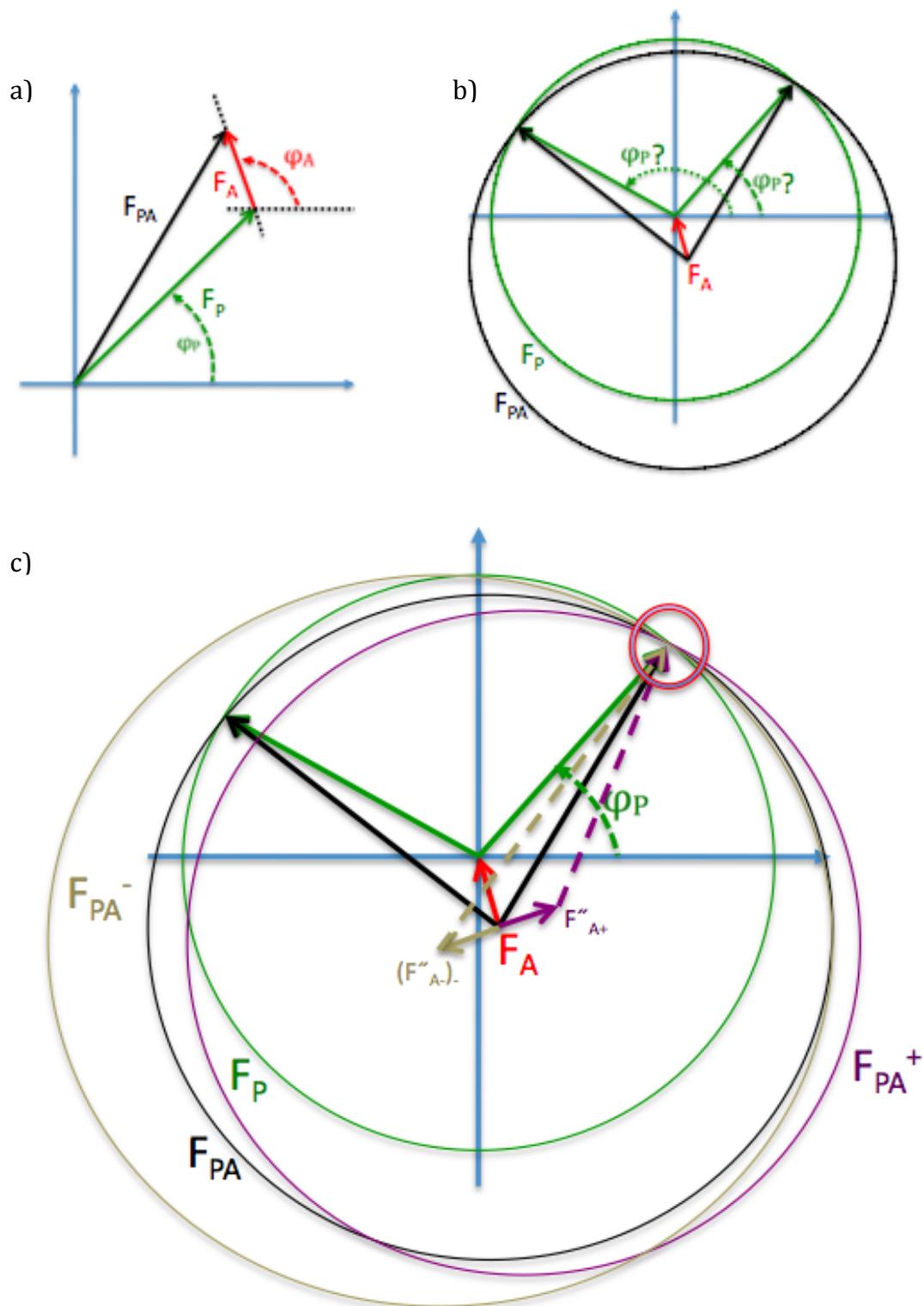


Figure 2.8: Phase solution using anomalous scattering and Harker plots.
 a) Relationship between protein (F_P), heavy atom (F_A) and heavy-atom derivative (F_{PA}) structure factor vectors for a generic reflection hkl . b) Circles of radius of $|F_P|$ and $|F_{PA}|$ magnitudes with unknown phase (φ) offset by $-F_A$ with known φ_A . Two possible φ_P are presented by intersecting circles. c) Anomalous scattering induces a $90^\circ \Delta\varphi$ due to the f'' absorptive contribution. Circles with radii of magnitude $|F_{PA+}|$ and $|F_{PA-}|$, with anomalous scattering origins at F''_{A+} and (F''_{A-}) , respectively, intersect at one correct phase angle φ_P . Adapted from Rupp, 2009.

best phase (φ_{BEST}) from soaked-in/co-crystallised heavy atoms is also subject to enantiomorphic variation, or ‘handedness’, resulting from the centrosymmetrical nature of the Pattersons used to calculate the heavy-atom substructure. In the case of MAD using selenomethionine (which is relevant to the experimental phasing used later in this thesis), the heavy atom substructure is non-centrosymmetrical, due to being subject to the chirality of the protein structure, and as such there is a 50% chance of obtaining the right ‘handedness’ (Sheldrick, 2008). An electron density map calculated on the incorrect enantiomorph will give uninterpretable noise.

2.7 Calculations and Modification of Electron Density

2.7.1 Calculation of Electron Density

Electron density, annotated as $\rho(\mathbf{r})$ or $\rho(x, y, z)$, can be calculated by the Fourier transform of phased structure factor amplitudes. For a more in-depth discussion of Fourier transformation the author refers the reader to the paper by Lynn Ten Eyck (Ten Eyck, 1985). Once the phases of the complex structure factors have been solved (also considered as resolving the ‘phase-problem’), they can be combined with the structure factor amplitudes and used in the following calculation:

$$\sum F(\mathbf{h}) \cdot \exp[-2\pi i(\mathbf{h} \cdot \mathbf{r}) + i\varphi(\mathbf{h})] = \rho(\mathbf{r})$$

where $F(\mathbf{h})$ = amplitude of each measured reflection and $\varphi(\mathbf{h})$ = the phase for each available reflection. Calculated electron density can now be back-Fourier transformed (FT^{-1}) to structure factors (Cowtan & Main, 1996) without any loss of information, via:

$$\text{FT}^{-1} = F(\text{hkl}) = V \sum_{x, y, z=0}^{a, b, c} \rho(x, y, z) \cdot \exp [2\pi i (\text{hx} + \text{ky} + \text{lz})]$$

and back to electron density by:

$$FT = \rho(x, y, z) = \frac{1}{V} \sum_{hkl = -\infty}^{+\infty} \sum \sum F(hkl) \cdot \exp [-2\pi i (hx + ky + lz)]$$

Once an electron density map has been calculated, the quality of it can be assessed. Visual investigation allows confirmation of correct enantiomorphs if the maps are clearly interpretable and combined with the identification of secondary structure elements such as α -helices and β -sheets allows for the judgement of 'handedness' (i.e. a right-handed helix is correct) and the assessment of quality due to phasing or source data. Once an initial electron density map has been calculated, several methods can be employed to improve it prior to model building, and a couple (not an exhaustive list) are discussed in the next sections.

2.7.2 Solvent Flattening

This method of density modification relies on the different physical properties between the immobilised protein molecules in a crystal and the disordered solvent that surrounds them (Wang, 1985). On average, 50% of a protein crystal consists of solvent channels (Heras & Martin, 2005). This technique exploits the fact that solvent regions of electron density should be essentially 'flat' (that is to say mainly water, with maybe slight variation due to the components of the crystallisation condition). Pure water has electron density of approximately $0.33 \text{ e}^-/\text{\AA}^3$ (a concentrated salt solution will be higher than this) and protein has an averaged approximate electron density of $0.44 \text{ e}^-/\text{\AA}^3$. These values are calculable via the application of Avogadro's number to each molecule (or average protein) mol/cm^3 with correction to \AA values, to give number of molecules per \AA , which can then be used to divide the number of e^- per molecule to give $\text{e}^-/\text{\AA}^3$.

An initial electron density map that has distinguishable, contiguous solvent regions compared to the molecular envelope is the minimum requirement for solvent flattening. With a 3-D grid placed over the density, spheres of defined

radius ($\sim 10 \text{ \AA}$) are plotted at grid points and the density within the sphere assessed against a predefined cut-off value. If this value is not reached, then the contents of the sphere are deemed part of the solvent region. Assuming the map is of a useful quality, definition between solvent and protein should be found (Wang, 1985). Maps that fail this process are likely to be of no use anyway, with re-processing data, re-calculating phases or even re-collecting better data likely to be the source of improvement.

The solvent region is then set to the solvent average ($0.33 \text{ e}^-/\text{\AA}^3$ if the crystal were grown from pure water, more if from salt solution) and new structure factors are calculated by inverse Fourier transform (FT^{-1}). The new structure factor phases are then combined with the original phases for FT back to electron density. Because the solvent regions really should be flat and not noisy, this leads to a map with improved phases therefore better defining the regions of protein. This process can then be repeated until a more interpretable map is achieved. Parameters involved may need adjusting, such as the radius of solvent-masking spheres, in order to explicitly flatten solvent and not regions of protein. The calculation of a Matthews probability (in which the solvent content of a given unit cell is estimated based upon the mass and number of copies of the protein present) may aid in the judgement of excessive or conservative solvent-masking (Kantardjieff & Rupp, 2003). Often the resulting electron density map post-solvent flattening is a viable map for manual or even automated model building.

2.7.3 NCS Averaging

Many unit cells contain asymmetric units (ASUs) that contain multiple copies of the protein, sometimes in the form of oligomers. This Non-Crystallographic Symmetry (NCS) means that the diffraction patterns collected contain redundant information due to this local symmetry, and the electron density maps show two or more approximately identical copies of the same molecule. If averages of these copies are taken and combined, then the electron density may be improved. The alignment of such NCS averaged molecules requires the application of transformation functions, in the case of MAD or other heavy-atom substructure-

phased density maps, this can often be facilitated by the rotation and translation of the heavy-atom substructure. In other cases, alignment of Patterson map peaks can calculate the required transformations. In this short introduction to the basics of crystallographic theory, the statistics of NCS averaging transformations are not discussed, but the author directs the reader to the text of Bernhard Rupp (Rupp, 2009) for a detailed description.

As with solvent-flattening (Section 2.7.2) a Matthews solvent-probability calculation helps determine the number of molecules present, and thus a target for the number of molecules to be averaged. In modern times, with an ever-expanding PDB (Berman et al, 2000), the calculated solvent content can be compared to thousands of other structures at the same resolution to more accurately estimate the number of molecules present with such a calculated Matthews coefficient (Kantardjieff & Rupp, 2003).

With the number N of identical molecular envelopes aligned (in real space), the improvement in signal:noise ratio is proportional to $N^{1/2}$. With this improvement in signal:noise increasing with each copy in the ASU, it can be seen how powerful this method can be for highly repetitive oligomers, especially with crystallised, highly symmetrical virus particles (Stehle et al, 1996).

2.8 Structure Refinement

Upon the calculation of a phased electron density map that has been subjected to density modification, a chemically viable protein structure can be modelled into the lumen of regions of connected electron density. This section describes some of the methods used to refine and validate an accurate model that remains true to the experimentally observed data. The techniques described are not an exclusive list of methods and are applied after the building of an initial model, either by manual or automated methods not described here. For model-building methods and processes, the author directs the reader to the program literature of CCP4i for the running of refinement programs (Potterton et al, 2003), ARP/wARP for automated model fitting to electron density (Morris et al, 2003)

and Coot for the manual building of protein models into experimentally determined electron density (Emsley & Cowtan, 2004).

2.8.1 The Maximum-Likelihood Function

Maximum-likelihood (ML) works on the basic principle that the best model built is the one that is most consistent with the experimentally derived observations (Terwilliger, 2000). The model built into the observed density can be changed to match the experimental data (practically done within a 3-D model-building computer program), and thus its likelihood for correctness increases and the model is considered better. This in effect 'sharpens' the probability of correctness. The refinement process can also be restrained by the application of set bond-lengths, bond-angles, isotropic or anisotropic B-factors/temperature-factors and occupancies, with weights applied to prevent the building of models that fall out of chemically acceptable parameters (Vagin et al, 2004).

Translation/Libration/Screw (TLS) restraint-parameterisation is a more advanced interpretation of B-factor restraints in which molecular motion (in terms of horizontal and vertical movement combined with potential rocking-and-twisting motion) between a defined region of atoms is weighted (Winn et al, 2003). Constraints on the number of parameters are also applied, perhaps only allowing for variation in a certain number of restraints, such as allowing only torsion angle-changes instead of all x, y, z coordinates.

The repositioning of atoms within the electron density alters the calculated phases, as the position of atoms within Bragg planes gives the origin of phase differences from the experimental data. This can in turn improve the structure if an atom is moved in one direction, and perturb accurate phases if the atom is modelled further away from a correct Bragg plane. As such, F_{obs} are assigned a figure of merit statistic (m where $m = \langle \cos(\Delta\phi) \rangle$) that represents the probability of a phase angle to be correct in relation to the probability-averaged, best F_{calc} (Rupp, 2009). In addition, to account for positional errors a term of variance (σ_A) is applied to the structure factor probability distribution. σ_A is determined by cross-validating structure factor amplitudes calculated from a partially built

model with the complete set of amplitudes from the experimentally observed structure factors (Read, 1997). The better the model being built into the electron density is, the lower the σ_A variance that is calculated.

When a model in which all parameters are variable is built into the initial density (calculated from observed structure factor amplitudes, F_{obs} , with experimental phases), it allows for the calculation of structure factors from the model placed into the density (F_{calc}). The best fit between these two sets of values can then be determined by a mathematical process of likelihood functions that assess whether the probability of the F_{calcs} is proportional to the experimentally calculated F_{obs} . These functions also attribute a scaling factor (D) to F_{calcs} in relation to F_{obs} . Another way of looking at this is to ask whether the F_{calcs} could be back calculated to give the amplitudes measured from the square-roots of the original diffraction pattern intensities.

ML is carried out over many cycles with variations in restraints and constraints, with new electron density maps via FT. However, these new maps are subject to newly calculated differences between F_{obs} and F_{calcs} , and are thus termed difference maps with:

$$(m|F_{obs}| - D|F_{calc}|) \exp(2\pi i\varphi_{calc})$$

where m is the figure of merit for F_{obs} phases, D is a scaling factor and F_{calc} and φ_{calc} are structure factor amplitudes and phases calculated from the model. ML values are derived from the gradient of the likelihood function between the observed structure factors and the ones from the current model.

In practice, model-building programs represent positive (likely region of unaccounted density) and negative density (should be no density) calculated by these differences to facilitate further model-refinement. However, it is worth considering the quality of data before any confident refinement by ML can occur. Poor data means poor structure factors (F_{obs}), which also means poor phases.

Therefore poor data also means poor refinement, and as such refinement procedures require validation.

2.8.2 Validation of Refinement by R_{work} and R_{free}

The R-factor measurements are an assessment of how well the crystallographic model agrees with the experimentally observed data, that is they measure how well a refined structure predicts observed diffraction. In order for this to function, a FreeR set of experimental data (usually 5%) is excluded from any model refinement and used for cross-validation of the model with unrefined experimental data (Brunger, 1993). The working, refined data set (giving the R_{work} value) and the unrefined FreeR dataset (giving the R_{free} value) are then calculated for agreement against the entire model-fitted data using the calculations:

$$R_{\text{free}} = \frac{\sum_{h\text{free}} |F_{\text{obs}} - kF_{\text{calc}}|}{\sum_{h\text{free}} F_{\text{obs}}}$$

and

$$R_{\text{work}} = \frac{\sum_{h\text{free}} |F_{\text{obs}} - kF_{\text{calc}}|}{\sum_{h\text{free}} F_{\text{obs}}}$$

Both the R_{work} and R_{free} should improve (decrease) as refinement progresses and the model becomes more complete, however the R_{free} value will always be greater than R_{work} . A change to the model that improves the way it interprets physical reality (i.e. the actual data) will be reflected in the way the model fits to the data contained in R_{free} , which was not used in the refinement process (Brunger, 1993). This can also be used as an assessment of restraint requirements, as if a certain number of enforced parameters give no change in

R_{free} compared to a less or unrestrained model then a model is being over-parameterised, maybe even to the point of bias. A decreasing R_{work} with a stable or increasing R_{free} is also an indication of bias, perhaps symptomatic of changes being made that the experimental data does not agree with. A judgement of R_{free} may also be useful in determining the optimum modelling of a protein structure too as a rapid loss of improvement in R_{free} , regardless of restraint or constraint variations, may mean that the model now fits the experimental data to the best ML.

Chapter 3: Materials and Methods

3.1 Materials

Throughout the experiments carried out in the construction of this thesis all equipment and reagents used were common to most laboratories and available for purchase from a range of suppliers. The grades of chemicals used were matched to the specific requirements of individual experiments, and as such any deviation from standard usages will be mentioned later on in this work on a case-by-case basis. As it was generally considered that the source of chemicals would have a negligible effect on the outcome of the experiments undertaken, an extensive description of all reagents does not appear in this work.

3.2 General Methods

This section documents the standard methods used during the experimental processes of this thesis. Any variations from the typical details of each method are described here.

3.2.1 PCR

PCR (Bartlett & Stirling, 2003) was used for 2 different applications: (I) amplification of inserts for cloning, and (II) colony PCR to confirm the insert presence in the purified plasmid post-cloning. In both cases the following reaction conditions or multiples thereof were used, and were occasionally varied by the addition of 10% DMSO to reduce primer secondary structure formation:

0.5 μ l sense primer (5 μ M)
0.5 μ l antisense primer (5 μ M)
0.5 μ l genomic template (approx 2 μ M)
6.25 μ l 2xBioMix
4.74ul sterile, deionised water
12.5 μ l reaction

BioMix is a premixed PCR cocktail containing *Taq* DNA polymerase, reaction buffer and free dNTPs available from Bioline. This reaction mix was subjected to the following conditions using a Techne TC-312 Thermocycler. A variation of the annealing temperature by no more than 5°C was occasionally employed to optimise specific binding of some primers:

94 °C Initial Denaturation	}	30 Cycles
94 °C Denaturation		
55 °C Annealing		
72 °C Extension		
72 °C Final Extension		

In the case of colony PCR the genomic template was replaced by boiled cell-lysate from colonies picked from an agar plate and resuspended in 30 µl sterile water.

3.2.2 Agarose Gel Electrophoresis

Separation of DNA vectors or fragments for analysis or purification was carried out by running samples on a 1% agarose gel and visualising their migration through the gel matrix. Gel production was commonly achieved by melting 0.5g agarose into 50ml of TAE buffer (40mM Tris pH 8.0, 20mM glacial acetic acid, 1mM EDTA) into a 200 ml Duran using a microwave, and adding 5 µl of x10,000 GelRed or ethidium bromide once the vessel had cooled down enough to comfortably touch with un-gloved hands.

Ethidium bromide was added to intercalate with the base-pairs of the DNA under investigation to allow fluorescent visualisation of migration under UV light. The melted agarose was allowed to cool slightly to prevent the inhalation of vapours, as ethidium bromide is a carcinogenic agent. GelRed is available from Biotium and is a safer alternative to ethidium bromide that cannot cross cell membranes and leaches-out less from gels to give improved resolution of results.

The gels were poured into BioRad Mini-Sub Cell gel-beds with a comb inserted to form wells, and were allowed to set for approximately 20 minutes. Once set, the gels were submerged in TAE buffer so that a thin layer of buffer covered the gel's surface. Appropriately diluted samples were combined with loading buffer (0.25% Bromophenol Blue, 30% Glycerol) from a x6 stock and carefully loaded into the designated well. Gels also had a molecular weight marker lane included such as Hyperladder I from Bioline or 1KB-Geneladder from NEB in order to identify DNA fragments by their migration compared to fragments of known molecular weight. Gel-tank lids were connected to their correctly orientated electrodes and were run at 100V. Time was varied to account for the separation of differently sized samples. Smaller PCR products were run for 40 minutes whilst larger samples such as digested plasmids were run for 60-70 minutes. After the electrophoresis had concluded, the gels were visualised using a UV lamp and a photographic record was made using a digital camera, a Polaroid camera, or a photomultiplier set-up when available.

3.2.3 SDS-PAGE

This method was used to analyse the presence and purity of proteins by separating proteins based upon their individual molecular weights. The protocol used was adapted from the protocol established by Laemmli (Laemmli, 1970). All SDS-PAGE gels in this thesis were made to have a 13.5% resolve with 5% stack for well formation, following the recipe:

<i>Reagent</i>	<i>Vol used for 13.5% Resolve</i>	<i>Vol used for 5% Stack</i>
30% Acrylamide sol ⁿ	2.81ml	0.63ml
1M Tris pH 8.8	2.35ml	
1M Tris pH 6.8		0.47ml
Deionised water	0.97ml	2.6ml
10% SDS	62.5µl	37.5µl
TEMED	6.25µl	3.75µl
10% APS	62.5µl	37.5µl

Samples were created to contain 10-15 µg of protein, mixed to dilute x4 LDS Loading Buffer (40% glycerol, 4% lithium dodecyl sulfate, 0.025% phenol red, 0.025% Coomassie G250, 2mM EDTA, 0.8M triethanolamine-Cl pH 7.6 – available from Invitrogen) and x10 reducing agent (1M DTT) and made up to 30 µl using deionised water. The samples were then boiled for 2 minutes. After centrifuging at 13kpm for 20 seconds in an MSE Micro-Centaur rotor the samples were loaded into the wells of the stacking gel alongside the Mark 12 molecular weight ladder from Invitrogen.

SDS-PAGE gels were ran at 60V for 10 minutes to ensure samples entered the resolves at the same time, before being run at 200V for approximately 45 minutes or until the running-front marker Coomassie G250 began to run off the bottom of the gel. Gels were stained overnight in 1:1:1 methanol:acetic acid:water with 0.1% Coomassie Blue and washed then soaked in destain (10% methanol, 10% acetic acid) overnight, changing into fresh destain several times. Gels were then dried by soaking 2 squares of DryEase cellophane from Invitrogen in drying solution (4:2:2:1 methanol:ethanol:isopropanol:glycerol) and using these to sandwich the drying solution-rinsed gel, which was then clip-sealed into a drying-cassette and left to dry naturally. SDS-PAGE gels were then scanned to be included in this thesis.

3.2.4 Determination of Protein Concentration

Protein concentrations were determined by the Bradford Assay (Bradford, 1976), in which the Bradford reagent binds to the hydrophobic regions of proteins causing an increase in absorption at 595nm proportional to protein concentration. Samples were analysed in a spectrophotometer using the following composition:

200 µl Bradford Reagent
800 µl deionised water
1-10 µl protein sample

The following equation was used to establish the concentration of the protein samples from the absorption read from the spectrophotometer:

$$\frac{\text{OD 595nm} \times 15}{\mu\text{l protein}} = \text{mg/ml}$$

where 15 L mol⁻¹ cm⁻¹ is the absorption coefficient for the Bradford reagent.

3.3 Bioinformatical Methods

Throughout this work, nucleotide and protein sequences were extracted from the databases of the NCBI or GeneDB.org. Gene and protein sequences were analysed for comparison of homologues and the detection of their presence in specific bacterial species using the NCBI BLAST server (Altschul et al, 1990). Gene sequences were analysed for potential restriction sites using the NEBcutter server (Vincze et al, 2003), whilst protein sequences were analysed for predicted regions of disorder using the RONN server (Yang et al, 2005) and were input to the ExPASy ProtParam server for calculation of mass, theoretical pI and atomic composition (Gasteiger et al, 2003). Prof. Pete Artymiuk generated genome-wide, individual Kyte-Doolittle Scale hydrophobicity plots (Kyte & Doolittle, 1982) using a program he wrote that extracted protein sequences from the translated *Burkholderia pseudomallei* genome data from the NCBI database. He then wrote a script to automatically create hydrophobicity plots using the Staden HYDROLOT program (Staden, 1994). Fold-recognition investigations were carried-out using the GenTHREADer server from Psipred (Jones, 1999a).

3.4 Cloning of Constructs

3.4.1 Oligonucleotide Primers

All the primers designed for this thesis were designed using the program AmplifX 1.5.4 (<http://ifrjr.nord.univ-mrs.fr/AmplifX>), which also ran a mock PCR reaction based upon calculations of T_m, potential secondary structure predictions and self-dimerisation likelihoods against the target gene sequence.

The following primers were designed for the production of PCR fragments, and were synthesised by Sigma-Aldrich:

Name	Sequence (+Restriction Site)	Gene	Vector
BPSL1204-F	ATGCAGCCGTGGCAGCAATTC	BPSL1204-MA	pETBlue-1
BPSL1204-R	GCGTGTCAAATGACTTCGGGCG	BPSL1204-MA	pETBlue-1
wtBCAL2351F	TTAAATTTCCATGGGACGCATTTCGCATTCCCGCG (NcoI)	wtBCAL2351	pBAD/hisB
wtBCAL2351R	TAGTAGGTTTCGAATCAGAACAGAACAGGGTGCGGCGAT (HindIII)	wtBCAL2351	pBAD/hisB
nH2351-F	AATTACATATGCAGCAATACCAGGCCGGCCA (NdeI)	His-tagged BCAL2351-	pET28a
nH2351-R	ATTAACCTAGGTCAGAACAGGGTGCGGCGAT (BamHI)	His-tagged BCAL2351-	pET28a
wt0998F_	CATTTAGACCATGGGATCCGCATCCGACCTCACTTC (NcoI)	BPSL0998	pBAD/hisB
wt0998R	ATTAAAGCTTCGAATCACTTCACGCGCATGCCGG (HindIII)	BPSL0998	pBAD/hisB

3.4.2 Sub-cloning of PCR products

PCR products were purified from the PCR reaction mix using either the Qiagen PCR Purification Kit, or were agarose gel-extracted using the Qiagen Gel Extraction Kit if the PCR produced contaminating bands. Both the PCR insert and the corresponding purified vector of choice were treated with the same two enzymes in a double-digest reaction at 37°C for 1-6 hrs. Pure plasmid-vectors were obtained using the Qiagen Miniprep Kit from maintained cell-line stocks. Digested plasmids were purified by Qiagen Gel-Extraction and inserts were purified by Qiagen PCR Purification. Some problematic digestions were further treated with Antarctic Phosphatase from NEB to remove 5' phosphate groups to prevent plasmid re-circularisation. Overnight ligations were carried out using T4 Ligase from NEB by incubation overnight at either 4°C, 16°C or a using a gradient

of 0-20°C. All kit-based methods were carried out as directed by the manufacturers manual.

In cases where blunt-end-ligation into the pETBlue vector was used PCR products were amplified and then gel extracted if not already pure, then ligated into the supplied, pre-linearised vector using T4 ligase overnight at 4°C. After transforming into the sub-cloning strain, colonies with correctly orientated inserts were identified using the 3'PCR primer and the upstream-of-insert pETBlue-1 specific 5'T7F primer (Studier & Moffatt, 1986).

3.4.3 *E. coli* Strains Used For Sub-cloning and Over-expression

The following bacterial strains were used in the cloning and over-expression of target genes:

Sub-cloning strains and genotypes:

NovaBlue – endA1 hsdR17 ($r_{K12^-}m_{K12^+}$) supE44 thi-1 recA1 gyrA96 lac F'[proA⁺B⁺ lacI^qΔM15::Tn10] (Tet^R) (from Merck Biosciences)

DH5α - fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 (from NEB)

XLBlue – recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI^qΔM15 Tn10 (Tetr)] (from Stratagene)

Over-expression strains and genotypes:

Tuner – F⁻ompT hsdS_B ($r_B^-m_B^-$) gal dcm lacY1 (from Merck Biosciences)

BL21(DE3) - F⁻ompT hsdS_B ($r_B^-m_B^-$) gal dcm (DE3) (from Invitrogen)

Top10 - F⁻ mcrA (mrr-hsdRMS-mcrBC) 80lacZM15 lacX74 recA1 ara139 (ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG (from Invitrogen)

All strains were transformed with ligated plasmids as suggested in the manufacturers literature.

3.5 Protein Expression

3.5.1 Media and Agar

For all standard bacterial growths on all scales liquid LB media was used (Bertani, 2004), using the following contents per litre of media:

10 g Tryptone

5 g Yeast Extract

10 g NaCl

made to 1 L with deionised water

For the creation of solid LB agar media, the above list was supplemented with 15 g Bactoagar per litre. All standard LB media were autoclaved to sterilise before use. This also promoted the solubilisation of agar into the media to allow proper mixing. Stocks of liquid and solid media were stored at room temperature in an aseptic room with a UV bulb used when unoccupied. All media were handled using proper aseptic technique to avoid contamination wherever possible.

3.5.2 Minimal Media for Seleno-methionine Incorporation

For growth experiments attempting to incorporate seleno-methionine into over-expressed protein, minimal media was created to cause the inhibition of methionine synthesis by the induced *E. coli* and also to promote the forced inclusion of seleno-methionine into protein synthesis in the place of naturally occurring methionine. The following chemicals were used to create the minimal media per litre:

10.5 g Dipotassium phosphate

1 g Ammonium sulphate

4.5 g Potassium dihydrogen phosphate

0.5 g Tri-sodium citrate

5 ml 100% Glycerol

0.5 g Adenine

0.5 g Guanosine

0.5 g Thymine

0.5 g Uracil

made to 1 litre with deionised water

The above mixture was split into two 2-litre baffled-flasks containing 490 ml each, or multiples thereof for larger growths, and was then autoclaved to sterilise. Baffled-flasks were used to promote better aeration as a combatant against the negative effects on growth of using lower nutrient media. To the above autoclaved bulk of the media, the following non-autoclavable components were added to each 490 ml:

1 ml 50 mg/ml L-Lysine

2.5 ml 20 mg/ml L-Phenylalanine

1 ml 50 mg/ml L-Threonine

1 ml 25 mg/ml L-Isoleucine

2 ml 12.5 mg/ml L-Leucine

1 ml 25 mg/ml L-Valine

1 ml 0.5 mg/ml Magnesium Sulphate

1 ml 2 mg/ml Thiamine

1 ml 20 mg/ml Seleno-methionine

All stock solutions were stored at -20°C with the exception of seleno-methionine which was made fresh every time.

When used for growth of bacterial culture, all media were supplemented with the appropriate antibiotic at the established effective concentration for the specific strain used.

3.5.3 *E. coli* Over-expression – General Overview

As a standard all initial expressions were carried out in a similar fashion. Specific details of individual over-expressions are given in the appropriate Results chapter. 250 µl of a 50 ml overnight culture was used to 5%-inoculate 4.75 ml fresh LB media, which was allowed to grow to an OD_{600nm} of 0.6-0.7 at 37°C on a 250 rpm shaking platform. OD was measured against a blank of sterile LB media in a spectrophotometer. The culture was then induced for 3 hours with either 1 mM IPTG for genes under the control of the *lac* operon or 0.01% arabinose for genes under the control of the *ara* operon. A non-induced control culture was carried out alongside induction samples. Cells were pelleted by centrifugation at 5000 rpm at 10°C then the supernatant was decanted and pipetted off. Pellets were subjected to Novagen BugBuster fractionation as directed by the manufacturer's manual to generate soluble and insoluble fractions that were analysed for protein over-expression by SDS-PAGE gel. Based upon the results of the gel, over-expressions were optimised by varying one or more of: induction OD, concentration of inducing agent, time-period of induction and temperature of induced samples. The specific conditions for each protein target are described individually later on in this thesis.

After optimum over-expression conditions were discovered for achieving sufficient soluble protein, the growth was then scaled up and repeated in bulk volumes of multiples of 500 ml. After large-scale growth, cells were pelleted in 500 ml centrifugation canisters by centrifugation at 5000 rpm at 8°C using a Beckman Avanti and a J10.500 rotor. Then the pellet was transferred to a 50 ml Falcon tube by scooping out most of the pellet and re-suspending the rest in the lowest volume of media possible and pipetting into the Falcon. Several canisters' pellets were combined into one Falcon, which was then spun for 20 minutes at 5000 rpm at 8°C in a Sigma 11133 rotor to re-pellet into the stock cell-paste. Supernatant was removed and the cell-paste stored at -80°C until used for protein purification. Before initial use in purification, a Bugbuster fractionation and an SDS-PAGE analysis were carried out on a sample of the paste to ensure the scale-up did not change the desired over-expression profile.

Large-scale over-expression of seleno-methionine-incorporated protein began with the growth of cells to an OD_{600nm} of 0.6 in LB as with the standard protocol. However at this point, cells were pelleted by centrifugation for 20 minutes at 5000 rpm using a J10.500 rotor at 8°C and the LB supernatant was decanted off and the pellet washed gently in minimal media so as to not re-suspend cells, which was then gently removed by decanting and pipetting. Each pellet was then re-suspended in 50 ml fresh minimal media taken from 500 ml in a 2 L flask, now containing 450 ml. The 50 ml re-suspension was then added back to the flask of origin, and the cells allowed to equilibrate and begin growth again for 30 minutes at the optimum over-expression temperature. Cells were then induced and left for 21 hrs to over-express. This extended expression period was required for sufficient yield of both cell count and maximum protein produced per cell, due to the growth-lagging effect of minimal media and the cytotoxic effect of seleno-methionine presence (Strub et al, 2003).

3.5.4 Glycerol Stocks

In order to maintain cell-lines proven to be over-expressing target proteins, and to facilitate future culturing, stocks of the original overnight cultures were combined with glycerol to a final concentration of 30% (700 µl culture vortexed with 300 µl 100% glycerol). These stocks were then stored at -80°C for up to 18 months and used to create 50 ml overnight cultures for future large-scale over-expressions. Fresh 30% glycerol stocks were created after every culture to replace older stocks.

3.6 Protein Purification

The following procedures are a generalisation on the techniques used in the experimental protein purifications carried out during the research on which this thesis is based. The specifics of individual purifications are given later on in this body of work.

3.6.1 Preparing Cell-free Extract

In order to achieve accessible soluble protein, cell membranes, insoluble proteins and other cell-debris needed to be separated from the soluble cytoplasm of the cell-paste used. The weight of cells to be used was re-suspended from the pellet kept at -80°C in 1:10 g:ml of a suitable buffer such as 50 mM Tris pH8.0. For example this meant that 3 g of cell-paste was re-suspended in 25-30 ml of buffer. This re-suspension was split between 3x20 ml pots, and on a bath of ice each pot was subjected to 3x20 seconds sonication at 16 microns using an MSE Soniprep 150 machine and a blunt-ended sonication probe. Each pot was returned to ice after each 20-second sonication to prevent overheating, thus avoiding denaturation of the protein sample. Pots were combined into a centrifuge tube and disintegrated cell-debris was pelleted by centrifugation at 24000 rpm for 15 minutes at 8°C using a Beckman Avanti and a JA25.50 rotor. The soluble protein supernatant was decanted away from the pellet and used immediately for the next step of purification.

3.6.2 Ion-Exchange Chromatography

Ion-exchange chromatography utilises surface net charge differences present between different proteins. As such it separates proteins based upon their ability to resist ionic-perturbation from a charged resin when a salt-gradient is introduced to the column used. Cell-free extract was applied to a DEAE anion-exchange column pre-equilibrated with a salt-free buffer such as 50mM Tris pH8.0 by gentle injection from a 50 ml Superloop connected to a ÄKTAprime purification system. The rate of sample application was determined by the pressure-capacity limits of the specific column used. After samples were applied and the flow-through collected for analysis later on, the column was washed with 3 column volumes of salt-free buffer until no more protein was eluted, detectable by a UV trace. All proteins left were bound to the positively charged column via their anionic surface charges. A 300ml salt gradient of 0-0.5 M NaCl in the same buffer as that used to wash was then applied to gradually increase the Cl⁻ concentration in the column to perturb protein:column interactions. The more negatively charged the protein, the more salt required to prevent interaction

with the column, thus allowing separation based upon charge. 4 ml fractions were collected, and those that fell under the peaks on the UV-trace were analysed by SDS-PAGE for the presence of the target protein. The fractions that contained the correct peak were combined for the next stage of purification.

3.6.3 Ammonium Sulphate Precipitation

By adding high concentrations of $(\text{NH}_4)_2\text{SO}_4$ it was possible to alter the solubility of a target protein to purify it away from contaminants by precipitation.

Combined fractions from a purification step such as DEAE ion exchange were diluted with 4 M $(\text{NH}_4)_2\text{SO}_4$ to a final concentration of 1.6 M. Precipitated proteins were spun into a pellet by centrifugation in a Beckman Avanti JA25.50 rotor at 24000 rpm at 8°C for 15 minutes. The supernatant was kept for analysis, and the pellet re-suspended in the minimal possible volume of stabilising buffer. The re-suspension was spun at 13000 rpm in a desktop micro-centrifuge for 5 minutes to re-pellet any permanently insoluble material, and then the soluble supernatant was taken for SDS-PAGE analysis of purity and further purification if required.

3.6.4 Gel-Filtration

Gel filtration or size-exclusion chromatography was used to separate proteins based upon their molecular weight and stoichiometric ratios. By passing a mixed protein solution through a gel-matrix of porous beads, a larger protein would access fewer of the pores of the gel, therefore the protein eluted from the column faster via a shorter elution route. If a protein naturally formed a polymeric complex, this would be reflected in the elution profile, with estimated molecular weights calculable against a standard curve. A Superdex200 GF column from GE Healthcare was pre-equilibrated with the stabilising buffer of choice (50 mM Tris pH8.0, 0.5 M NaCl as a standard) and the sample was applied in a volume of 1-2 ml at a rate of 1 ml/min. At the same rate, the column was run for its entire inclusion volume, with 2 ml fractions collected and monitored by a UV trace for protein peaks. Peaks were assessed by SDS-PAGE for presence of target protein,

and fractions under the peak combined for the next stage of purification or for experimental use as a pure protein sample.

3.6.5 Viva-Spin Concentration

In order to achieve highly concentrated protein solutions for crystallographic experiments and to perform buffer-exchanges, Vivaspin centrifuge columns from Sartorius Biotech were used. The cut-off molecular weight of the columns filter was varied according to the molecular weight of the target protein in order to maximise sample-retention and purity. Samples were concentrated by centrifugation through the filter at 4500 rpm in a Sigma 11133 rotor at 5°C. The period of time concentration occurred for varied between experiments, and was determined by the desired final concentration as measured by the Bradford Assay. This method was also used to exchange the buffers of protein samples when required, as the much lower molecular weight buffers easily passed through even the lowest molecular-weight filter. Replacing the old buffer in the concentrated protein sample with many-fold more volumes of new buffer and re-concentrating could change the buffers. This was repeated several times to complete the exchange.

3.7 Crystallographic methods

3.7.1 Initial Screening

Using the HydraII crystallisation robot from Matrix Technologies, 96-well sitting-drop vapour-diffusion experiments were used to perform initial crystallisation experiments using a selection of pre-mixed screens bought from Qiagen (JCSG+, Classics, PACT, PEG, Ammonium Sulphate, MPD and pHClear screens). The robot pipetted 100 µl of each screen condition into a micro-plate's reservoir-well, and then ~200 nl was transferred into the sitting-drop indented-ledge of each well. The nano-needle then dispensed ~200 nl of the supplied pure protein solution into each sitting-drop ledge and the completed microplates were sealed with a transparent film and spun at 2000 rpm for 1 minute to ensure sitting-drops were mixed at the bottom of each ledge. Trials were left at either 7°C or 17°C for

water-vapour equilibration and with the aim of crystallising the target protein. The micro-plates were gently checked under a light-microscope once every 2 days for the first fortnight, then weekly after that.

3.7.2 Siliconisation of Coverslips

100-200 glass cover slips from Menzel-Gläser were placed into a glass-dish and the dish was rested on top of a glass beaker containing 10 ml of dimethyldichlorosilane siliconising solution. This was placed into a glass-vacuum chamber, and the entire procedure was carried out in a fume cupboard with the extraction-fan on and the shield down as far as practically possible. Gloves were worn due to the toxic nature of the siliconising solution. With the chamber lid in place, a tube was connected from the chamber to a Buchner flask, from which a tube ran intersected by a water pipe. As water was run, air was removed and the chamber interior became a vacuum. The chamber exhaust was sealed, and the cover slips left in the siliconising vacuum for 20 minutes, after which the chamber was depressurised by releasing the exhaust and the cover slips removed for storage in a sterile Petri dish. Each cover slip was polished using lens-tissue in an air-stream before use in crystallisation experiments.

3.7.3 Condition Optimisation

Following the identification of successful crystal growth in the initial screens, conditions were optimised to focus in on the best crystallisation condition by varying the reservoir components and volume, drop size and drop ratio, protein concentration used and incubation temperature in 24-well hanging-drop trays. Drops of protein and reservoir solution were mixed together on a siliconised coverslip, gently inverted to suspend the drop over the reservoir and then sealed as a lid to the well by using vacuum-grease to line the lip of the well. Initial optimisations were typically done in duplicate to allow investigation of the effect of temperature on crystallisation, with one duplicate left at 17°C and one left at 7°C during the length of the experiment. Sometimes several rounds of condition variation and optimisation were required to focus in on the best possible crystal growth.

3.7.4 Mounting of Crystals onto an X-ray Source

Once growth of sufficiently large, individual crystals had been achieved they required looping from the mother-liquor that they had grown in so they could be put onto the in-house X-ray source for testing. Pre-mounted polyimide litholoops from Molecular Dimensions of widths appropriate for the crystal to be looped were immobilised on a magnetic base using superglue. In order to be compatible with future, automated experiments at the Diamond Synchrotron Light Source (DSLS) the base used was bar-coded and of a specific Crystal Cap HT design, from Hampton Research, for the sample-loading robot used at Diamond.

Under a microscope, crystals were gently manipulated into a loop from their growth-drop and smoothly transferred to a 2 μ l drop of cryo-buffer, described in Section 3.7.5, and left to soak for 1-2 minutes. The angle and direction of the looping motion was dependant on the crystal form to be looped, as thinner, longer crystals with potential shatter-points needed different angles of leverage than fatter, evenly dimensioned crystals. Using the same loop, or a sterile loop of the same size if multiple crystals or crystal debris were originally transferred, the individual crystal was then transferred to the head of the goniometer quickly to avoid evaporation of water from the loop to the air, but also as carefully as possible as to avoid physical damage to the sample via collisions or vibrations. For this reason also, the goniometer magnet was not engaged until the loop was in position on the head as to avoid the sudden pull of magnetism as the base entered the goniometers proximity.

This now placed the isolated crystal in the nitrogen stream and attached to a goniometer that allowed for orientation changes to the crystal while it was aligned in the X-ray beam. For long-term saving after in-house testing, the crystal could be carefully removed from the X-ray machine surrounded by a cryo-tube full of liquid nitrogen and stored in a nitrogen-filled Dewar. From this storage it was then possible to unload the crystal in the loop and base into a crystal transport 'puck' using a magnet-tipped rod under a pool of nitrogen. This could now safely be transported to the synchrotron in a liquid nitrogen-filled Dewar

for automated loading onto a beam-line using a robotic arm for more advanced data collection.

3.7.5 Cryo-protection of Crystals

To reduce radiation damage to crystals due to exposure to prolonged high-intensity radiation, crystals were cooled in a nitrogen stream during data-collection. In order to prevent crystallisation of water ice when data are collected at temperatures below the freezing point of water, typically 100 Kelvin, the solution in which the crystals were looped onto the X-ray machine needed to be supplemented with a cryo-protectant (Garman, 2003). Compounds such as glycerol, ethylene glycol and MPD were added to the mother-liquor in gradients of 5% from 0-25%. The ability of the solution to act as a cryo-buffer was assessed by the looping of cryo-buffer followed by exposure to high-intensity X-rays whilst under a nitrogen stream. The cryo-buffer used was the lowest % cryo-agent that prevented formation of ice crystals (detected by diffraction) without compromising the physical integrity of the crystal.

3.7.6 Metal-Soaking of Crystals

In order to experimentally calculate phases to break the phase ambiguity, measurements of anomalous scattering from the crystal were required. Where it was not possible to use selenomethionine (Section 3.5.3), in some cases attempts to generate this anomalous signal were carried out by the soaking in of heavy-atom containing compounds. In general, preformed crystals were gently looped from the mother-drop and transferred to a drop containing cryo-buffer + 10 mM metal-compound where they were left to soak for 30-40 minutes. Crystals were then looped onto an X-ray machine for testing, and anomalous data were collected at a synchrotron light-source. The individual soaks and resulting diffraction experiments are described later on in this work.

3.7.7 Data Collection

Crystals were tested for protein diffraction, resolution limits and if possible the calculation of cell-parameters using an in-house Mar345dtb goniometer and detector set-up linked to a computer running the Mar345dtb 6.1.0 program and iMOSFLM (Battye et al, 2011). When suitable crystals were identified for each specific dataset desired, they were transported in liquid nitrogen to the DSLS in Oxford. At the synchrotron, X-rays of much higher intensity were used to collect the most complete, highest resolution datasets possible for native data, and beam lines were attenuated to vary the wavelength of the radiation used for the collection of anomalous data for MAD phasing. Diffraction images were collected on an ADSC Q315r CCD detector and saved to a portable hard-drive for future processing and analysis.

3.7.8 Data Processing

Data images were indexed and integrated as complete datasets in the program iMOSFLM (Battye et al, 2011) and output as .mtz files. This program was also used for the calculation of unit cell parameters and symmetry space-group from test images and the design of a strategy to optimise data-collection in advance of dataset-length exposure to radiation.

Datasets were then transferred to the CCP4 suite of programs (Winn et al, 2011). Datasets of experimental intensities were scaled using SCALA (Evans, 2006) and converted to amplitudes by running Ctruncate (Padilla & Yeates, 2003), and were output in Scalepack format as unmerged, scaled .sca files. For MAD phase solution, these data were then input into ShelX C, D and E (Sheldrick, 2010), run through HKL2MAP (Pape and Schneider, 2004), for the preparation of data from .sca to .hkl files for phasing, heavy atom location and then phasing and electron density modification respectively.

Using the CCP4i F2mtz import program, each of the two output ShelXE enantiomorph .phs files of phased electron density were converted back to .mtz files. They were then subjected to automatic model-building using ARP/wARP

(Langer et al, 2008), which when supplemented with the expected amino acid sequence as a .pir file attempted to fit as many residues into the calculated electron density as possible and output the result as a .mtz file of density alongside a .pdb file representing the model built. The resulting density map and output model were viewed in the molecular graphics program Coot (Emsley & Cowtan, 2004). The enantiomorph chosen was the solution that calculated an electron map in which protein features could be most reasonably deduced, such as α -helices, β -sheets and solvent boundaries.

After the auto-fitted model was viewed in the electron density in Coot, any residues left un-built were fitted in the electron density according to the predicted sequence to complete as much of the model as the calculated electron density allowed. In addition, each residue either built by ARP/wARP or manually was further analysed to correct residue orientation utilising chemical knowledge of charged residues and bond lengths for correct hydrogen bonding or ionic interactions. Water molecules were also built into 'blobs' of density where chemically possible. For every significant batch of changes made to the model, a cycle of refinement in the Refmac program (Murshudov et al, 1997) was used to refine the density to the model, using a FreeR dataset of 5% of the data to reduce data-bias, producing new density .mtz and model .pdb files.

Once the most complete model was built in Coot, the structure coordinates in the .pdb file were structurally analysed using the MolProbity server (Davis et al, 2007) and ProCheck (Laskowski et al, 1993) in CCP4i. The outputs from these analytical programs were used to assess the Ramachandran distribution (Ramachandran et al, 1963) of residues and any potential errors in the structure such as non-biological conformations of residues and impossible chemistry such as irregular B-factors. Any flagged residues were then fixed in Coot and then the model was re-analysed until deemed as correct. With a complete model built into the calculated electron density, the final .pdb file was loaded into the molecular graphics program PyMol (Schrödinger, LLC) for the graphical representation of the completed protein molecule in the form of a cartoon for illustrative purposes of key features and the overall 3D structure.

3.8 Assays

3.8.1 β -lactamase Inhibition Assays

Assays were carried out for this thesis to identify the effect of one of the proteins of interest on the hydrolysis of β -lactam-containing antibiotics such as penicillin by the general β -lactamase TEM-1 (available from Invitrogen). Inhibition analysis was carried out by spectrophotometrically measuring the absorption change of 390nm to 486nm of the penicillin analogue Nitrocefin, product and protocol both available from Calbiochem, upon β -lactam ring hydrolysis by TEM-1. Measurements were carried out using an Agilent Cary Eclipse spectrophotometer linked to a PC for recording of results as a Microsoft Excel file.

100 μ l of a 0.5 mg/ml Nitrocefin, 5% DMSO, 100 mM PBS pH7.0 stock was added to 900 μ l of 50 mM Tris pH 7.0 in a sterile, 1 ml glass cuvette to create the substrate solution for the reaction. The entire preparation procedure and assay was carried out with as little exposure to light as possible to minimise any effects caused by the photoreactive nature of Nitrocefin. With the Cary machine's darkened measurement chamber ready for recording, the substrate-solution was supplemented with a 1 μ l premix of 1:1, 1:2, 1:10 or 1:4000 mix of 5 ng TEM-1: χ ng potential inhibitor in 50 mM Tris pH 7.0. Control experiments were carried out with just 5 ng TEM-1 and without TEM-1. Immediately upon the addition of the protein premix and a quick, vigorous inversion to mix, measurements were taken every 30 seconds over a period of 90 minutes and recorded as an Excel file, for statistical interpretation and representation as line graphs.

3.8.2 Affi-Gel15 Pull-Down Assays

Affi-Gel15 beads, available from BioRad, were used to covalently bind free primary amine-groups of acidic bait proteins. The following method was applied to both the bait protein under investigation and an unrelated control protein. 50 μ l of the cross-linked agarose beads were transferred to a 1.5 ml Eppendorf by using a 1000 μ l pipette tip with the end 1 mm cut off to reduce beads sticking to

the tip. The beads were washed from their isopropanol storage-solution by allowing beads to settle then gently removing isopropanol by pipetting until the surface of the beads was nearly exposed. The tube was then topped up with 1.5 ml of sterile, deionised water, inverted to mix and then the beads were allowed to settle again. Beads were gently pelleted at the bottom of tubes if stuck on the sides by centrifugation at 500 rpm for 30 seconds. This was repeated two further times, and then an equilibrating wash-step was carried out for the final time using 50 mM HEPES pH7.5. It was important to not use Tris as the reaction buffer, even if the bait proteins were usually purified into this buffer, as the primary amine in the Tris molecule would otherwise swamp the now active N-hydroxysuccinimide esters coating the beads.

3 mg of bait protein in the highest concentration possible was prepared in 50 mM HEPES pH8.0 and added to the equilibrated beads. Over a period of 2 hours the beads and protein were left to covalently bind in a 4°C cabinet on a mixing wheel. The beads were then gently pelleted and the solution removed and assayed for protein concentration using the Bradford method. The beads were then washed with 50 mM HEPES pH7.5, 0.1 M NaCl, and then allowed to settle before the solution above was removed. Then 1 ml 50 mM HEPES pH7.5, 2 M NaCl was applied, followed by 3x1 ml 50 mM HEPES pH7.5 washes until the removed solution no longer contained protein. At this point the mass of protein removed was subtracted from the mass originally added to the beads to calculate the mass of protein bound by the beads. 100 µl of 1 M ethanolamine was added to block any remaining binding sites and left to incubate for 1 hour at 4°C on a wheel. Finally, the ethanolamine was removed and the beads were left in 50 mM HEPES pH7.5 and stored at 4°C until further use.

2 g of cell-paste to be used for the pull-down was re-suspended in 3 volumes of 50 mM HEPES pH7.5 and split into 2x3 ml aliquots and each was sonicated 5x7 seconds on ice at 16 microns, and cell-debris was pelleted by centrifugation at 70000xg for 10 minutes at 8°C. The 50 µl of beads with bound bait protein were split into 3 by re-suspending into 850 µl and with a 1000 µl pipette tip cut off by 1 mm, 300 µl of the bead suspension was put into 3 fresh 1.5 ml Eppendorfs and allowed to settle. The excess solution above was removed and the beads re-

suspended in 1.5 ml of the concentrated cell-free extract, with the addition of final concentrations of 0, 0.1 or 0.2 M NaCl to investigate affinity–strength of the pull-down. The pull-down interactions were allowed to happen for 2 hours at 4°C on a wheel before being washed 5x1.5 ml in the corresponding concentration of NaCl and 50 mM HEPES pH7.5.

Each set of beads were re-suspended in 1 ml of buffer and split further into 2x0.5 ml, spun for 500 rpm for 30 seconds to gather at the bottom of the tube and had the buffer removed to the surface of the beads. One set of beads were eluted once with the addition of 20 µl 2 M NaCl to approximately 1.5 M NaCl when diluted by the beads, with gentle mixing for 1 minute, and then after the removal of the first eluate this was repeated to elute at 2 M NaCl. SDS-PAGE samples were made by addition of 2 µl 1 M DTT and 5 µl 4xLDS loading buffer to 10 µl of each eluate. The other portion was boiled in 10 µl water, 2 µl 1M DTT and 5 µl 4xLDS loading buffer to elute based upon LDS denaturation. Samples of each different NaCl molarity pull-down, eluted by both 1.5 M NaCl or 2 M NaCl or LDS, were ran on a 13.5% SDS-PAGE gel against the corresponding experiments with the control protein and the beginning bait, control and cell-extract protein samples used. After staining the gel with colloidal Coomassie and then de-staining the gel, the lanes were compared to discover any bands that appeared only in the bait protein experiments, which were then identified using mass spectrometry of trypsin-digested fragments.

3.9 Localisation and Antigenic Studies

3.9.1 Fractionation of *Burkholderia cenocepacia*

A 500 ml overnight culture of *B. cenocepacia* was grown following the protocol for *E. coli* culturing described in Section 3.5.3. LB media was supplemented with a final concentration of 100 µg/ml ampicillin. Cells were harvested by centrifugation as described in Section 3.5.3. The sub-sections that follow document the chronological process of acquiring cell fractions from a cell pellet. The methods described were adapted from a periplasmic and outer membrane preparation protocol for the Gram-negative bacterium *Campylobacter jejuni*,

provided by Professor David Kelly (University of Sheffield, Dept. of MBB). All work was carried out in a BioMat2 Class II Microbiology Safety Cabinet with laminar airflow, sealable protective screen and UV sterilisation.

3.9.1a Secreted Fraction: The media supernatant left after centrifugal pelleting of cells contained secreted proteins. 100 ml of the media was taken from the top of the supernatant prior to decantation from the pellet to avoid a contaminating re-suspension of cells. 50 ml of this was treated with a final concentration of ~3.5 M ammonium sulphate by the gradual addition of 23 g of solid ammonium sulphate, to precipitate proteins into a concentrated pellet. BPSL1204, and thus by inference BCAL2351, precipitated in ≥ 1.6 M ammonium sulphate (see Section 3.6.3). The remaining 50 ml was used as a 'native' secreted fraction, to be loaded on analytical SDS-PAGEs post-boiling in LDS to concentrate. Samples were stored at -20°C .

3.9.1b Periplasmic Fraction: The cell pellet was gently re-suspended in 20 ml STE buffer (20% sucrose, 30 mM Tris pH8.0, 1 mM EDTA) and shaken at room temperature for 30 minutes, following which cells were harvested by centrifugation in a Sigma 11133 rotor at 5000 rpm for 10 minutes at room temperature. The supernatant was discarded and osmotic shock was applied by re-suspension in 10 ml of ice-cold 10 mM Tris pH 8.0, followed by incubation with gentle shaking at 4°C for 2 hours. The suspension was then centrifuged at 11000 rpm in a JA 25.50 rotor for 25 minutes at 4°C , and the periplasm-containing supernatant was collected and stored as 0.5 ml aliquots at -20°C .

3.9.1c Cytoplasmic Fraction: The post-osmotic-shock pellet was re-suspended in 20 ml of 10 mM HEPES pH 7.4, and sonicated on ice for 6 x 15 seconds at 16 microns, using a MSE SoniPrep 150 with a sealable screen, to destroy intact cell membranes. The sonicate was then centrifuged in a JA25.50 rotor for 20 minutes at 15000 rpm at 4°C to pellet unbroken cells and debris. The supernatant was carefully removed and transferred to a pre-chilled, polyallomer Beckman ultra-centrifuge tube on ice, and the pellet was discarded. Against a same-tube water-balance (to within 0.05g), the supernatant was ultra-centrifuged in a pre-chilled

Beckman 70.1 Ti angle-rotor at 100,000xg for 1 hr at 4°C. The supernatant was removed as the purified cytoplasmic fraction and stored as 0.5 ml aliquots at -20°C.

3.9.1d Inner Membrane Fraction: The ultra-centrifuged pellet had its surface washed by the gentle addition of a full tube of 10 mM HEPES pH7.4 that was gently poured off. This was repeated twice. 2 ml of 10 mM HEPES pH7.4 was then added, and the membrane was fully suspended by gentle agitation with a glass rod. To selectively solubilise the inner membrane (Buchanan, 1999), 2 ml of 2% sodium N-lauroyl sarcosinate (Sarkosyl) dissolved in 10 mM HEPES pH 7.4 was mixed with the membrane-solution thoroughly (but not so that foam formed), and was then incubated at 37°C for 30 minutes. After centrifugation in a JA25.50 rotor for 30 minutes at 20,000 rpm at 15C, the supernatant was carefully removed as the soluble inner membrane fraction and was stored as 0.5 ml aliquots at -20°C.

3.9.1e Outer Membrane Fraction: The outer membrane pellet was washed twice by the gentle addition and removal of a full tube of 10 mM HEPES pH 7.4. The outer membrane pellet was then suspended in 0.5 ml of 10 mM HEPES pH7.4 and stored at -20°C.

3.9.2 Buffers used in Immuno-blotting

For the Western blotting (Burnette, 1981) and immuno-blotting methods several buffers were required to be prepared, which are summarised in this section:

Western Transfer buffer: 25mM Tris (do not alter pH), 192 mM glycine and 10% methanol. Stored at 4°C.

X10 Phosphate-buffered Saline (PBS) pH 7.4: 80 g NaCl, 2.0 g KCl, 14.4 g Na₂HPO₄ and 2.4 g KH₂PO₄ dissolved in 850 ml deionised water, pH adjusted to pH 7.4 and then water added to 1 L.

M-T-PBS: 5% non-fat milk powder, 0.1% Tween-20 in 1xPBS pH 7.4. Stored at 4°C.

T-PBS: 0.1% Tween-20 in 1xPBS pH 7.4.

PBS: 1xPBS pH 7.4.

3.9.3 Western Blotting

A 13.5% SDS-PAGE gel was created and samples were run through it as described in Section 3.2.3. A pre-stained protein marker was used, as this would transfer to the blotted nitrocellulose membrane to aid in molecular weight determination of any bands seen. An Invitrogen XCell Surelock tank with an XCell II Blot Module was used for all Western blots, kept cold by surrounding with ice in a polystyrene icebox. Prior to assembly of the blot, seven 10.5x9.5 cm sponge pads were soaked and left submerged in cold Transfer buffer, ensuring that all air bubbles were pressed out. A 9x9 cm nitrocellulose membrane, available from Invitrogen, and two 9x9 cm sheets of filter paper were also soaked in cold Transfer buffer. The negative-electrode compartment of the XCell II module was laid down first, and then stacked horizontally with 3 buffer-soaked sponge pads. Using gloved hands only, a piece of filter paper was then placed on the sponges, on top of which the SDS-PAGE gel was gently lain, ensuring no tearing or creasing occurred. The nitrocellulose membrane was then placed on top of the gel, followed by the second piece of filter paper and 4 buffer-soaked sponge pads. The stack was created ensuring that no bubbles were caught within any layer, and ordered so that the current ran through the stack to the positive-electrode, transferring the SDS-coated negatively charged proteins from the gel to the nitrocellulose membrane.

The positive-electrode lid of the module was then used to gently compress the Western stack between the two electrode panels, sealing against the rubber gasket. The whole module was then placed into the XCell Surelock tank and locked into place by the plastic lever-arch mechanism. Cold transfer buffer was

used to fill the module to approximately 1 cm above the stack, to prevent drying out whilst creating the minimum electrical conductivity, and thus the lowest heat possible. The buffer-compartment was filled with cold Transfer buffer in order to further dissipate heat, and account for any module-leakage. Blots were run using cold equipment and buffers, as excessive heat may have damaged the module electrodes and melted or distorted the gel and nitrocellulose. The wired-lid was attached to the tank so that electrodes were correctly orientated, and then the blot was run at 65 V for 120 minutes.

3.9.4 Immuno-blotting

Using gloved hands, the Western-blotted nitrocellulose membrane was transferred to a clean, plastic gel-box containing 100 ml of M-T-PBS, and was left sealed on a gently rocking platform overnight at 4°C, to block the regions of the membrane not occupied with electro-blotted proteins. After the blocking incubation, the membrane was rinsed in T-PBS, and then was transferred into 50 ml of M-T-PBS containing an experimentally determined dilution of primary antibody. This primary binding reaction was allowed to incubate for 60 minutes at room temperature on a gently rocking platform. No primary antibody was added to control membranes.

After incubation with the primary antibody, the membrane was washed with 3x50 ml of T-PBS, each for 5 minutes, to remove any residual unbound primary antibody. Next, the membrane was incubated with 50 ml M-T-PBS containing an experimentally determined dilution of secondary reporter antibody, on a gently rocking platform at room temperature for 30 minutes to allow for high-affinity secondary binding to the primary antibody. The membrane was then again washed with 3x50 ml of T-PBS to remove any unbound secondary reporter antibody, before being rinsed in PBS to remove residual Tween-20 detergent.

3.9.5 Chemiluminescence Assays

To facilitate safe and measurable exposure of the membrane to photoreactive chemicals and X-ray film, the chemiluminescent reaction was carried out

contained within plastic wrap. With gloved hands, a piece of plastic wrap several times larger than the membrane was laid totally flat onto a flat, clean lab surface. A solution of Enhanced Chemiluminescence (ECL) reagent, from Pierce ThermoScientific, was pipetted onto the centre of the plastic wrap, forming a dome of solution. The horseradish peroxidase (HRP, the reporter used on the secondary antibody) ECL substrate was made by mixing a 1:1 solution of the 2 components of the reagent, with care taken due to the highly toxic nature of the solution in the concentrations used. For a 9x9 cm membrane, a 500 µl final volume was ample for total coverage. The membrane was smoothly laid down on top of the ECL solution, blotted-side down as indicated by the presence of the pre-stained protein markers, and using a glass rod to gently roll back-and-forth the membrane was coated bubble-free in ECL reagent. The plastic wrap border was carefully folded over the back of the membrane and sealed with tape. A marker was drawn on the tape to indicate the orientation of the membrane in the plastic wrap.

The ECL reagent was highly reactive within the first 30 minutes of mixing; therefore the sealed membrane was immediately taken to a photographic dark room. When in the dark room and when not in use for exposure, the membrane was kept in a clean, opaque container to prevent bleaching of unprotected X-ray film. When lit by only red-filtered safety lights, a piece of CL-XPosure X-ray film from Pierce ThermoScientific was removed from its protective sleeve and placed onto a black-foam positioning tablet. The membrane was then placed upon the film and its edges were marked for later reference. Exposure was allowed to occur for a range of time periods from 10 seconds to 5 minutes in order to achieve the optimal exposure time for the best signal:noise. Film was then developed using a Konica SRX-101a Automatic X-ray film-processing unit. Film could then be aligned with the membrane edges to compare the molecular weights of the bands recorded on the developed film against the pre-stained protein markers on the membrane.

The immuno-blotting and chemiluminescent assay were then optimised by the alteration of primary antibody and secondary antibody dilutions used and then

further improved by variation in the length of X-ray film exposure time, until the most interpretable signal:noise was achieved (Kricka, 1991).

3.9.6 Preparation of Slides for Fluorescence Microscopy

A 10 ml overnight culture of cells was grown in LB media, and then the following morning 100 µl was used to 2% inoculate a fresh 4.9 ml of LB that was allowed to culture at 37°C and 200 rpm until an OD_{600nm} of 0.6 was reached. This was with the aim of collecting a sample of cells within which existed cells in various stages of the cell cycle. 1 ml of this culture was taken into a sterile 1.5 ml microfuge tube and pelleted by centrifugation in an MSE Micro-Centaur rotor for 5 minutes at 13000 rpm at room temperature. The supernatant was discarded and the pellet was re-suspended in 0.5 ml of PBS, and then mixed with 0.5 ml of fresh fixation solution (Hopwood, 1972). Fixation solution was made by the addition of 420 µl of 15% formaldehyde and 500 µl of 25% glutaraldehyde to 2.08 ml PBS. The cell fixation reaction was allowed to continue for 30 minutes at room temperature on a rotary wheel.

Once fixed, the cells were collected by centrifugation at 13000 rpm at room temperature in an MSE Micro-Centaur rotor, and the pellet was gently washed twice with 1.5 ml of PBS. At this stage, fixed cells could be stored at -20°C if required. The fixed-cell pellet was re-suspended in 0.5 ml of PBS and 10 µl of the re-suspension was pipetted onto the centre of a poly-L-lysine-coated slide (Huang et al, 1983), fixed-cells were allowed to bind to the slide and then were air dried whilst protected by a cover to prevent environmental contamination. Once dry the slide was rehydrated with 50 µl PBS and unbound poly-L-lysine was blocked with 2% bovine serum albumin (BSA) in PBS for 15 minutes at room temperature. The fixed- and slide-immobilised cells were then incubated for 2 hours with an experimentally determined dilution of primary antibody in 2% BSA in PBS. No primary antibody was added to control slides.

Cells were washed 8 times by the addition of 0.5ml PBS followed by air aspiration. Then cells were incubated with an experimentally determined

dilution of secondary antibody FITC-conjugate in 2% w/v BSA in PBS for 30 minutes. From this stage onwards, slides were manipulated in the dark to prevent bleaching of the FITC fluorophore (Lichtman & Conchello, 2005). Cells were again washed 8 times with 0.5 ml PBS followed by air aspiration. The slide was then treated with 5 μ l of Slow Fade Gold AntiFade reagent, to suppress photobleaching and improve fluorescent signal. Finally, slides were sealed with DPX mountant media.

3.9.7 Fluorescence Microscopy Image Collection and Processing

Slides were viewed using a DeltaVision deconvolution fluorescence microscope and confocal Z-stack images were acquired and analysed using SoftWoRx 3.5.0 software, from Applied Precision Solutions. Fluorescence of FITC was achieved by setting a laser excitation wavelength of 490nm. Quick Projection image processing, in which images are stacked upon one another to create a single focussed image, was applied when only an overview of the data collected was required due to limited or lack of fluorescence in the experiment.

Chapter 4: Target-Selection, Over-expression and Purification of BPSL1204

4.1 Bioinformatical Selection of BPSL1204

4.1.1 Analysis of Hydropathy Plots

All the 5521 potential translated protein sequences from the proteome of *B. pseudomallei* were plotted as hydropathy plots by Prof. Peter Artymiuk as described in section 3.3 and then each was individually examined by eye. Each plot and related sequence were then categorised as either having a significant N-terminal or C-terminal peak of hydrophobicity perhaps representative of a membrane-anchoring sequence, or as either having no significant hydrophobicity or multiple hydrophobic peaks representative of a multi-pass membrane protein. This left 525 proteins from the 5521 proteins of both chromosomes that appeared to potentially have hydrophobic N- or C-terminal tails.

The list of proteins with hydrophobic termini was further reduced by only accepting proteins with a terminal peak greater than 2.0 on the Kyte-Doolittle scale (Kyte & Doolittle, 1982), leaving a list of 205 for further analysis and target selection (Figure 4.1).

4.1.2 Selection of Potentially Novel Folds

The amino acid sequences of the 205 proteins were batch-input in groups of around 30 sequences into the BLAST server from NCBI (Altschul et al, 1990), and aligned against the PDB sequence database to discover if any homologs with known structure and function were already known. As the aim was to discover the structures of novel, membrane-anchored proteins any significantly positive hits found during this search were rejected. This process of positive discrimination led to a shortening of the list to 102 proteins with a profile of an E-value > 0.1 and less than 50% ID when compared to the PDB sequences.

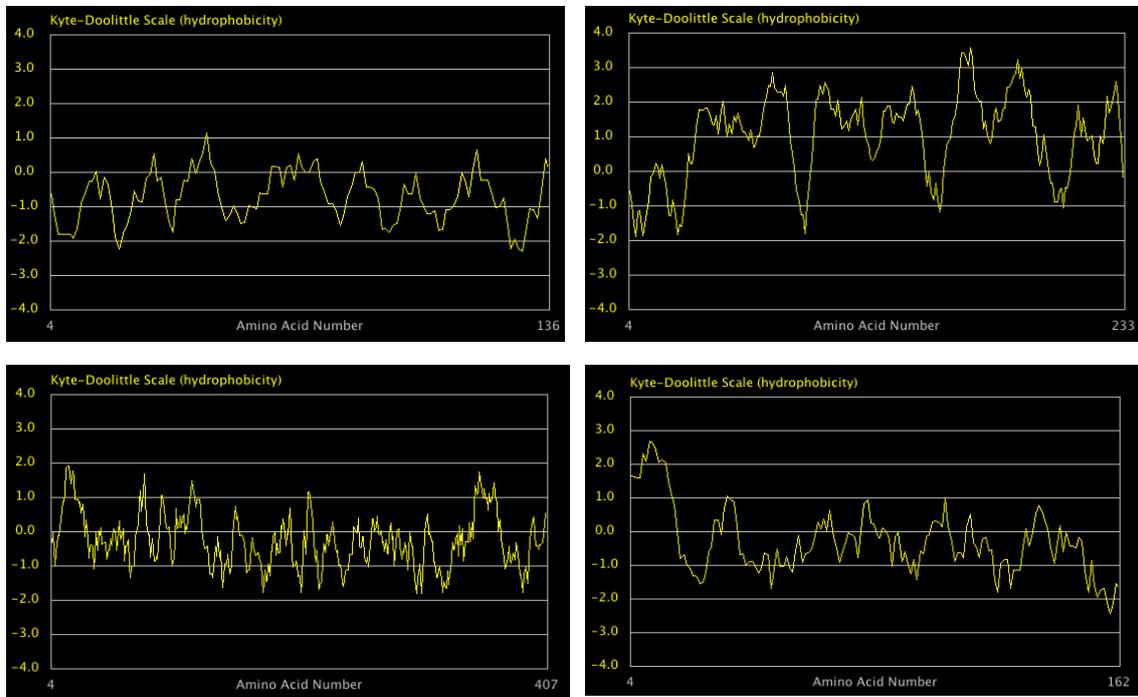


Figure 4.1: Analysis of hydropathy plots of proteins from *B. pseudomallei*, in order to identify targets with potential hydrophobic membrane-insertion sequences. Hydropathy was plotted using the Kyte-Doolittle scale (Kyte & Doolittle, 1982) and a window size of 9. *Top left:* BPSS1839, rejected as a soluble protein. *Top right:* BPSL1509, rejected as multi-pass membrane protein. *Bottom left:* BPSS0580, rejected as N-terminally hydrophobic <2.0 . *Bottom right:* BPSL1204, selected as N-terminally hydrophobic >2.0 . Plots created using the Molecular Toolkit from the University of Colorado (www.vivo.colostate.edu/molkit).

This still extensive list was therefore shortened further by demanding potential targets had a lowest E-value of >1.0 . Therefore any hits generated by the BLAST search could be considered as random and we could be confident that no structure with homologous sequence existed. This left a short-list of 43 targets to pass onto the next stage of selection.

4.1.3 Further Exclusion of Targets from the Short-List

Excluding targets based upon the following criteria further reduced the list of proteins that had no positive PDB result and a highly hydrophobic N- or C-terminus:

- Discard proteins with any region of potential disorder >0.5 probability, other than perhaps the potential hydrophobic membrane-anchor region, as predicted by the RONN server (Yang et al, 2005) (www.strubi.ox.ac.uk/RONN). This was done to increase chances of ordered crystal formation in the future.
- Compare the short-list to the Target Database (Chen et al, 2004) (<http://targetdb-dev.rutgers.edu/>) of targets currently under investigation by structural genomics consortia, and discard any that are already being investigated.
- Utilise the Psipred GenThreader tool (<http://bioinf.cs.ucl.ac.uk/psipred>) (Jones, 1999a) to probe for structural predictions based upon common sequence motifs for a specific domain fold. Reject any targets that produce a fold-recognition with any degree of certainty to further increase the chances of working with a novel protein fold.
- Use the BLAST server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al, 1990) to search for homologs of the *B. pseudomallei* short-listed targets in the paraphyletic, non-virulent *B. thailandensis*, and only select those with an E-value $<1e^{-15}$. This increased the likeliness of selecting a target protein that was safe to work with in the standard Class I lab environment.

- Reject any protein that does not contain 3 or more methionine residues (not including the N-terminal residue, which is often disordered or post-translationally removed). As the desired target was to be of a potentially new fold with no known homologous structure, the preferred method of solving the phase-problem would be to use MAD data collected from a crystal of a selenomethionine derivative of the target. In order for incorporation of selenomethionine and the generation of significant anomalous signal for phasing, multiple methionines were required to be in the chosen amino acid sequence.

4.1.4 Final Selection of BPSL1204 and Truncation

The following table describes the short-list of potential targets, narrowed down from a predicted proteome of 5521 proteins to 3 potentially novel membrane-anchored targets:

Gene ID	MW (kDa)	Membrane-anchor truncation MW (kDa)	E-value against <i>B. thailandensis</i>	No of Methionines
BPSL1204	19189	17197	$2e^{-86}$	4
BPSL1927	26379	24420	$2e^{-126}$	4
BPSL2756	19466	15904	$1e^{-74}$	4

Attempts to clone and over-express soluble protein of all three truncated proteins led to success only with BPSL1204-MA (i.e. BPSL1204 with its Membrane Anchor removed). The cloning and over-expression processes applied to BPSL1204 (described in Sections 4.2.1-4.3.2) were applied to BPSL1927 and BPSL2756 but did not result in soluble protein.

4.2 Cloning of BPSL1204-MA

4.2.1 Primer Design and PCR of Truncated BPSL1204

Primers BPSL1204-F (forward) and BPSL1204-R (reverse) were designed against the BPSL1204 gene sequence for amplification directly from purified *B. pseudomallei* K9263 genomic DNA. The primers have been described in Section

3.4.1. These primers omitted from the amplified fragment the initial 60 5' base pairs, resulting in a translated protein product truncated by the first 20 N-terminal amino acid residues. The decision to truncate by 60 nucleotides was based upon the comparison of the gene-product's hydropathy plot aligned with its sequence. At the point where the hydrophobicity of the amino acid sequence approached 0 on the Kyte-Doolittle scale was the point of truncation (Figure 4.2). The primers incorporated only the fragment of BPSL1204-MA to be over-expressed with no overhanging bases.

Primers were synthesised by Dr Arthur Moir of the University of Sheffield, and kept at -20°C as a 100 µM master stock from which 5 µl were diluted with 95 µl of sterile, deionised water to make a 5 µM working stock. The PCR method as described in section 3.4.1 was applied, with the only variation being the use of 55°C as an annealing temperature which was required for optimum BPSL1204-MA amplification. The purity of the PCR product was analysed by 1% agarose gel electrophoresis (Figure 4.3) and was discovered to be pure, and in order for efficient ligation into the pETBlue-1 vector the insert sample amplified by the PCR reaction was purified from the PCR buffer into 40 µl sterile, deionised water using the Qiagen PCR Purification kit.

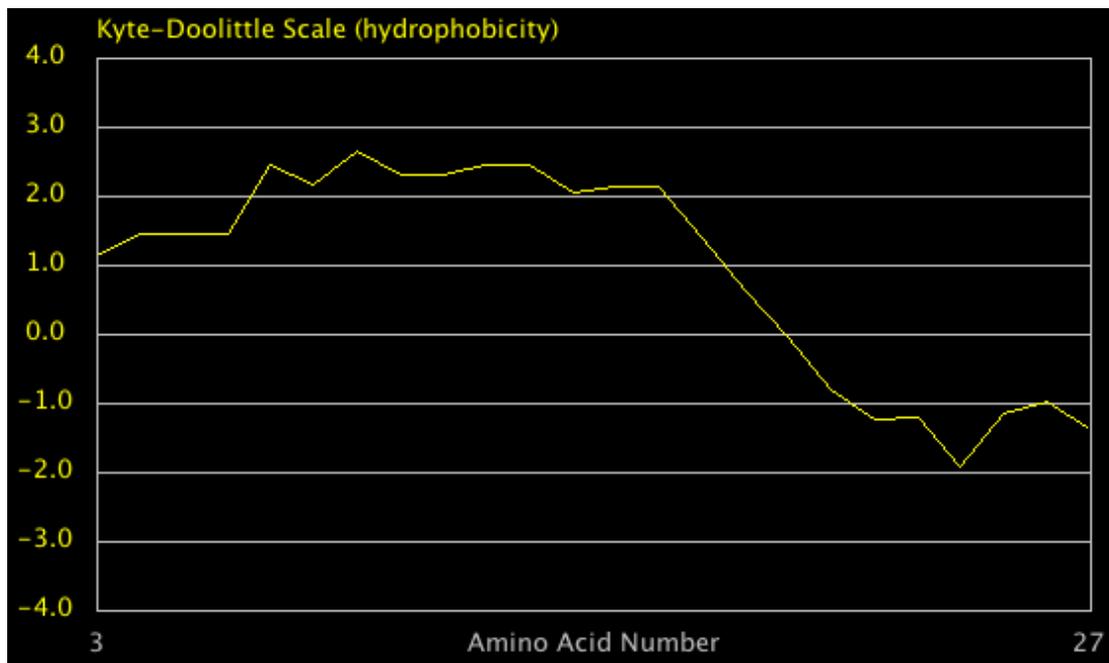
4.2.2 Blunt-End Ligation into pETBlue-1

Following the AccepTor Vector kit protocol, the following blunt-end ligation reaction was set up:

3 µl 50 ng/µl BPSL1204-MA PCR product
1 µl 50 ng/µl linearised pETBlue-1 vector
5 µl Clonables mix (contains ligase and ligase-buffer)
1 µl sterile, deionised water

10 µl reaction

The pETBlue-1 vector was supplied from Novagen as a pre-linearised vector.



MFARLPVLFACAAALLAGCAQPWQQFHAGDD
 SSAIVARLGPPREIYDLPNGGKRLMWPTQPMG
 EVTTAADIDAHDRIVSVRQVLQPLEFYRAEIGK
 WTKTDVLVNFGRP VETSYFPLMKREVWTYRYL
 EDNVWYMLYSFYFDDDGILRLTQKTPDPLHDP
 DRRTL F

Figure 4.2: Amino acid sequence of BPSL1204 (*bottom*). When the amino acid sequence was compared to the hydropathy plot of the 30 N-terminal residues of BPSL1204 (*top*), it was seen that hydrophobicity drops off after approximately 20 residues. Plots were created using the Kyte-Doolittle scale (Kyte & Doolittle, 1982) and a window size of 9, using the Molecular Toolkit from the University of Colorado (www.vivo.colostate.edu/molkit).

Investigation of the N-terminal 20 residues revealed a high occurrence of hydrophobic residues (Phe/F, Val/V, Cys/C and Leu/L). Therefore BPSL1204-MA was a truncate formed by the deletion of 20 N-terminal residues/60 5' base pairs during cloning.

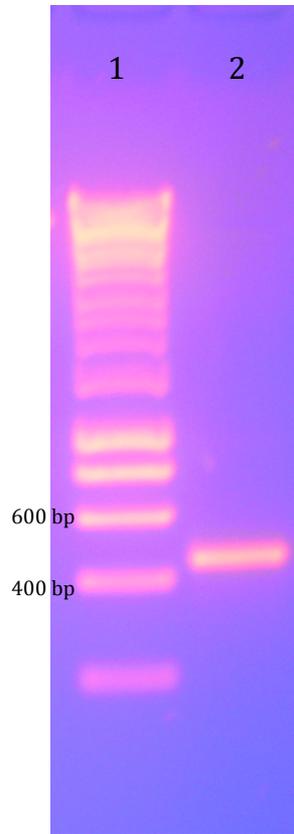


Figure 4.3: 1% agarose gel of pure BPSL1204-MA PCR product.
Lane 1: Hyperladder I (Bioline). Lane 2: BPSL1204-MA PCR (444bp)

The ligation reaction was allowed to proceed overnight at 4°C.

4.2.3 Sub-cloning into NovaBlue

The ligation of pETBlue-1:BPSL1204-MA was used to transform the sub-cloning *E. coli* NovaBlue, tetracycline-resistant cell-line. A pre-aliquoted 50 µl tube of cells was removed from storage at -80°C and allowed to gently thaw on ice for 5 minutes. 1 µl of the ligation was gently mixed into the thawed cells and left on ice for a further 10-15 minutes. The cells were then heat-shocked at 42°C for 30 seconds and then placed back on ice for 2 minutes. 250 µl of room temperature LB media was then added to a final volume of 300 µl and a 60 minute period of out-growth was carried out at 37°C and 200 rpm. Three LB agar plates containing 30 µg/ml tetracycline and 100 µg/ml carbenicillin were pre-soaked with 35 µl 50 mg/ml Xgal and 20 µl 100 mM IPTG and then were spread with the transformed culture using 200 µl, 100 µl and the excess liquid from the spreader from the 100 µl plate. This was in order to establish at least one plate with a usable colony count. Double antibiotic-resistance selection was used as pETBlue-1 carries the ampicillin/carbenicillin resistance gene. Plates were allowed to soak-in and then inverted overnight at 37°C for growth of colonies and further analysis.

4.2.4 Identification of Successful Clones

The Novablue cell-line was compatible with the pETBlue-1 vector to facilitate blue-white screening of colonies. The linearised pETBlue-1 was constructed so that the incorporated *lacZ* gene was broken, and upon re-circularisation a functional gene was restored to allow the expression of *lacZ* as induced by the 1mM IPTG incorporated into the plates. This allowed for functional β-galactosidase production, established by the production of blue colonies as the enzyme cleaved the Xgal in the plate media. However, when the BPSL1204-MA insert was blunt-end ligated into pre-linearised pETBlue-1 this prevented the production of active β-galactosidase, therefore any colonies containing vector with ligated-insert were selected as they presented as white colonies.

Due to the blunt-ended nature of ligation, ligated vectors could exist in 3 different forms leading to white colonies; BPSL1204-MA correctly orientated 5'-3' in the forwards reading-frame, BPSL1204-MA incorrectly orientated 3'-5' in the forwards reading-frame or not incorporated at all and the white colony was an unexplained false-positive actually containing only re-circularised pETBlue-1. Therefore each white colony was subjected to colony-PCR using the method described in section 3.4.1 but changed so that alongside the insert-specific 3' reverse primer the 5' primer used was 0.5 µl of 5 mM pETBlue-1-specific T7F primer. This primer would only bind the T7 promoter-region upstream of the insertion site so that only colonies with correctly orientated 5'-3' inserted pETBlue:BPSL1204-MA would generate PCR products. Colonies that generated positive PCR products were identified by 1% agarose gel electrophoresis (Figure 4.4).

Once a white, positive colony-PCR colony was identified, it was grown up overnight in 5 ml LB media + 30 µg/ml tetracycline + 100 µg/ml carbenicillin at 37°C and 250 rpm. The cloned-vector was then purified by Qiagen Miniprep to be sequenced using the T7F primer by Geneservice, to confirm again the presence of the insert in the correct orientation and to double-check that no mutations had occurred by comparison of the sequenced data to that of the original sequence taken from the BLAST server.

4.2.5 Cloning of Full-length BPSL1204

The restriction cloning protocol described in Section 3.4.2 was applied to full-length BPSL1204, with NcoI and BamHI restriction enzymes, using the primers in Section 3.4.1 and using the standard PCR reaction described in Section 3.2.1 with an annealing temperature of 55°C. The digested, full-length BPSL1204 gene was ligated into the arabinose-inducible pBAD/hisB vector and sub-cloned in DH5α, following the protocol in Section 3.4.3. Successful clones were identified by colony PCR using the original primers.

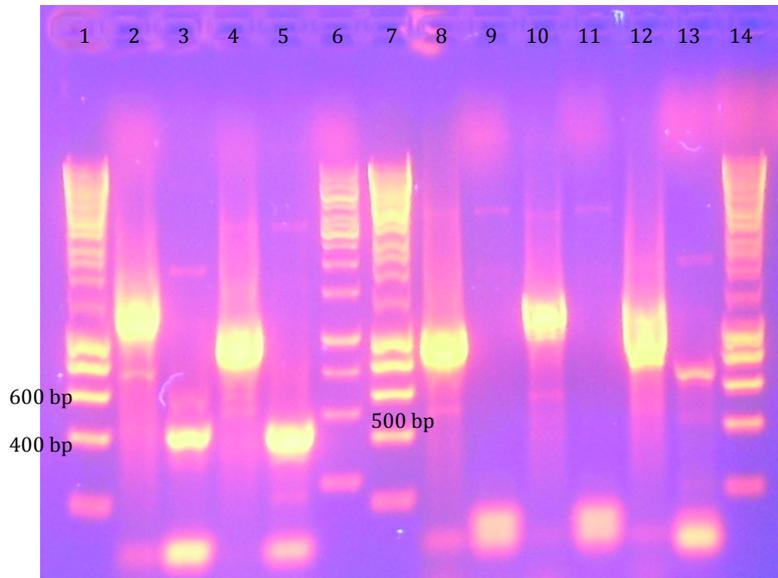


Figure 4.4: Colony-PCR agarose gel of white Novablue:pETBlue:BPSL1204 colonies.

Lane 1, 7 and 14: Hyperladder I (Bioline).

Lane 6: 1 Kb ladder (Fermentas).

Lane 2-5 and 8-13: white colonies 1-10.

Colonies 2, 4 and 10 gave expected bands of ~444 bp.

Colonies 1, 3 and 5-9 contained contaminating bands but no BPSL1204-MA.

4.3 Over-expression of BPSL1204-MA

4.3.1 Transformation into Tuner Cells

Correctly sequenced, cloned vector was transformed into competent, chloramphenicol-resistant Tuner(DE3)LacI cells for investigation of the over-expression of the cloned gene-product. A 20 μ l aliquot of competent cells in a 1.5 ml Eppendorf was removed from storage at -80°C and was gently thawed on ice for 5 minutes and the cells were re-suspended by gently flicking twice. 1 μ l of a 1:50 dilution of a 50 ng/ μ l stock of pETBlue:BPSL1204-MA was pipetted into the aliquot of cells and gently mixed on ice using the tip of the pipette. The Tuner cells and plasmid were allowed to incubate on ice for a further 10 minutes.

The cells were then heat-shocked for 30 seconds at 42°C in a water-bath to promote the uptake of DNA before being placed back in ice for another 2 minutes. 80 μ l of sterile LB media at room temperature was then added and the cells were left for an outgrowth step at 37°C and 250 rpm for 60 minutes. 2 plates of LB agar containing 100 $\mu\text{g}/\text{ml}$ carbenicillin and 34 $\mu\text{g}/\text{ml}$ chloramphenicol were prepared, onto one of which all 100 μ l of the transformation were spread, after which the excess liquid left on the spreader used was spread onto the second plate in order to guarantee at least one plate would supply a suitably diluted, usable colony count. Plates were left to soak-in, then inverted and left at 37°C for 15-18 hours to grow.

4.3.2 Over-expression and Solubility Tests

A colony from a successfully transformed Tuner:pETBlue:BPSL1204-MA plate was picked and used to inoculate 10 ml LB media, supplemented with 34 $\mu\text{g}/\text{ml}$ chloramphenicol and 100 $\mu\text{g}/\text{ml}$ carbenicillin which was left to culture at 37°C and 200 rpm overnight. Once uniformly re-suspended by inversion, this starter-culture was then used to 5% inoculate 7 x 4.75 ml of fresh LB in sterile 15 ml Falcon tubes, supplemented with the same antibiotics, by the addition of 250 μ l of starter-culture to each to a final volume of 5 ml. This was then allowed to culture at 37°C and 250 rpm until an $\text{OD}_{600\text{nm}}$ of 0.6-0.7 was reached in the

seventh tube. The other six tubes were left to culture untouched with the assumption of uniform growth throughout all seven tubes. Six 1 ml, 30% glycerol stocks were created using a portion of the excess starter-culture and were stored at -80°C for future use in repeating the culture.

When the correct optical density of cells was measured, three of the six tubes were induced by the addition of 5 µl 1 M IPTG to a final concentration of 1 mM, whereas the other three tubes served as non-induced controls. Three pairs consisting of one induced and one non-induced tube were created and each pair were left to over-express protein at 25°C, 30°C or 37°C for three hours at 250 rpm. Cells were then pelleted by centrifugation at 4500 rpm for 20 minutes at 10°C in a Sigma 3-16K 11133 rotor and the supernatant discarded by decanting and then removing the liquid left on the surface of the pellet by pipetting.

The pellets were subjected to Bugbuster and Benzonase (Nestle & Roberts, 1969) fractionation using the method described by Novagen, with the insoluble pellets re-suspended in 4% SDS using half the original Bugbuster volume, to increase the concentration of the samples. The protein concentrations of each fraction were calculated using the Bradford method described in Section 3.2.4 and samples containing 15 µg of protein were boiled with LDS loading-buffer and DTT, then loaded onto a 13.5% resolving SDS-PAGE gel and ran for 45-50 minutes at 200 V, before staining for several hours and then de-staining until bands were visible. The gel in Figure 4.5 shows that the optimum temperature for soluble over-expression of BPSL1204-MA over a three hour induction is 37°C when compared to the uninduced control, the other soluble variations in temperature and the corresponding insoluble fractions.

4.3.3 Large-scale Over-expression

Based upon the findings of the previous over-expression tests, large-scale over-expression was undertaken with a few variations in order to provide sufficient material for purification of BPSL1204-MA. A 100 ml starter-culture was set off

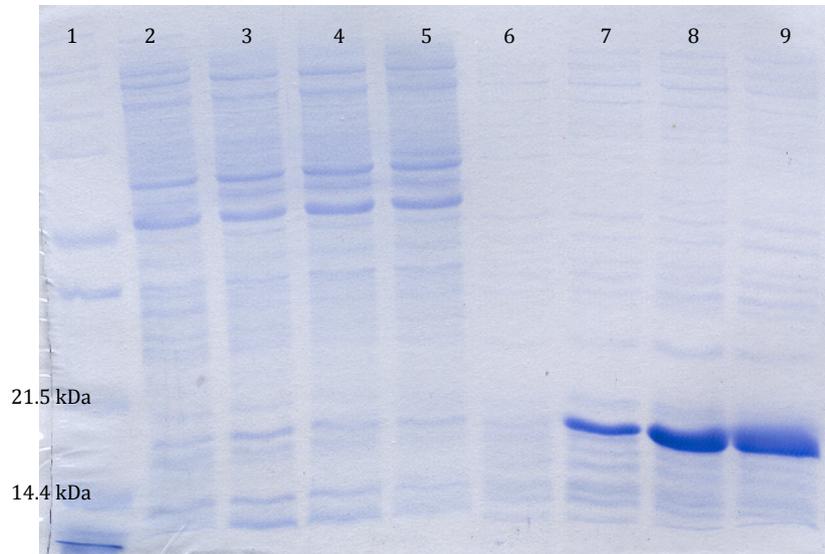


Figure 4.5: SDS-PAGE analysis of Tuner BPSL1204-MA over-expression.

- Lane 1: Mark 12 ladder (Invitrogen).
- Lane 2: Non-induced Insoluble
- Lane 3: 25°C Insoluble
- Lane 4: 30°C Insoluble
- Lane 5: 37°C Insoluble
- Lane 6: Non-induced Soluble
- Lane 7: 25°C Soluble
- Lane 8: 30°C Soluble
- Lane 9: 37°C Soluble

overnight at 37°C and 200 rpm the night prior to the scaled-up over-expression. Six 2 L flasks containing 490 ml LB were autoclaved, supplemented with antibiotics and warmed to approximately 37°C to reduce the lag-phase of bacterial growth. Each was then 2% inoculated with 10 ml of the starter-culture to a final volume of 500 ml and left to culture at 37°C and 250 rpm until an OD_{600nm} of 0.6 was reached. They were then induced by the addition of 0.5 ml of 1 M IPTG to a final concentration of 1 mM and left to over-express for four hours at 37°C and 250 rpm, after which cells were collected by centrifugation as described in Section 3.5.3 and stored as cell-paste in 50 ml Falcon tubes at -80°C until further use in purification experiments.

The lower inoculation percentage used was in the hope of generating more 'fresh' cells than in the original tests that would grow within an experimentally practical time-period. The use of a four-hour instead of three-hour induction was based upon the total lack of insoluble BPSL1204-MA in the three-hour tests, thus again aiming to achieve more soluble material per gram of cells within an experimentally practical time-period. A scraping of cells was removed from the cell-paste pellet and subjected to SDS-PAGE analysis post-Bugbuster fractionation, which showed that the scaling-up of the over-expression had no negative effect on the production of soluble BPSL1204-MA.

4.3.4 Over-expression of Full-length BPSL1204

After the transformation of purified pBAD/hisB:BPSL1204 into the TOP10 cell-line, variations in temperature, induction-time and the percentage of arabinose used for induction were applied to the over-expression protocol described in Section 3.5.3. As can be seen in Figure 4.6, full-length BPSL1204 was found to be insoluble. However this was considered a positive result in comparison to colleagues' success at producing soluble protein via similar methods applied to a pBAD/hisB construct (Robert Salmon and Professor David Kelly, University of Sheffield, Dept of MBB, personal communication). As full length BPSL1204 was predicted to consist of a targeting sequence followed by a glycerylcysteine lipoprotein moiety before the soluble BPSL1204-MA domains, the presence of

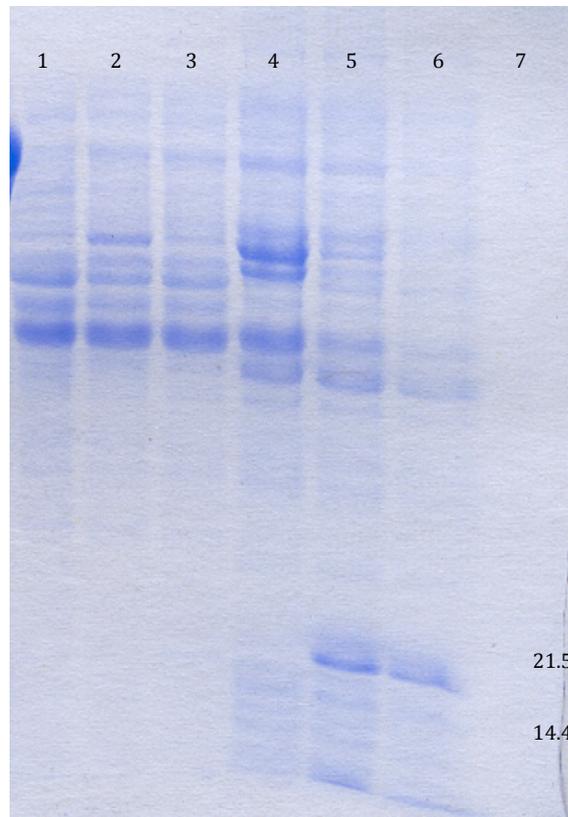


Figure 4.6: SDS-PAGE analysis of full-length BPSL1204 (19.2 kDa) over-expression at 37°C for 3 hours.

Lane 1-3 = No induction, 0.1 and 0.01% arabinose soluble

Lane 4-6 = No induction, 0.1 and 0.01% arabinose insoluble

Lane 7 = Mark12 (kDa)

Full-length BPSL1204 is insoluble when over-expressed in *E. coli*. This was assumed to be due to the presence of an N-terminal anchoring sequence usually cleaved in native *B. pseudomallei*.

insoluble BPSL1204 suggests the presence of a non-folded, non-localised pre-prolipoprotein. This led to strengthen the hypothesis that BPSL1204 functions whilst anchored to a membrane.

4.4 Purification of BPSL1204-MA

4.4.1 Fractionation of Cell-Paste

A portion of the cell-paste was re-suspended in 10 ml Tris pH 8.0 per 1 g of cells. To achieve enough pure protein per round of purification in order to carry out efficient screening or optimisation it was established that 3.5 g of cells were to be used at the start of each purification, re-suspended in 35 ml buffer. The suspension was then sonicated on ice for 3 x 20 seconds at 16 microns and the insoluble fraction was pelleted by centrifugation, as described in section 3.6.1. The 35 ml of soluble supernatant was used immediately for DEAE ion-exchange chromatography.

4.4.2 DEAE Ion-Exchange Chromatography

The 35 ml cell-free extract was applied to a Superloop attached to the injection valve of an AKTAprime purification machine connected to a 20 ml HiPrep 16/10 DEAE-FF column from Amersham Pharmacia that had all been pre-equilibrated with 50 mM Tris pH 8.0. The cell-free extract was loaded onto the column from the Superloop by injecting at 4 ml/min, and the flow-through was kept for analysis. Once loaded, the column was then washed with 60 ml of 50 mM Tris pH 8.0 to remove any unbound material and the eluate was collected for analysis.

Once the UV-trace continuously read at the pre-equilibrated base line, a gradient of 0-0.5 M NaCl (by increase of 0-50% of a 50 mM Tris pH 8.0, 1 M NaCl buffer off-set against 50 mM Tris pH 8.0) was applied over 300 ml whilst 4 ml fractions were collected. The peaks that appeared on the UV-trace were analysed by 13.5% resolve SDS-PAGE to identify the peak that contained BPSL1204-MA. Figure 4.7 shows the analysis of the peaks post-DEAE chromatography.

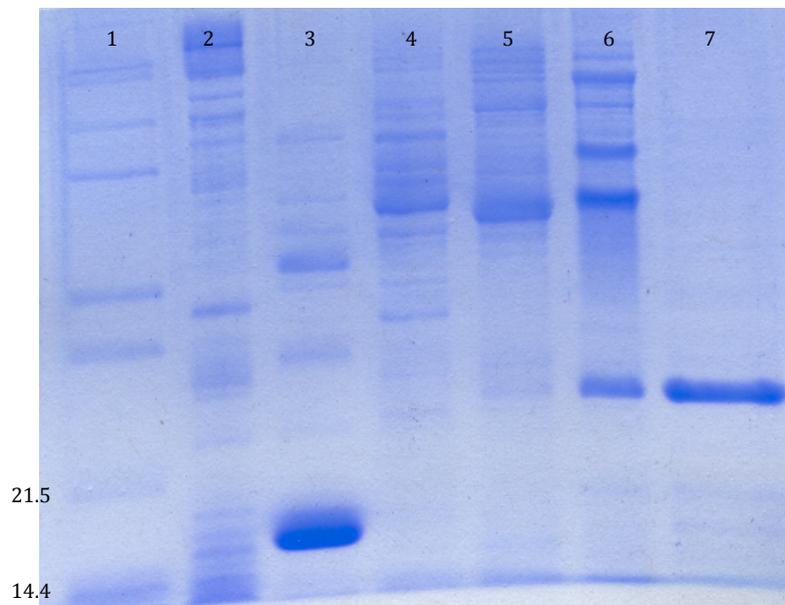
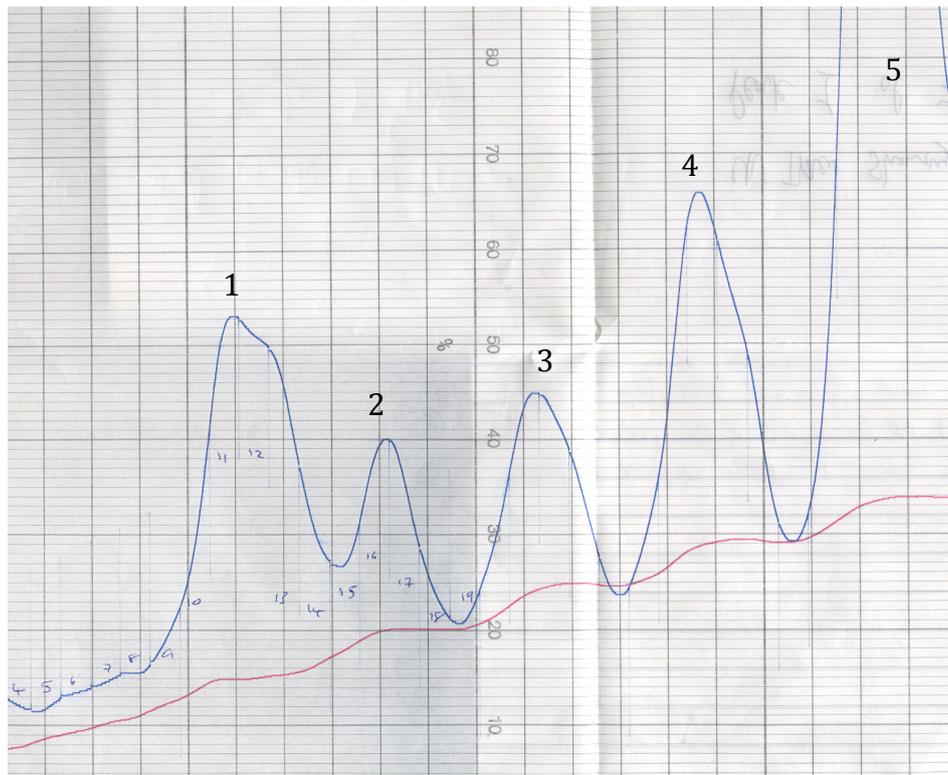


Figure 4.7: DEAE ion exchange of BPSL1204-MA. The UV trace of fractions collected from a DEAE column loaded with cell-free extract (*top*) showed 5 peaks eluted over a 0-0.5 M NaCl gradient. SDS-PAGE analysis of peak fractions (*bottom*) showed that BPSL1204-MA eluted in the first DEAE peak.

Lane 1 = Mark 12 (kDa)
 Lane 2 = Crude cell-free extract
 Lane 3-7 = Fractions under peaks 1-5.

BPSL1204-MA elutes at approximately 80 mM NaCl in 50 mM Tris pH 8.0 in the first peak of the experiment. The 5 fractions contained within the BPSL1204-MA peak were combined for further purification by $(\text{NH}_4)_2\text{SO}_4$ precipitation. A standard combined concentration was 0.9 mg/ml, as determined by Bradford Assay, thus taking forwards 18 mg of protein in 20 ml.

4.4.3 Ammonium Sulphate Precipitation

The combined 20 ml taken forwards from ion-exchange chromatography was transferred to a sterile centrifuge tube and combined with 8 ml of 4 M $(\text{NH}_4)_2\text{SO}_4$ to create a final 28 ml mixture of protein in 1.6 M $(\text{NH}_4)_2\text{SO}_4$. At this stage it was possible to see protein precipitating out of solution as the mixture became a cloudy, white suspension. Precipitation was allowed to occur for 5-10 minutes on ice before the insoluble precipitate was pelleted by centrifugation at 24,500 rpm in a JA25.50 rotor at 8°C for 15 minutes.

The soluble fraction was removed by decanting the bulk of the solution and carefully pipetting from the surface of the pellet and stored at 4°C for future analysis. The pellet was resuspended in 1.5 ml of 50 mM Tris pH 8.0 by pipetting and then transferred to a 1.5 ml Eppendorf which was then spun at 13000 rpm in a Sigma 3-16K 12154 rotor for 5 mins at 8°C to pellet permanently insoluble material. This second supernatant was then assessed for the presence of BPSL1204-MA by SDS-PAGE as shown in Figure 4.8 and then taken forward to the next stage of purification by gel filtration. A standard concentration after re-suspension of the $(\text{NH}_4)_2\text{SO}_4$ pellet was 6.5 mg/ml, as determined by Bradford Assay, thus taking forwards 9.75 mg of protein in 1.5 ml of buffer.

4.4.4 Gel Filtration

A 120 ml HiLoad 16/60 Superose 6 gel filtration column from Amersham Pharmacia was connected to an AKTAPrime purification machine and equilibrated with a buffer of 50mM Tris pH 8.0, 0.5 M NaCl at 1 ml/min for 1.5 column volumes. The 1.5 ml re-suspension from the $(\text{NH}_4)_2\text{SO}_4$ precipitation was

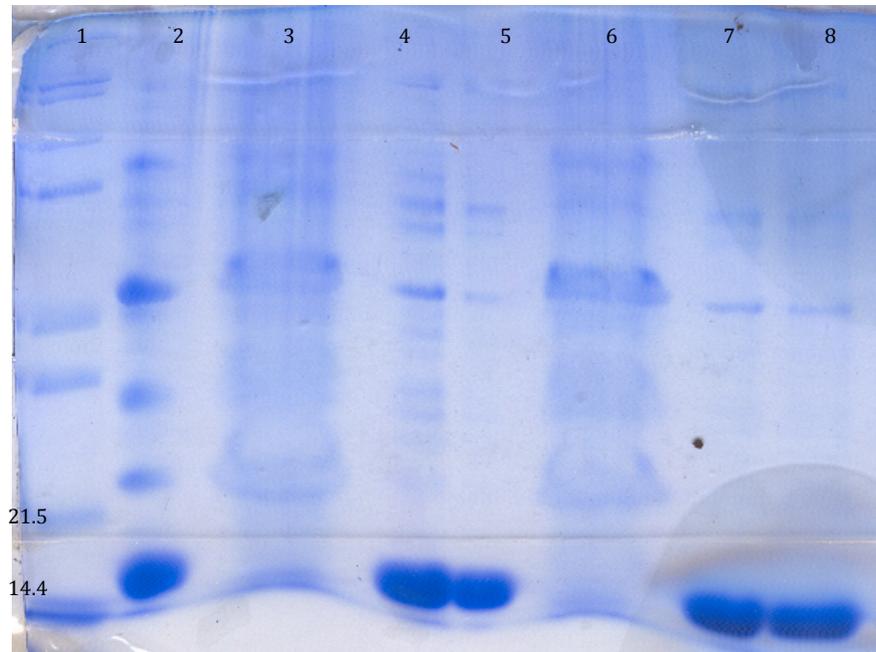


Figure 4.8: SDS-PAGE analysis of $(\text{NH}_4)_2\text{SO}_4$ -precipitation of BPSL1204-MA from combined DEAE fractions.

Lane 1 = Mark12 (kDa)

Lane 2 = Combined DEAE fractions

Lane 3 = Supernatant post-2 M $(\text{NH}_4)_2\text{SO}_4$

Lane 4 = Re-suspended 2 M $(\text{NH}_4)_2\text{SO}_4$ pellet

Lane 5 = Insoluble material post-2 M $(\text{NH}_4)_2\text{SO}_4$ pellet re-suspension

Lane 6 = Supernatant post-1.5 M $(\text{NH}_4)_2\text{SO}_4$

Lane 7 = Re-suspended 1.5 M $(\text{NH}_4)_2\text{SO}_4$ pellet

Lane 8 = Insoluble material post-1.5 M $(\text{NH}_4)_2\text{SO}_4$ pellet re-suspension

It was considered that 1.5 M $(\text{NH}_4)_2\text{SO}_4$ precipitation purified BPSL1204-MA from more contaminants than 2 M $(\text{NH}_4)_2\text{SO}_4$. However, both concentrations led to the generation of some insoluble material (Lanes 5 and 8). This effective technique was used as a compromise between purification speed and yield.

loaded onto the top of the column and then the column was immediately re-connected and the gel filtration program run on the AKTAPrime. At 1 ml/min a 45 ml void-elution was allowed to run into waste-disposal, and then the remaining 75 ml was collected in 2 ml fractions alongside the recording of a UV trace. The peak fraction on the trace was assessed for the presence of pure BPSL1204-MA. The 4 fractions under the peak were combined for concentration and buffer-exchange by Viva-spin centrifugation.

A standard concentration after gel filtration was 0.5 mg/ml, as determined by Bradford Assay, thus taking forwards 4 mg of protein. The gel filtration also acted as a method of buffer-exchange away from residual $(\text{NH}_4)_2\text{SO}_4$. Analysis of the gel-filtration trace against a standard curve from the column used suggested the elution of pure, totally monomeric BPSL1204-MA (Figure 4.9) with the purity of the final protein sample and the effectiveness of the purification protocol assessed by SDS-PAGE (Figure 4.10). BPSL1204-MA was purified to a high purity.

4.4.5 N-terminal Sequencing

The presence of BPSL1204-MA in the final purified sample was confirmed by Edman-degradation N-terminal sequencing (Niall, 1973). There was little to no background contamination detected by this procedure. Dr Arthur Moir of the University of Sheffield carried out the sequencing.

4.5 Synthesis of a Selenomethionine Derivative

4.5.1 Large-scale Over-expression and Purification of SeMet BPSL1204-MA

Following the protocol described in section 3.5.2, 4 x 500 ml cultures at 0.7 OD 600nm, each in 2 L baffled-flasks, were induced with 1 mM IPTG for 24 hrs at 37°C and 250 rpm. 7 g of cell-paste was harvested and stored as 2 x 3.5 g pellets at -80°C until further use in purification. Soluble over-expression was confirmed by SDS-PAGE as shown in Figure 4.11. The purification protocol employed for the native BPSL1204- protein was applied to the seleno-met derivative, and was

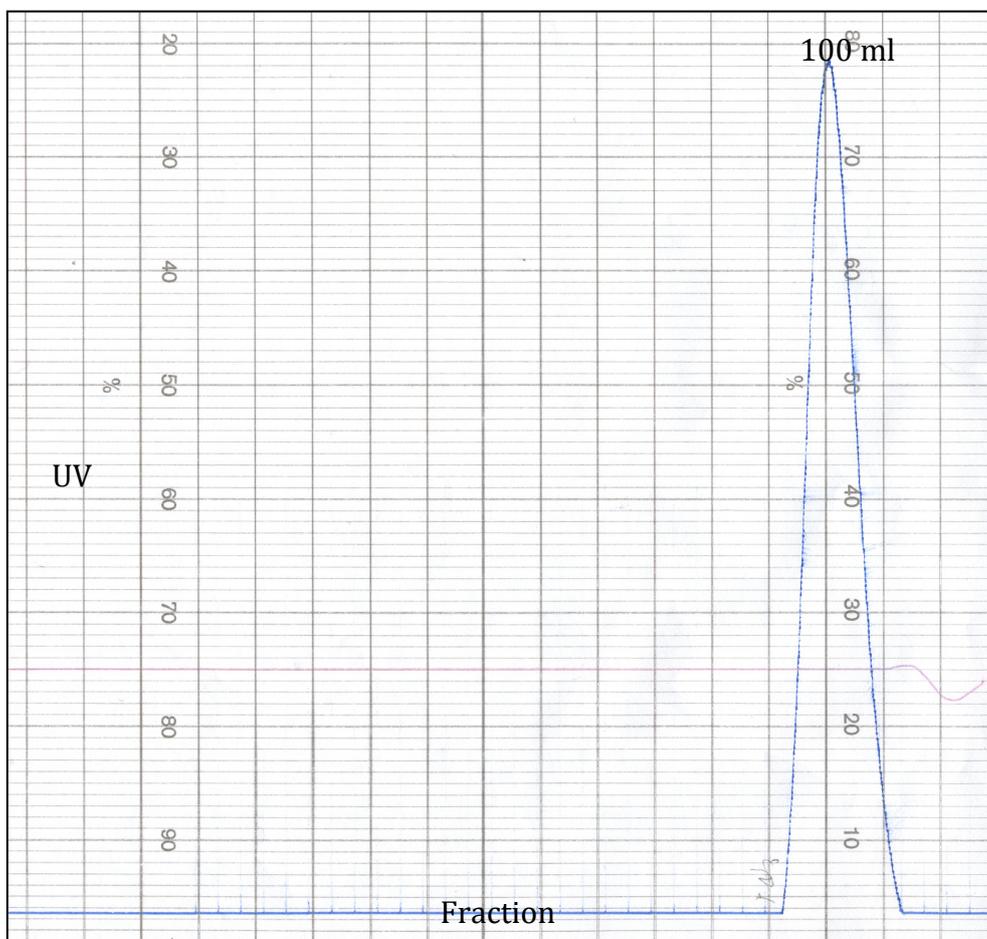


Figure 4.9: Gel filtration trace of BPSL1204-MA purification The UV sensitivity was set to 1 V and 2 ml fractions were collected. A single monomeric peak of BPSL1204-MA was witnessed in Fraction 27, after an elution volume of 100 ml.

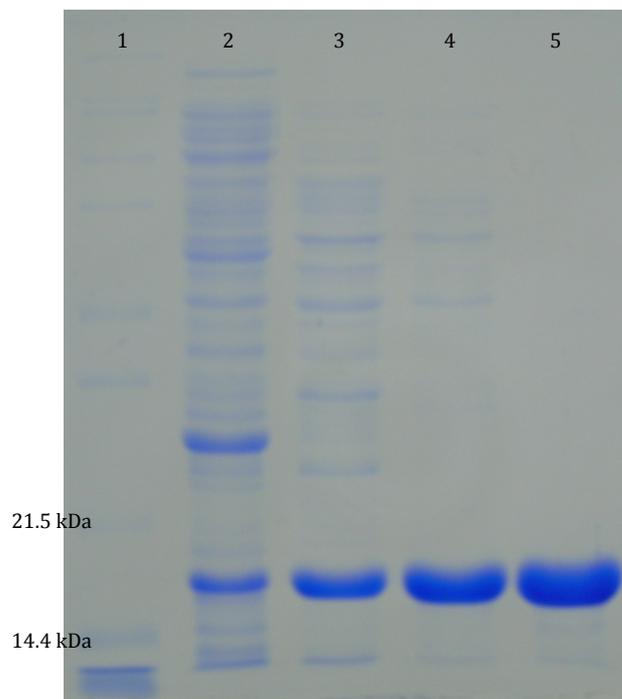


Figure 4.10: SDS-PAGE analysis of the BPSL1204-MA purification protocol.

- Lane 1: Mark 12 (Invitrogen)
- Lane 2: Cell-free Extract
- Lane 3: Post DEAE
- Lane 4: Post $(\text{NH}_4)_2\text{SO}_4$
- Lane 5: Post Gel Filtration

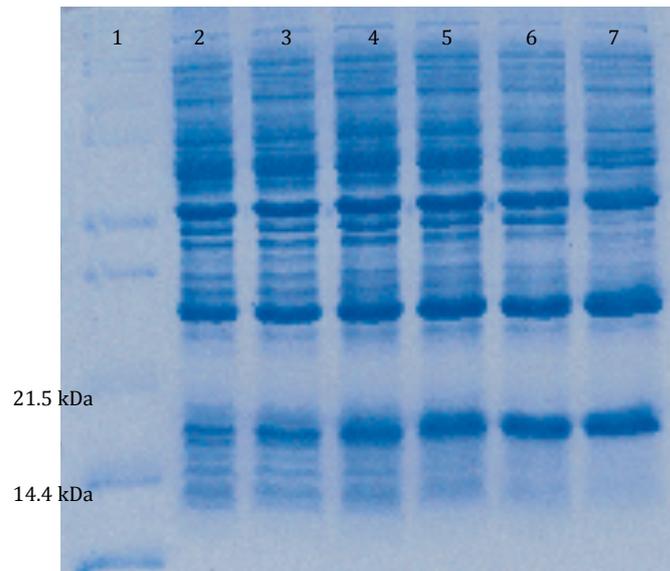


Figure 4.11: SDS-PAGE analysis of Selenomethionine BPSL1204-MA over-expression.

Lane 1: Mark 12 (Invitrogen)

Lane 2: 4 hours

Lane 3: 6 hours

Lane 4: 12 hours

Lane 5: 16 hours

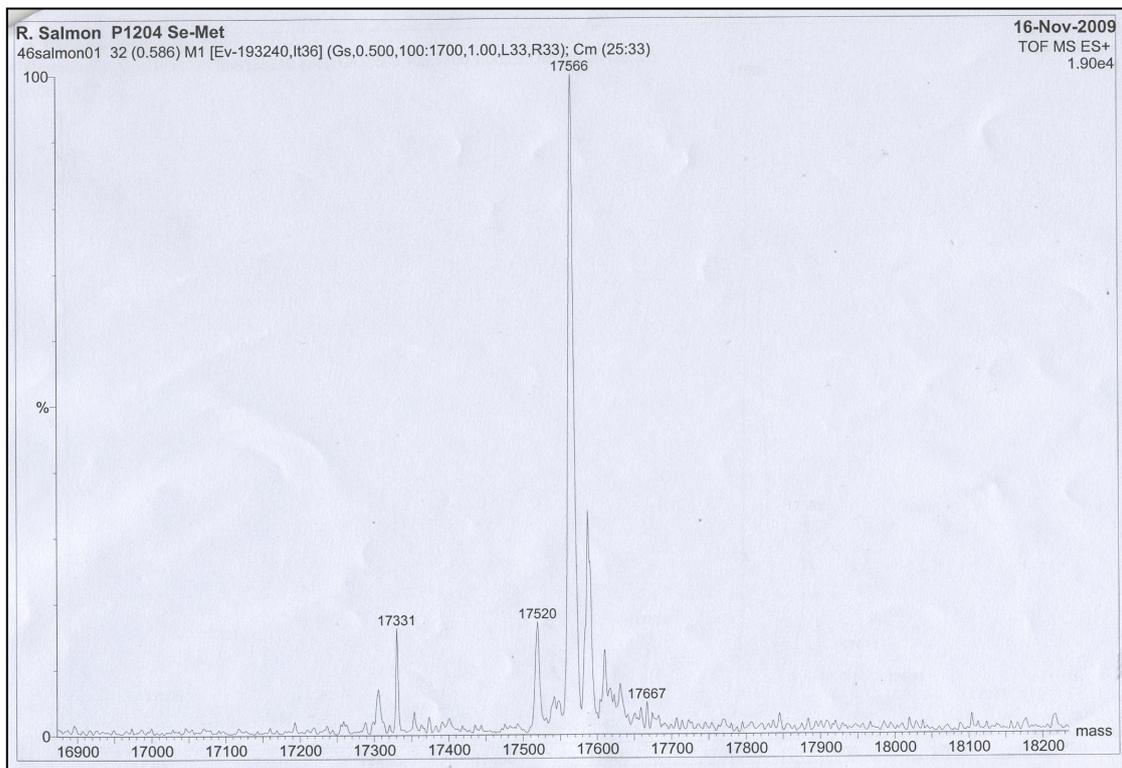
Lane 6: 20 hours

Lane 7: 24 hours

successful in purifying 3 mg of pure, seleno-met BPSL1204-MA in 100 μ l from 3.5 g of cell-paste.

4.5.2 Analysis of SeMet Incorporation

3 μ l of 30 mg/ml seleno-met BPSL1204-MA was analysed by mass spectrometry (Domon & Aebersold, 2006) by Simon Thorpe of the University of Sheffield Chemistry Department. As can be seen in Figure 4.12, the mass spectra produced confirmed the 100% incorporation of 5 selenomethionine into BPSL1204-MA. This was something of a surprise because it implied that the N-terminal methionine not present in the native BPSL1204-MA was present in the seleno-met derivative, presumably reflecting differences in the processing of expressed polypeptides in the presence of seleno-methionine.



	4 Met BPSL1204-MA (Da)	5 Met BPSL1204-MA (Da)
S-Met	17197	17329
+1 Se-Met	17244	17376
+2 Se-Met	17291	17423
+3 Se-Met	17338	17470
+4 Se-Met	17385	17517
+5 Se-Met	N/A	17564

Figure 4.12: Mass spectra of Se-Met BPSL1204-MA (*top*) and table of possible Se-Met incorporation masses (*bottom*). The mass spectra show a mass of 17566 Da recorded, which corresponds to the incorporation of 5 (100%) Se-Met residues into BPSL1204-MA

Chapter 5: Crystallisation and Structure Determination of BPSL1204

5.1 Protein Sample Preparation

5.1.1 Buffer-exchanging the Final Sample

After the combination of gel filtration fractions under the peak of the UV trace (6-8 ml), pure BPSL1204-MA was reduced to a volume of approximately 300 μ l using a 6 ml VivaSpin 10 kDa centrifugal filter-column, available from Sartorius, by centrifuging in a Sigma 3-16K 11133 rotor for 20-30 minutes at 4500 rpm at 10°C. For crystallographic screening the lowest concentration buffer with no contaminating salt was required. In order to remove the 0.5 M NaCl from the gel filtration buffer and exchange the 50 mM Tris pH 8.0 buffer into 10 mM Tris pH 8.0, the 300 μ l was then re-suspended to 6 ml by the addition of 5.7 ml of 10 mM Tris pH 8.0. This was then reduced to 300 μ l again, and then re-suspended back to 6 ml with fresh 10 mM Tris pH 8.0. This left approximately 4.5 mg of pure protein in 6 ml of 10 mM Tris pH 8.0 with less than 1.5 mM NaCl.

5.1.2 Concentration of Protein Sample for Crystallography

In order to raise the concentration of the protein solution to a level that would allow super-saturation of protein leading to crystallisation via vapour-diffusion (Chayen & Saridakis, 2008), the pure protein in low concentration buffer was ideally to be of a final sample concentration of greater than 8-10 mg/ml. Using the same VivaSpin column as in the buffer-exchange step, the 6 ml sample in 10 mM Tris pH 8.0 was centrifuged at 4500 rpm at 10°C for 50-60 minutes until the solution reached a suitable concentration. The sample was then removed by pipetting from the VivaSpin. The sample was pipetted up and down from the bottom of the filter several times before removal, to ensure maximum retrieval of protein from the filter-base, where protein concentration was highest post-centrifugation. From 3.5 g starting mass of cell-paste it was typically possible to achieve 4 mgs of pure BPSL1204-MA in a 200 μ l solution at 20 mg/ml protein, buffered by 10 mM Tris pH 8.0. The Bradford Assay was used to determine protein concentration.

5.2 Automated Screening of Crystallisation Conditions

5.2.1 Creation of Sitting-drop Robot Trials

The pure, concentrated sample of BPSL1204-MA was used to carry out initial screening of potential crystallisation conditions using a Hydra II crystallisation robot manufactured by Matrix Technologies. Each of the following screens used, bought from Qiagen Nextal, consisted of 96 different conditions per screen: JCSG+, Classics, pHClear, PEG, PACT, Crystal Screen I, Ammonium Sulphate, MPD. The contents of each screen can be found via the Nextal website (www.qiagen.com). For each condition 200 µl were pipetted into the 96 reservoirs of a plastic MRC sitting-drop plate (Molecular Dimensions), from which approximately 200 nl were transferred to the raised platform-dimple of each plate. Onto the drop in each dimple the robot's nano-needle aspirated approximately 200 nl of the pure protein sample. Each 96-condition plate was then sealed using a sheet of airtight and adhesive ClearVue (Molecular Dimensions) to create 96 individually sealed vapour-diffusion experiments. Trays were centrifuged at 2000 rpm in a Grant-Bio LMC-3000 R-2 rotor for 2 minutes at 19°C to ensure proper mixing of drops in the raised platforms. These screens were each repeated several times over an extended period of several months, in order to vary incubation temperature, protein concentration used and freshness of protein sample, dependant on the short-term results of the previous repeat of the screen.

5.2.2 Incubation of Screens

Screens were set up in duplicate, and left at 2 different temperatures. One was left to equilibrate in a temperature-controlled room set to 17°C whilst the other copy of the screen was left refrigerated at 7°C. This was in case nucleation happened too quickly, too slowly or not at all at one temperature in differing conditions, but was appropriate at another temperature (Judge et al, 1999).

5.2.3 Examination of Growth Conditions

Screens were investigated immediately post-sealing to note and establish those conditions that led to immediate precipitation. These were then discounted from the list of potential conditions straight away. In approximately one quarter of all conditions, BPSL1204-MA precipitated immediately. Each screen was then gently removed from its incubator 1-day post-trial, to again either establish further conditions that had precipitated over a short incubation period or to witness the fast over-night growth of any crystals. In the case of BPSL1204-MA, several more conditions per screen had precipitated, leaving behind roughly half the conditions still with clear drops and thus potentially still able to yield crystals. No crystals were witnessed growing overnight in any BPSL1204-MA trial. From then on, screens were checked twice weekly for the first month, then weekly for the months after that.

5.2.4 Selection of Potential Hits

Over a period of several months, screening was carried out and the resulting sitting-drop trays monitored. Screens in sitting-drop robot trays were labelled as rows A-H with columns 1-12. Table 5.1 summarises the eight conditions that generated low-quality crystals from all variations of all screens done for BPSL1204-MA.

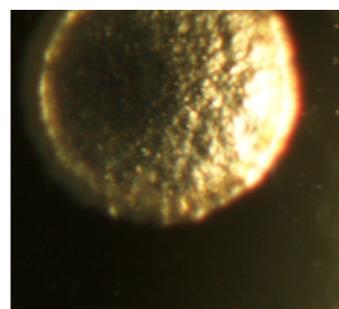
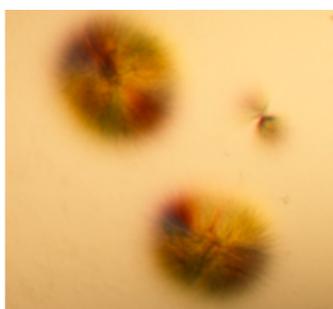
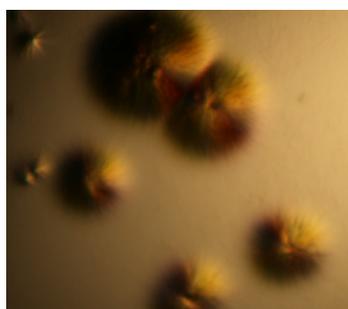
5.3 Optimisation to Hanging-drop Trays

5.3.1 Broad-range Screening Around Potential Hits

Broad-range screens of hanging-drop trials were created, centred on the initial hits described in 5.2.4, following the method described in 3.7.3. The first hanging-drop screens consisted of $\pm 6\%$ precipitant (in 2% steps), variation of

Screen	Well	Condition	Protein mg/ml	Incubation Temp (°C)	Approx. Time to Crystallise (days)
JCSG+	C3	0.2M Ammonium nitrate pH6.3 20% PEG3350	13.5	17	60
JCSG+	C4	0.1M HEPES pH7.0 10% PEG6000	8	17	20
JCSG+	C11	2M AmSO4 0.1M Sodium acetate pH4.6	13.5	17	3
MPD	E2	0.1M Sodium acetate pH5.0 10% MPD	24.5	17	10
PEG	A8	0.1M MES pH6.5 40% PEG300	24.5	17	15
PEG	B8	0.1M Tris.HCl pH8.5 30% PEG300	24.5	17	15
PEG	F2	0.2M Potassium thiocyanate 20% PEG3350	24.5	17	15
PEG	G5	0.2M Potassium acetate 20% PEG3350	24.5	17	15

Table 5.1: The table shows the potential crystallisation conditions found after investigation of initial screening attempts. The images below show initial hits, from left to right, of JCSG+ C3, MPD E2 and JCSG+ C11.



buffers by ± 0.5 pH (as well as the original pH) and variations of additives by ± 0.1 M (as well as the original concentrations). Hanging-drops were created as 1 μ l protein:1 μ l reservoir solution, using 20 mg/ml BPSL1204-MA. Of the initial hits described in 5.2.4 only JCSG+ C3, C4, C11 and MPD E2 were reproducible from sitting-drop to the first stage of hanging-drop optimisation over a 17°C incubation period of less than 4 weeks. Each formed thin, ball-like needle clusters that seemed a little thicker than those discovered in the initial robot-screens. Attempts to optimise PEG A8, B8, F2 and G5 were taken no further after the lack of success in this first stage of optimisation.

5.3.2 Focussing onto Optimal Growth Conditions

After reproduction and slight improvement of the thin rods in the initial robot-screens, the condition from each that gave the thickest looking clusters, or sometimes thin individual rods, were further optimised by $\pm 3\%$ precipitant (in 0.5% steps), ± 0.25 pH (as well as the original optimised hit pH) and the same concentration of additive as in the original optimised hit. Fresh protein was purified for each round of optimisations. Each hanging-drop condition of the second round of optimisations was done in triplicate in order to both increase the number of crystals produced should they form useful crystals, and also to screen for success even if one drop failed to equilibrate or nucleate properly for an unexpected reason. These were then incubated at 17°C and checked twice weekly.

Of the 4 screens taken forward to the second-round of screening, only JCSG+ C4 and MPD E2 showed any marked improvement in forming clusters of thicker needles and rods over a period of 3 weeks. These were then optimised further by $\pm 2\%$ of the best precipitant % (in 0.25% steps) and ± 0.25 pH (as well as the current best pH). Each condition was repeated in triplicate, and for each member of a triplicate 4 drops were set up of μ l protein: μ l reservoir ratios of 1:1, 1:2, 2:1 and 2:2 to investigate the effects of these ratios on nucleation speed, crystal quality and the number of crystals produced per drop. These were then incubated at 17°C and checked twice weekly.

5.3.3 Production of High Quality BPSL1204-MA Crystals

Investigation after 2 weeks of the third round of optimisations showed no improvement on the relatively low quality needle/rod-cluster formed by optimisations of JCSG+ C4, so this condition was pursued no further. However, MPD E2 optimisations had generated large rod-clusters with individual rods reaching approximately 30-40 μm in width and 300-400 μm in length. The clusters were all tightly packed with crystals growing in all directions, but each seemed to grow as a single rod that shared a nucleation point with the others (Figure 5.1a). The best rod-clusters were grown in 1:1 μl drops of 9% MPD and 0.1 M sodium acetate pH 5.0 using a 1 ml reservoir, although the production of crystals with reasonable visual quality occurred in 8.5-10.5% MPD with 0.1 M sodium acetate of pH 5.25-5.75 (Figure 5.1b). Increasing the drop size ratio to 2:2 seemed to proportionately increase the number of nucleation points leading to more rod-clusters like those in 1:1, which in practical terms just led to 'busier' drops, making the process of obtaining an individual rod more difficult. These crystals (1:1, 9% MPD, 0.1 M sodium acetate pH 5.0 with 17°C incubation for 2 weeks) were considered the best possible for this condition, and so samples were taken for X-ray diffraction studies.

5.4 In-house Diffraction Studies

5.4.1 Establishing Cryo-buffers

In order to screen for an appropriate cryo-protectant buffer that varied the mother-liquor of the crystal under investigation the least, the 9% MPD/0.1 M sodium acetate pH 5.0 condition was supplemented with 0-25% glycerol or ethylene glycol in 5% increases, to generate 12 cryo-buffers for testing. As MPD itself could act as a cryo-protectant, it was hoped that relatively little would need to be added to prevent the freezing of water within the condition selected. A 300 μm loop was filled with each cryo-buffer in turn, mounted onto the goniometer so that the loop was in the 100 Kelvin nitrogen-stream and in the path of X-ray radiation, and then was exposed to X-rays for 2 minutes. If the resulting

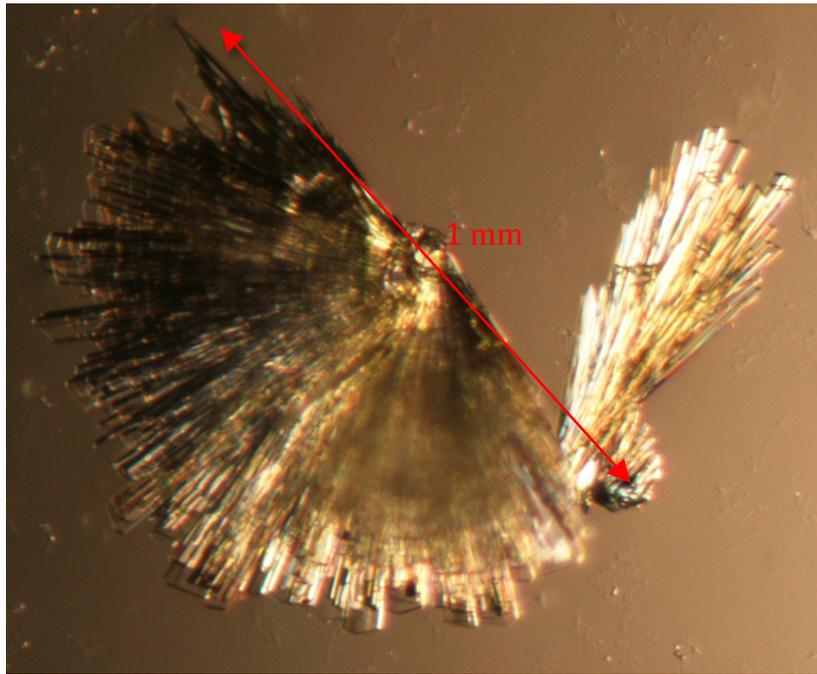


Figure 5.1a: Condensed rod-cluster grown in 1:1 6% MPD and 0.1 M sodium acetate pH 5.5

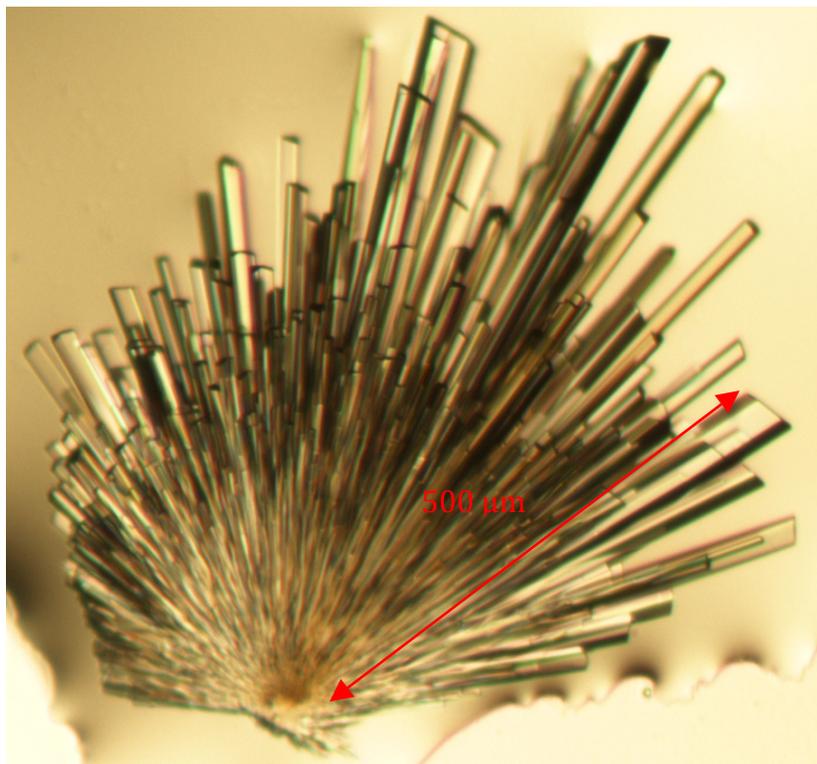


Figure 5.1b: Cluster of thick rods grown in 1:1 9% MPD and 0.1 M sodium acetate pH 5.0. A single rod was taken from this cluster for native data collection

detector-image showed ice-rings (commonly most prevalent at $\sim 3.5 \text{ \AA}$ resolution) the buffer was discarded and the next 5% increase tested. Detector-images of cryo-buffers with 0-5% glycerol and 0-10% ethylene glycol presented ice-rings, so therefore mother-liquor supplemented with 10% glycerol was used as a cryo-buffer.

5.4.2 Mounting Individual 1204-MA Crystals

Under a microscope, large rod-clusters were shattered at their core using a sterile acupuncture needle. Then by combining physical but gentle separation of crystals using the needle and manipulating the surface tension of the drop to align suitable candidates, several conjoined, thick rods were looped from the mother-drop and quickly but carefully transferred to a 3 μl drop of fresh cryo-buffer in place on a siliconized, polished coverslip. Usually the act of transferring the crystals led to further separation for easier mounting of an individual rod. Often attempting to loop an individual crystal directly from the mother-drop led to the generation of lots of crystal-debris and the loss of other potentially useful crystals. Crystals were left to soak in the cryo-buffer for no more than 120 seconds before being looped onto the X-ray machine, ensuring that the loop chosen supported the crystals in potential 'shatter-zones' whilst providing a suitable volume of cryo-protectant. Rods were aligned so that the length of the crystal was parallel to the direction of the neck of the mounted loop. This way a long crystal could be rotated and shot along its length using the finer focus beams available at synchrotrons (Axford et al, 2012).

5.4.3 Detection of Diffraction

Crystals were only tested in-house, with full data collection left for a synchrotron light source. Two test-images of 1° oscillation through φ , 90° apart, were collected using an exposure of 5 minutes at 8.014 KeV/ 1.546 \AA CuK α , with the detector set to 200mm distance. The best diffraction achieved in tests, later to become the high-resolution native dataset, was to a maximum resolution of 2.1 \AA in-house. When crystals that when tested led to initial calculation of symmetry space-group and cell-dimensions were detected, they were saved in a vial of

liquid nitrogen and stored in a nitrogen-filled dewar for transportation to a synchrotron.

5.4.4 Space-group, Cell-refinement and Collection Strategy Calculation.

In-house test images were loaded into iMOSFLM (Battye et al, 2011) and subjected to auto-indexing of reflections once parameters such as beam-centre and resolution limit had been selected. iMOSFLM then calculated the potential space-groups from the two images supplied, and the most likely selected was the order of symmetry with the least number of penalties after discounting the zero-penalty P1. In the case of BPSL1204-MA, the most likely space-group was C2 (C_{121}). These images were then processed further by iMOSFLM to refine cell-parameters using the chosen space-group, with BPSL1204-MA in C2 given dimensions of $a=99.96 \text{ \AA}$, $b=32.91 \text{ \AA}$, $c=48.35 \text{ \AA}$, $\alpha = 90.0^\circ$ $\beta=116.14^\circ$ $\gamma = 90.0^\circ$ (see Table 5.2 in Section 5.10.4). Combined, this information allowed the generation of a data-collection strategy using iMOSFLM, which suggested a full dataset minimising overlaps would require the collection of 180° of data consisting of 180 images, each image having an oscillation of 1° about φ .

5.5 Collection and Analysis of Native Diffraction Data.

5.5.1 Collection of Diffraction Images Using a Synchrotron Light Source

All native data were collected at the Diamond Synchrotron Light Source (DSLS) in Oxford, UK. Using the beamline I02 at the standard energy of 12.658 KeV (0.979 \AA), data were collected for each crystal under experiment using the strategy predicted by iMOSFLM, following the same process as in Section 5.4.4. 200° were collected for each data set to ensure completeness, using an exposure of 0.5 seconds per image at 20-25% beam-transmission. The detector plate was set to allow for data-collection to resolutions of 1.2 \AA . Test-images taken showed that from the best crystals diffraction occurred to a resolution limit of 1.295 \AA (Figure 5.2), compared to the 2.1 \AA achievable in-house.

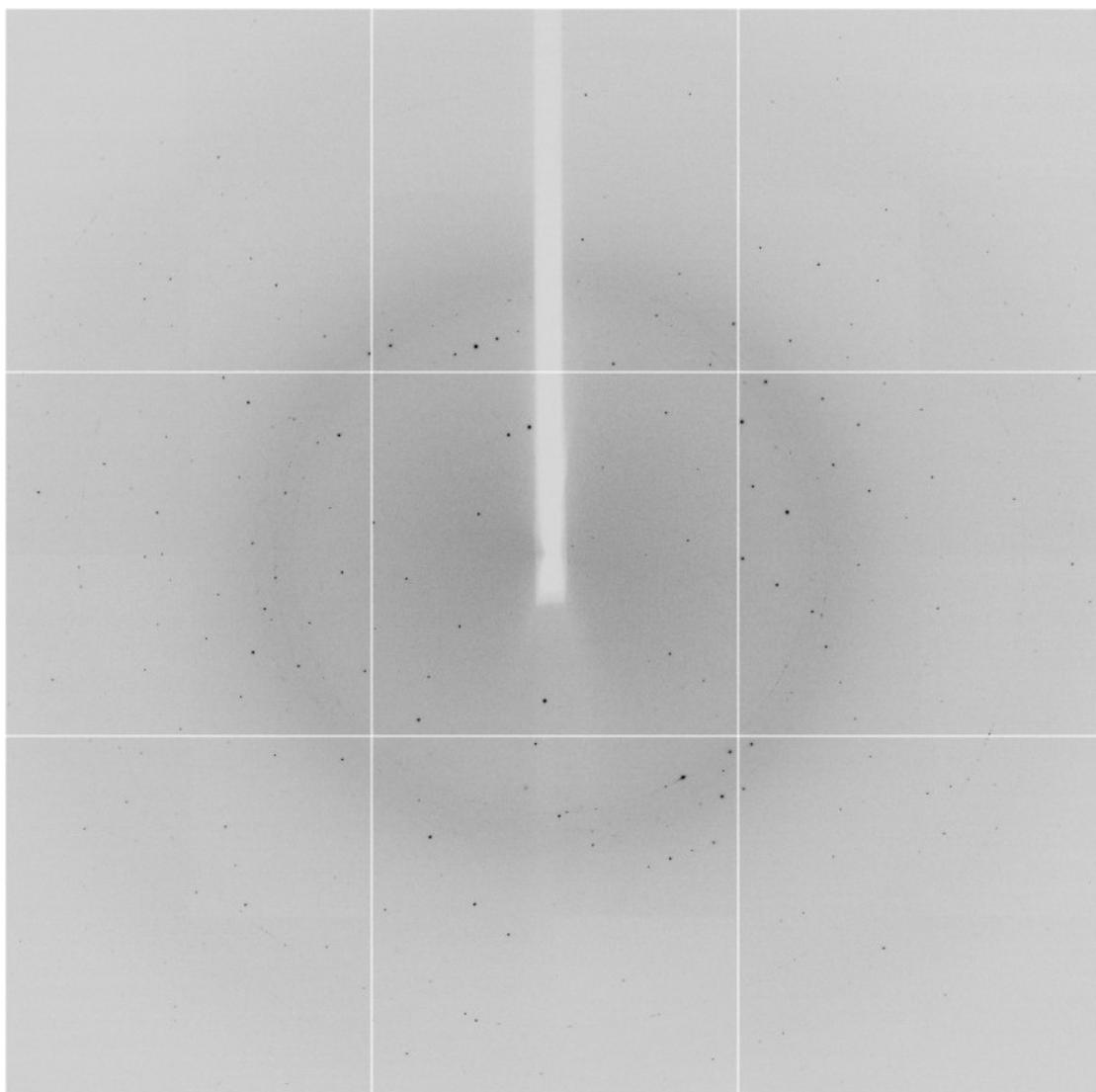


Figure 5.2: A diffraction image of native BPSL1204 crystals to 1.295 Å, collected on the I02 beamline at DSLS. The detector plate was set to a 1.2 Å resolution limit at the edges of the image above

5.5.2 Indexing, Integration and Reduction of Native Data

Native data images had their reflections indexed by iMOSLM using the default settings for $I/\sigma(I)$ spot discrimination threshold, spot size and separation thresholds and the box-size of the local background. The beam-centre was set and the region of no diffraction caused by the beam-stop was masked and applied to all images collected. The resolution limit of spots to be indexed was also set to cut off at 1.2 Å. Reflections not adhering to these parameters were rejected as part of data reduction. Indexed images, with mosaicity of 0.58° were then integrated into an .mtz file once cell-refinement had calculated a sensible unit-cell. The output nat.mtz file was then taken forward to the CCP4 Suite (Winn et al, 2011) for further data processing. A table of cell parameters (Table 5.2) is included later on in this chapter, for comparison with seleno-met derivative crystals (Section 5.8.2)

5.5.3 Scaling of Data, Resolution Limits and Data Quality

The .mtz file containing the intensities of integrated reflections from the entire native dataset were then scaled and merged together using the Scala program (Evans, 2006) in the CCP4i GUI (Potterton et al, 2003). Upon investigation of the output log file (Figure 5.3), it was decided that the resolution limit of data that gave the best signal-to-noise ($I/\sigma(I) = 2.2$, accepted as >2.0), completeness (99.6%) and acceptable high resolution R_{merge} (0.34, accepted as <0.35) was 1.3 Å. With the resolution limits now set, the data were scaled back to only merge intensities <1.3 Å, with a re-run of the Scala process formatting the output file into .sca (Scalepack) format, for further processing in HKL2MAP (Pape and Schneider, 2004) running ShelXC/D/E (Sheldrick, 2010) when datasets for phasing became available.

With the known cell-dimensions and known molecular weight of BPSL1204-MA as 17200 Da, the Matthews Coefficient program Matthews_coef was also run in CCP4i to predict the number of molecules in the unit-cell using the input scaled .mtz file (Kantardjieff & Rupp, 2003). It was predicted that in the calculated cell-

```

By 4SINTH/LASQ bins (all statistics relative to Mn(I))
-----
capplet width=" 700" height=" 300" code="JLogGraph.class"
;odebases="/usr/local/ccp4-6.1.2/bin"><param name="table" value="
$TABLE: Analysis against resolution nat 3
$GRAPHS: I/sigma, Mean Mn(I)/sd(Mn(I)):N:2,11,13:
: Rmerge v Resolution:N:2,4,5,7:
N 1/resol^2 Dmin(A) Rmrg Rfull Rcum Ranom Nanom Av_I SIGMA I/sigma sd Mn(I/sd) Nmeas Nref Ncent FRCBIAS Nbias
bias $$
N 1/d^2 Dmin(A) Rmrg Rfull Rcum Ranom Nanom Av_I SIGMA I/sigma sd Mn(I/sd) Nmeas Nref Ncent FRCBIAS Nbias
1 .0594 4.10 .035 .032 .035 - 0 18589 1157 16.1 1498 22.3 3933 1114 174 .017 717
2 .1187 2.90 .039 .037 .037 - 0 16304 1101 14.8 1290 21.6 7084 2042 190 .006 1389
3 .1781 2.37 .064 .059 .043 - 0 5864 664 8.8 496 19.0 9187 2608 177 -.015 1736
4 .2374 2.05 .069 .063 .047 - 0 4360 516 8.5 394 17.6 11222 3124 188 .002 2102
5 .2968 1.84 .065 .060 .049 - 0 2529 296 9.9 266 14.6 12667 3489 186 .005 2479
6 .3561 1.68 .087 .078 .051 - 0 1254 157 8.0 170 11.4 14034 3855 181 .024 2722
7 .4155 1.55 .120 .110 .054 - 0 824 138 6.0 150 8.8 15241 4183 182 .028 2891
8 .4748 1.45 .171 .156 .057 - 0 568 138 4.1 144 6.5 16355 4511 193 -.003 3208
9 .5342 1.37 .241 .220 .060 - 0 391 126 3.1 141 4.8 17312 4791 196 .020 3334
10 .5935 1.30 .341 .315 .064 - 0 275 125 2.2 140 3.5 18141 5047 205 .004 3494

```

Figure 5.3: SCALA output file after scaling and data reduction. Native data was scaled and cut to 1.3 Å resolution gave a I/σ of 2.2 and an R_{merge} of 0.34

volume of 142910 \AA^3 one molecule of BPSL1204-MA existed with a solvent content of 40.8%.

5.6 Heavy-atom Soaks and Initial Phasing Attempts

5.6.1 Heavy-atom soaks

While attempts were being made to prepare and crystallise selenomethionine-incorporated protein, which proved difficult (see Section 5.7), it was decided to attempt some heavy atom soaks. With a reliable source of high quality, diffracting crystals attempts to solve the phase problem began. Phasing required the anomalous signal created by the presence of regularly distributed, electron-dense atoms incorporated into the crystal lattice sub-structure (Schneider & Sheldrick, 2002). A variety of water-soluble heavy-atom compounds were soaked into preformed native crystals in the hope that potential sites of occupation existed within the protein or within the protein-lattice. $\text{KAu}(\text{CN})_2$, KAuCl_4 , K_2PtCl_4 , K_2PtBr_4 , $\text{K}_2\text{Pt}(\text{CN})_4$, HgCl_2 , $\text{K}_3\text{UO}_2\text{F}_5$, $\text{UO}_2(\text{CH}_3\text{CO}_2)_2$, $\text{Pb}(\text{NO}_3)_2$ and $\text{Pb}(\text{CH}_3\text{CO}_2)_2$ were all soaked into preformed BPSL1204-MA crystals in 9% MPD and 0.1M NaCH_3CO_2 at pH5.0 to a concentration of $\sim 5 \text{ mM}$. From these, attempts were made to collect anomalous data at DSLs over a period of several months.

Due to the abundant supply of native crystals and the delicate nature of an individual rod, soaks were carried out by addition to the mother-drop to reduce the amount of physical disturbance that could destroy a target-crystal. 10 mM heavy-atom compound was incorporated into a fresh solution of reservoir solution, and a 1 μl drop was placed on a cover-slip about 1 mm away from an approximately 1 μl mother-drop containing preformed crystals. The two drops were then gently bridged using an acupuncture needle to allow the gentle diffusion of the heavy-atom compound into the crystal to a final concentration of 5 mM. The heavy-atom-containing drop was placed on the opposite side of the mother-drop from the best crystals whenever possible, to reduce the 'shock' of breaking surface tension when the drops were gently bridged together. The cover-slip was then resealed over its reservoir and left for 30-40 minutes to soak in.

Crystals were then looped from the mother-drop into a pre-determined cryo-buffer containing 5 mM heavy-atom for two minutes, to cryo-protect and prevent back soaking-out of non-covalent incorporated compounds, before mounting on a goniometer head held in a cryo-stream, for initial diffraction tests and saving into liquid nitrogen for transportation to a synchrotron.

5.6.2 Overview of Collection of MAD data and Anomalous Signal Detection

The detailed collection and processing of MAD data will be discussed for the collection of data at the selenium edge in Section 5.8. In general, a 200° dataset with an oscillation angle of 1° was collected at the peak and inflection energy of the L-3 edge of each potential heavy-atom in the derivative, with a high-energy remote collected at +150 eV of the peak energy. The edges of each heavy-atom were determined using a Vortex fluorescence detector at DSLs. Datasets were then processed through iMOSFLM and Scala programs to be assessed by the ShelXC/D/E programs for the presence of anomalous signal.

Of all the 10 heavy-atom soaks performed, only $\text{Pb}(\text{NO}_3)_2$ gave a very weak anomalous signal at its peak energy, assessed by a ShelXD d'' of 0.94 at a resolution of 4.5 Å. However when the resulting calculated, 'phased' density .mtz files of both possible enantiomorphs were loaded into Coot (Emsley & Cowtan, 2004), there was no properly phased density as judged by the lack of obvious solvent-boundaries and lack of distinguishable secondary-structural features such as α -helices. It was decided to solve the phase problem by the incorporation of selenomethione directly into the polypeptide of BPSL1204-MA.

5.7 Crystallisation of a Selenomethionine Derivative

5.7.1 Preparation of Protein Samples

Incorporation of seleno-met into BPSL1204-MA was carried out using the media and growth protocol described in sections 3.5.2 and 3.5.3 respectively. Seleno-met BPSL1204-MA was purified using the same methodology as that used for native BPSL1204-MA, as described in 4.5.1. After the Viva-spin concentration step as described in section 5.1.2, it was possible to achieve 30 mg/ml in 100 μ l of 10 mM Tris pH 8.0 as determined by Bradford Assay. 100% incorporation of seleno-met was confirmed by mass spectrometry, as previously shown in Figure 4.11. This revealed that an extra N-terminal seleno-met was present in the protein; presumably due to the presence of seleno-methionine rather than methionine preventing the action of an *E. coli* aminopeptidase that post-translationally removed the native sulphur-methionine (Ben-Bassat et al, 1987).

5.7.2 Attempts to Repeat Native Conditions

A hanging-drop screen of 8-13% MPD and 0.1 M sodium acetate pH 5.0 was set up using 1:1 μ l drops over a 1 ml reservoir using 30, 20 and 10 mg/ml seleno-met BPSL1204-MA protein samples. After 2 weeks of incubation at 17°C with bi-weekly investigation it was seen that no nucleation had occurred, the drops remaining clear. This created a problem because seleno-methionine BPSL1204-MA crystals were required for structure solution, and preferably they would be isomorphous to the native crystals that had already been processed to high-resolution. It was proposed that this inability to grow in the native conditions might be due to the extra N-terminal methionine present in the seleno-met derivative.

5.7.3 Original Streak-Seeding Experiments

In an attempt to avoid the potentially lengthy process of re-screening for crystallisation conditions, seeding experiments were attempted using native crystals in the native growth conditions. A 24-well hanging-drop plate was set up

with 8 multiplicates of 9, 10 and 11% MPD in 0.1 M sodium acetate pH 5.0, with each 2 μ l drop consisting of 1:1 20 mg/ml seleno-met protein:reservoir over a 1 ml reservoir. An autoclaved cats-whisker was stroked across the surface of a preformed native crystal from the relevant native condition, and then streaked through the each of the 8 corresponding drops containing selenomethionine-incorporated protein in succession without returning to the original 'donor' drop so that each would receive less nucleation than the last. Each dilution was then sealed and incubated at 17°C. Due to the supplement of preformed nucleation points transferred by the serrations of the cats-whisker (Guerin et al, 2003), the screen was checked daily due to the expectation of faster crystallisation. After 2 days numerous, condensed groups of rods, like hay-bales, appeared in all drops, but all were useless for data collection purposes in terms of size and number, as shown in Figure 5.4. It was considered that too high a concentration of nucleation points was being used, and that dilution by streak-seeding was not robust enough a method.

5.7.4 Serial Dilution of Seeding

A process of serial dilution of nucleation points achieved crystals of seleno-met BPSL1204-MA in the same form as the native protein. An entire native-drop of 9% MPD/0.1 M sodium acetate pH 5.0 mother-liquor containing several rod clusters of native BPSL1204-MA was repeatedly pipetted up and down to shatter the crystals into fragments. Then the entire drop (\sim 1 μ l after water-vapour equilibration) was transferred to a sterile Eppendorf containing 999 μ l of fresh mother-liquor and was vortexed to further shatter the transferred crystal fragments.

From this x1000-diluted solution of native BPSL1204-MA fragments (nucleation 'seeds'), 1 μ l was transferred to 999 μ l of fresh mother-liquor and vortexed again. This process was repeated eight times to leave a range of nucleation-point dilutions from 1×10^{-3} to 1×10^{-24} in decreases of 1×10^{-3} . A screen was then set up using fresh, 1 ml 9% MPD/0.1 M sodium acetate pH 5.0 mother-liquor reservoirs but instead of mixing 1:1 μ l protein:reservoir, 1 μ l 20 mg/ml seleno-met protein

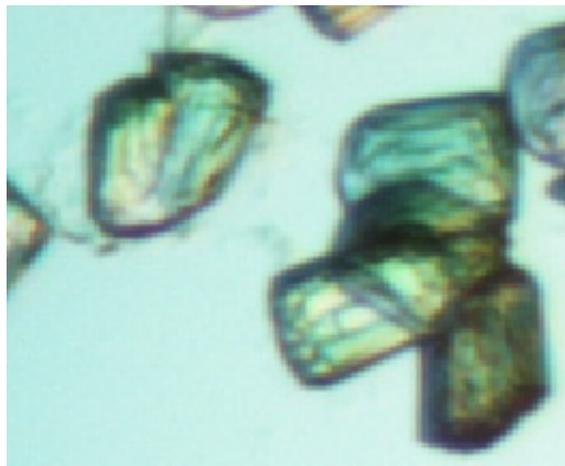
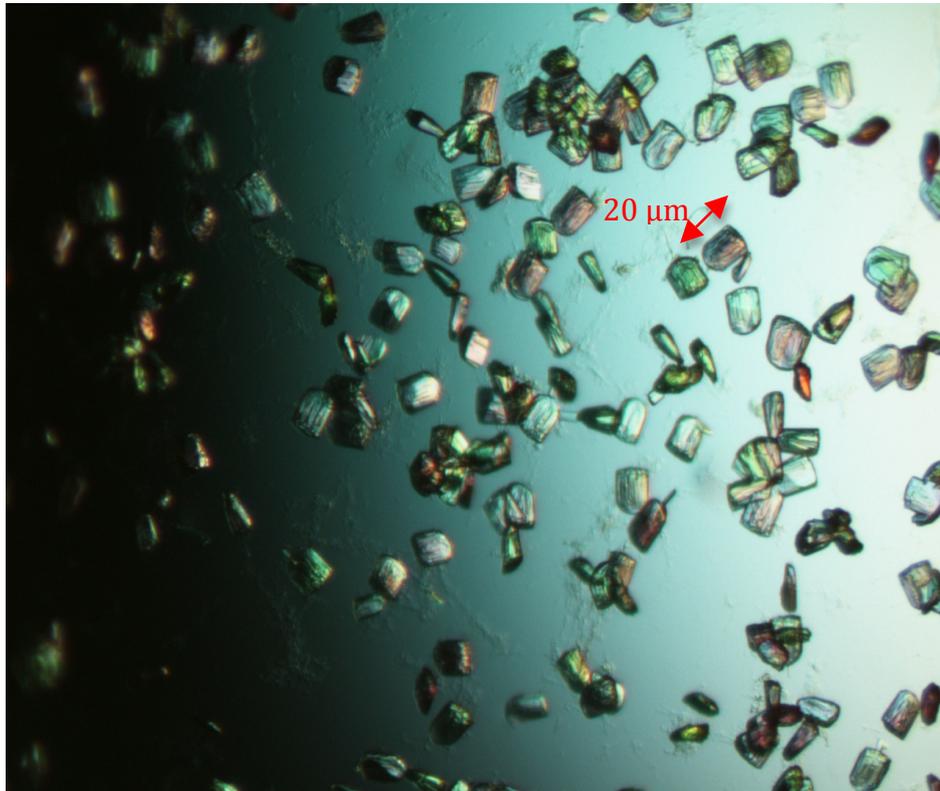


Figure 5.4: Streak-seeded seleno-met crystals. *Top:* numerous small crystals generated even after streaking through 24 2 μ l drops. *Bottom:* Closer inspection of crystals revealed them to be clusters of very small rods.

was mixed with 1 μl of a dilution of native BPSL1204-MA fragments in mother-liquor. Each drop was performed in triplicate, sealed and left to incubate at 17°C. After 2 days, crystals of the same morphology as the native ones grew in the dilution range of 1×10^{-9} - 1×10^{-15} (Figure 5.5). 1×10^{-3} and 1×10^{-6} gave similar results to streak-seeding experiments; whilst dilution of native fragments beyond 1×10^{-18} must have been diluted too far as to practically say there were no nucleation points transferred.

5.7.5 Confirmation of Crystallised Seleno-met Protein

Before collection of MAD data using a synchrotron beam-line, confirmation that what had crystallised was seleno-met BPSL1204-MA was carried out to prevent the wasting of synchrotron beamtime. A seleno-met crystal similar in size and shape to that from which the native dataset was collected was carefully looped from its mother-drop and transferred delicately to a 5 μl drop of fresh mother-liquor to 'wash' it. This was repeated 3 times, using a clean loop each time, to ensure that the final crystal was separated from any non-crystallised material or native fragments that might have been left in the original drop. The final, 'clean' crystal was then looped and dissolved into 5 μl of sterile, autoclaved, deionised water and was sent for analysis by mass spectrometry. The resulting spectra was identical to that of pure seleno-met BPSL1204-MA in section 4.5.2, suggesting that with the exception of the native nucleation-point (which likely formed the core of the rod-clusters) seleno-met BPSL1204-MA crystals of 100% selenium-incorporation had been successfully grown in the same crystal form as native, in spite of the extra N-terminal seleno-met.

5.8 Collection of MAD Data Using Synchrotron Light

5.8.1 Detection of Selenium Fluorescence and Determination of X-ray Wavelengths for MAD Data Collection

At DSLS, the robot within the beam-line experimental hutch carried out automated-mounting of crystals from the transport 'puck' in a nitrogen bath onto

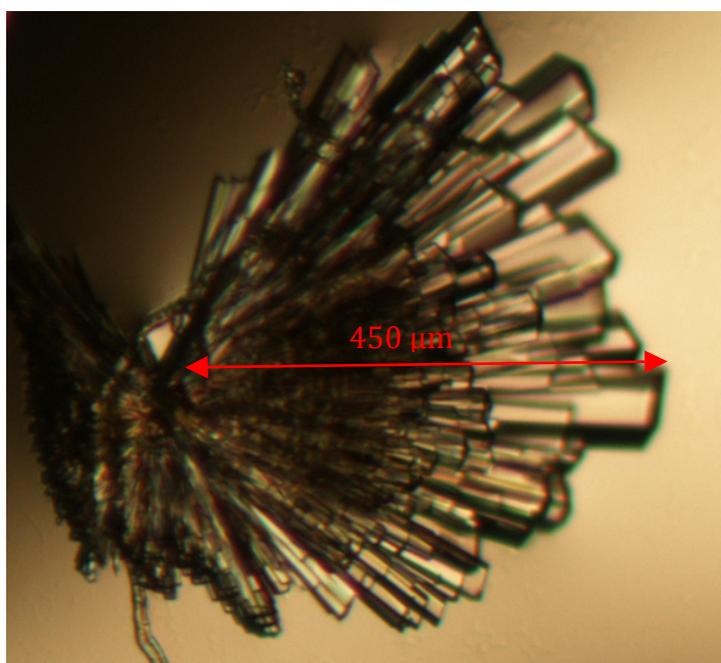


Figure 5.5: Crystals of seleno-met BPSL1204-MA similar to those of native protein. *Top:* 1×10^{-9} dilution-seeding. *Bottom:* 1×10^{-15} dilution-seeding from which MAD data was collected off an individual rod.

the goniometer and cryo-stream. A Vortex fluorescence detector was used for the determination of the emitting energy-edges, thus confirming again the presence of selenium within the lattice sub-structure and allowing for the determination of experimental wavelength energies around the L-3 edge (Figure 5.6).

The peak was determined to be of 12646.43 eV, therefore Peak data were collected just up of this value at 12648.0 eV to ensure peak anomalous signal was not missed. Inflection data were collected at 12644.0 eV, and High-Energy Remote data were collected well up of the Peak at 12796.4 eV in accordance with the theory presented in Section 2.6.2.

5.8.2 Data Collection of Peak, Inflection and High-Energy Remote Data

To check the isomorphous nature of the MAD data collected in comparison to the high-resolution native data already collected and processed, two test images of $\Delta\varphi = 1^\circ$ and 90° apart on φ were taken to estimate the resolution limit of the crystals and to calculate cell-parameters. With $a=99.71 \text{ \AA}$, $b=32.89 \text{ \AA}$, $c=48.34 \text{ \AA}$, $\alpha=90^\circ$, $\beta=115.89^\circ$ and $\gamma=90^\circ$ it was seen that a very similar unit-cell in C2 was present as discovered in the native data presented in section 5.4.4 (Table 5.2 in Section 5.10.4), leading to an assumption of isomorphism. Resolution using a transmission of 20% and 0.5 second exposure reached approximately 1.8 \AA in the test images.

Due to the radiation-damage that could potentially ensue from exposing the crystals to synchrotron radiation, combined with the need to collect three full datasets from the same crystal with the added radiation-damage of anomalous, heavy-atom scattering (Garman & Nave, 2002) it was decided not to push transmission or exposure time further to increase resolution. Instead, upon the survival of the crystal post-MAD data collection, a higher resolution Remote data set was collected, to complement the extension of resolution by phase-extending MAD data with the 1.3 \AA native dataset during processing data.

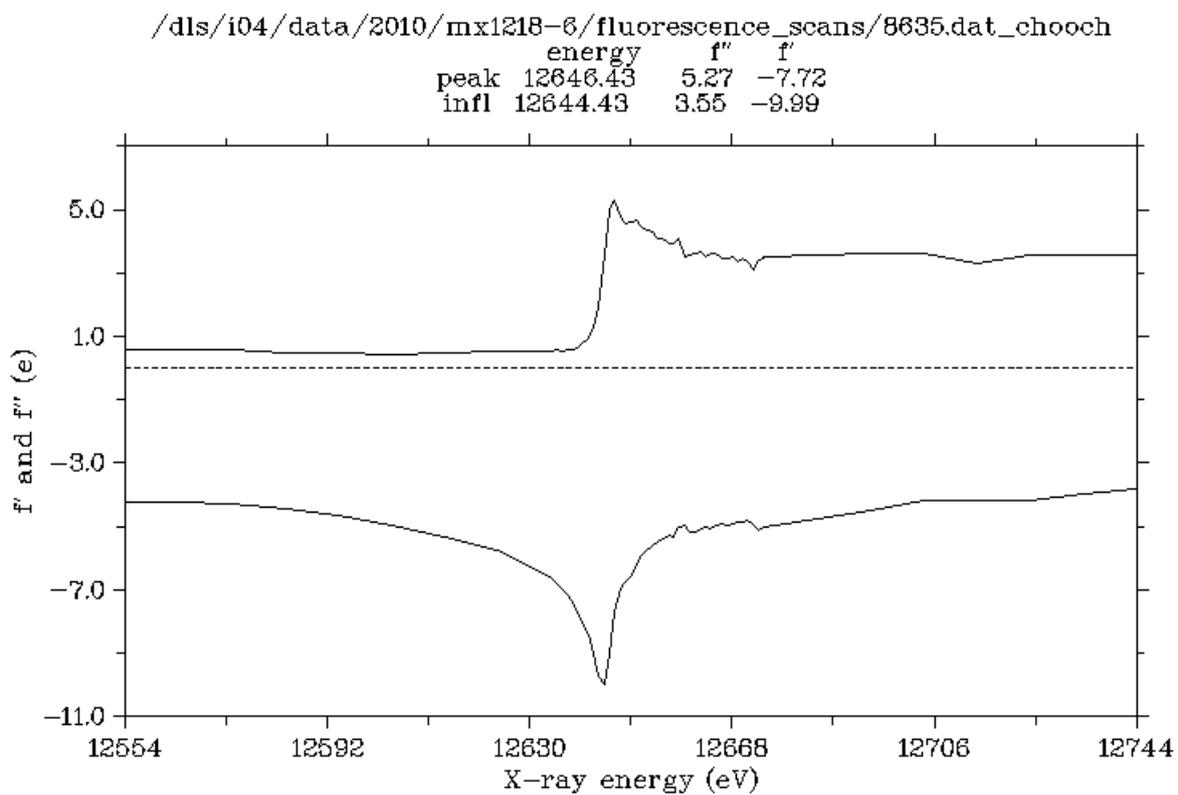


Figure 5.6: Determination of the seleno-met BPSL1204-MA peak (f'') and inflection (f') points. The image shows the CHOOCH output from selenium fluorescence at the L-3 edge, detected by a Vortex detector at Diamond Synchrotron Light Source.

A 200-image dataset with an oscillation angle of $\Delta\varphi = 1^\circ$ per image was collected using 30% beam transmission at the chosen Peak energy first with the detector distance set to a 1.5 Å limit, followed by re-attenuation of the beam to Inflection energy data collection. Then the beam was attenuated to the High-Energy Remote energy for a dataset to be collected with the same experimental parameters as the Peak and Inflection, followed by a repeated dataset collected with a 1.2 Å resolution detector limit at 60% transmission for 0.5-second exposure. This 2nd high-energy remote gave indexable diffraction with no visible radiation-damage to 1.4 Å.

5.9 Processing of MAD Data for Phasing

5.9.1 Indexing, Integrating and Scaling Measured Reflections

Each dataset was processed to the point of obtaining a scaled and merged .sca file of intensities as described for the native data in Section 5.5.2 and 5.5.3. Very little change was made to the original resolution limits, as estimated by eye and post-integration, as all data gave acceptable $I/(\sigma I)$, %completeness and high-resolution R_{merge} (Table 5.2 in Section 5.10.4) meaning there was now Peak and Inflection data cut to 1.8 Å, and High-Energy Remote data to cut to 1.4 Å. These could be used for the detection of anomalous scattering signal and the calculation of the solution to the phase-problem. The 1.3 Å native data were saved for phase-extension later on as part of structural refinement.

5.9.2 Detection of Anomalous Signal and Phasing

The scaled .sca files for each of the MAD datasets were loaded into ShelXC (Sheldrick, 2008); part of the ShelxC/D/E suite of phasing programs ran through the HKL2MAP platform. ShelxC then performed the necessary preparations for phase calculations, such as creation of ΔF and F_a statistics. A .hkl file of 'native' ΔF s and a _fa.hkl file of anomalous F_a 's were output. The d'' statistic that gives an indication of anomalous signal if >0.8 gave signals of: Peak=1.3, Inflection=0.9 and High-Energy Remote=0.98 for data to 2.4 Å resolution (Figure 5.7a), thus

9842 Unique reflections, highest resolution 1.988 Angstroms

Resl.	Inf	- 8.0	- 6.0	- 5.0	- 4.0	- 3.5	- 3.0	- 2.8	- 2.6	- 2.4	- 2.2	- 1.99
N(data)	171	221	273	602	615	1065	665	876	1187	1664	2503	
Chi-sq	0.23	0.76	0.82	1.44	1.90	0.76	0.58	0.80	0.62	1.07	1.02	
<I/sig>	21.8	22.3	20.6	22.0	20.9	19.4	18.1	16.2	14.7	13.8	10.7	
%Complete	100.0	100.0	100.0	99.7	99.8	100.0	98.5	99.7	99.4	99.9	96.7	
<d"/sig>	2.09	2.33	1.68	1.22	1.21	1.30	1.31	1.30	1.13	1.05	0.96	
CC(anom)	82.6	83.6	75.7	27.3	31.8	62.1	53.9	52.6	39.1	36.6	19.7	

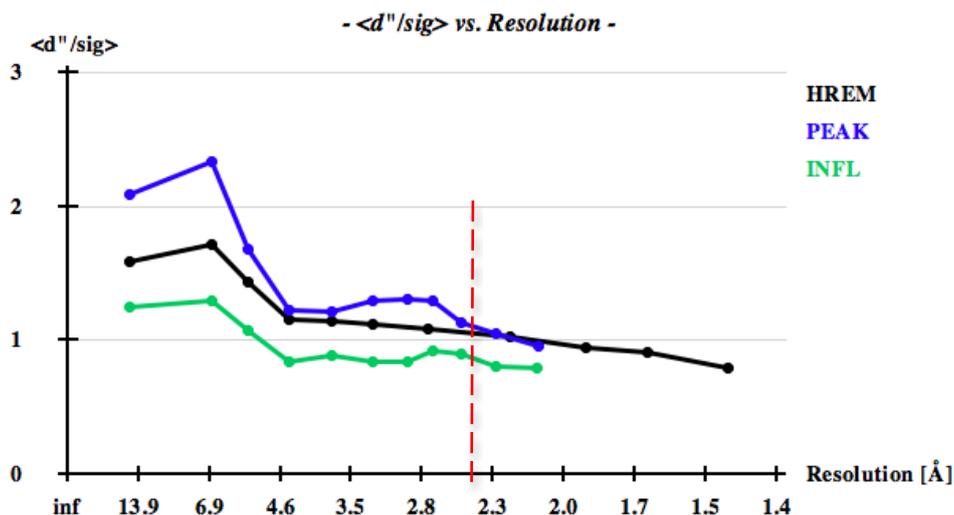


Figure 5.7a: Output from ShelxC. *Top*: ShelxC statistics for peak data. *Bottom*: ShelxC plot of signal Vs resolution. Anomalous signal was calculated to be of significance in the peak MAD dataset to a resolution of 2.4 Å with a d'' of 1.13

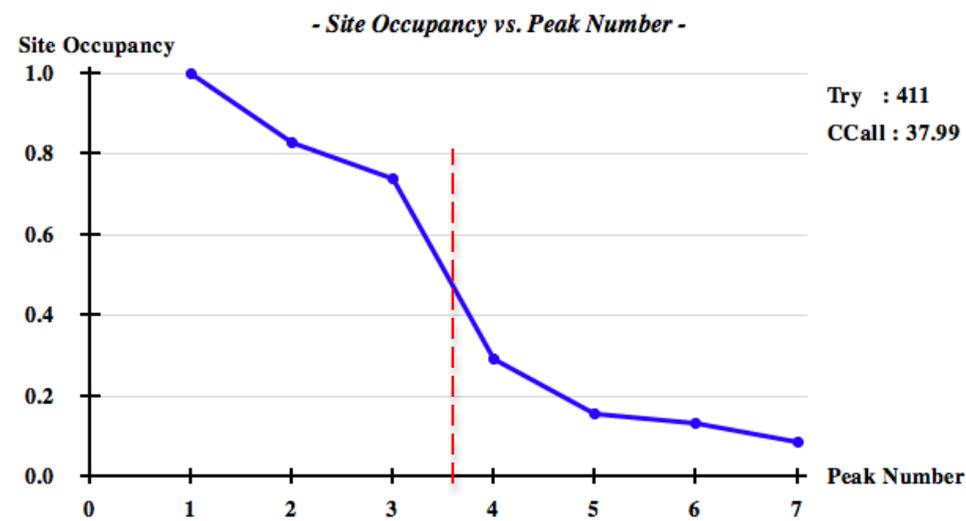


Figure 5.7b: Output from ShelxD. Plot of calculated heavy-atom site occupancies Vs heavy atom peaks. It was decided that 3 heavy atom sites were significantly occupied, described by the peaks found prior to a significant drop in occupancy.

giving confidence to the successful next step of calculating phases for the modification of electron density.

The `_fa.hkl` file was then input to ShelXD (Schneider & Sheldrick, 2002) for the location of heavy atom sites for phasing using maximum resolution of 2.4 Å, limiting the search to 500 tries. Three potential sites were found (Figure 5.7b), which combined with the Matthews estimation of one molecule in the unit cell with 40.7% solvent (which matched the native Matthews in Section 5.5.3) seemed acceptable if for 3/5 seleno-methionines were scattering and the N-terminal methionine and one internal methionine were in a state of disorder. The heavy-atom sites were then output as a `_fa.pdb`.

The `.hkl` and `_fa.hkl` files were then input to ShelXE (Sheldrick, 2008) for phasing using the identified anomalous signal and then the refinement of electron density using the new phase information. The density was phased and refined for 20 cycles, using a ShelX-calculated solvent content that matched that of the independently carried out Matthews analysis. Phasing was carried out for the substructures of both potential enantiomorphs of the heavy atom substructure. Both output `.phs` files were then assessed for the most likely enantiomorphic solution by identifying which file contained the largest pseudo-free Correlation Coefficient (pseudo-freeCC), with a value above 70% usually indicating an interpretable map (Figure 5.8). The pseudo-freeCC is a figure of merit calculated by the removal of a subset of 10% of random reflections, which are then independently density modified for one cycle and back-Fourier transformed for comparison to every Nth cycle of density modification, as an evaluation of phase improvement (Pal et al, 2009; Sheldrick, 2010). Ultimately the output `.pdb` coordinate file was loaded into the Coot electron-density modification program (Emsley & Cowtan, 2004) along with each enantiomorphs `.phs` or `_I.phs` (original and inverted hand respectively, converted to `.mtz` files using `Convert_to_MTZ` in CCP4i) file to investigate whether solvent-boundaries and secondary- or tertiary-structures were clearly visible in the calculated electron-density. The inverted enantiomorph gave the best statistics and the best-defined density into which model building could begin (Figure 5.9).

Pseudo-free CC = 52.77 %

Density (in map sigma units) at input heavy atom sites

Site	x	y	z	occ*Z	density
1	0.8839	0.0176	0.2575	34.0000	63.74
2	0.9097	-0.4718	0.0290	28.1010	48.26
3	0.9043	-0.9048	-0.2366	25.1566	40.21
4	0.9344	-0.6878	-0.4127	9.9654	14.56
5	0.8704	0.0111	0.2825	5.2938	6.44
6	0.8673	-0.9105	-0.2472	4.5594	6.39
7	0.8906	-0.4005	-0.4792	2.9206	3.34

Pseudo-free CC = 78.48 %

Density (in map sigma units) at input heavy atom sites

Site	x	y	z	occ*Z	density
1	0.1161	0.9824	0.7425	34.0000	68.18
2	0.0903	1.4718	0.9710	28.1010	52.39
3	0.0957	1.9048	1.2366	25.1566	50.63
4	0.0656	1.6878	1.4127	9.9654	18.40
5	0.1296	0.9889	0.7175	5.2938	4.66
6	0.1327	1.9105	1.2472	4.5594	2.82
7	0.1094	1.4005	1.4792	2.9206	0.76

Figure 5.8: ShelXE heavy-atom site density calculations and pseudo-freeCCs of calculated electron density map enantiomorphs. *Top:* Original enantiomorph. *Bottom:* Inverted enantiomorph.

Assessment of phased density modification statistics prior to the examination of calculated electron density maps suggested that the high pseudo-freeCC (>70%) and greater density for refined heavy atom sites that were calculated for the Inverted enantiomorph made it the most likely candidate for correctly phased density.

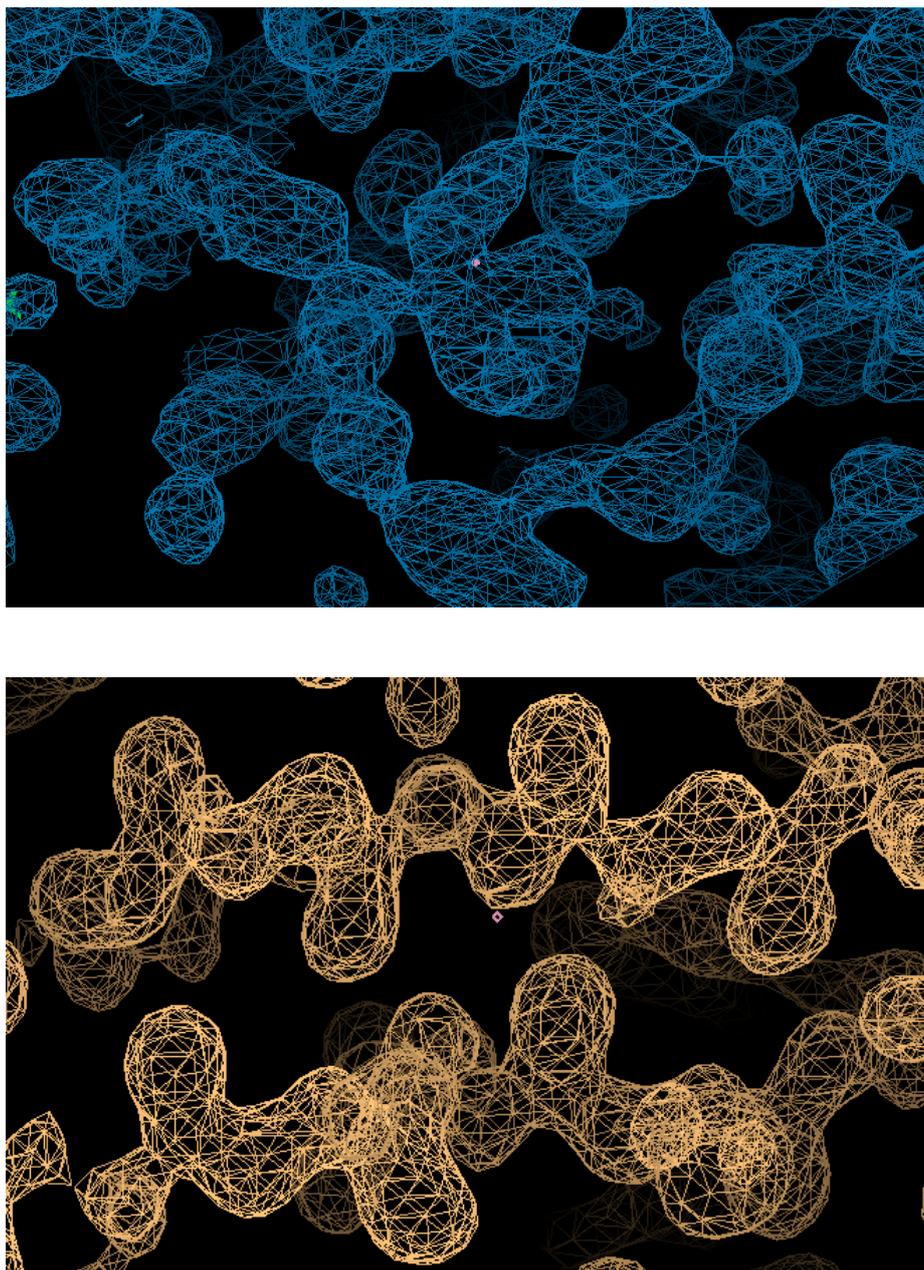


Figure 5.9: Investigation of the the .phs outputs from ShelxE in Coot (at a map sigma of 1.0). *Top:* Original enantiomorph density. *Bottom:* Inverted enantiomorph density.

The original density shows limited connectivity, with little evidence of tertiary structure. The inverted enantiomorph shows clear connectivity, and tertiary-elements such as the β -sheet shown are prominent. As suggested by the statistical outputs, the inverted density was the best-phased map.

5.9.3 Automatic Model Building

The phased electron density output from ShelX proved to be very clear and well defined, to 2.4 Å resolution. Without any model building yet undertaken, it was easy to recognise features of primary-, secondary- and tertiary protein structure, for example at a σ -level of 0.98 the individual atoms of a tyrosine ring were clearly visible (Figure 5.10). Therefore the data seemed a perfect candidate for automated model building. Using the CCP4i ARP/wARP 7.0.1 program (Langer et al, 2008), supplied with the expected amino-acid sequence for BPSL1204-MA in .pir format, automated model-building using the Maximum Likelihood target function for 50 cycles of build-and-refine were run to look for one molecule of 147 amino-acids per asymmetric unit.

ARP/wARP successfully modelled 100 residues of the expected 147 residues into the density. Initial investigation showed generally very good fitting, and the presence of disorder in the first and last 5 residues, accounting for ~7% of the atoms in the polypeptide. Disregarding the disordered residues meant an overall fitting of ~73% of the possible sequence, which was considered a very good start success. Phases were extended to 1.3 Å resolution native data using the CCP4i program Convert to/Modify/Extend MTZ program, applying the FreeR dataset used for the 2.4 Å resolution seleno-met data-processing, with refinement carried out using Refmac5 (Murshudov et al, 1997).

5.10 Refinement of the BPSL1204 Structure

5.10.1 Fitting of the Complete Model

Even though ARP/wARP generated a fairly complete model, it was still required to be manually investigated to find the missing 37 residues and later for correction of inconsistencies such as impossible stereochemical angles, unrealistic chemical bonds and the placing of true but unlikely residue conformations that may have been incorrectly orientated. A typical example was the rotation of the C γ of Asn⁸⁵ in order to align the nitrogen to ensure chemically

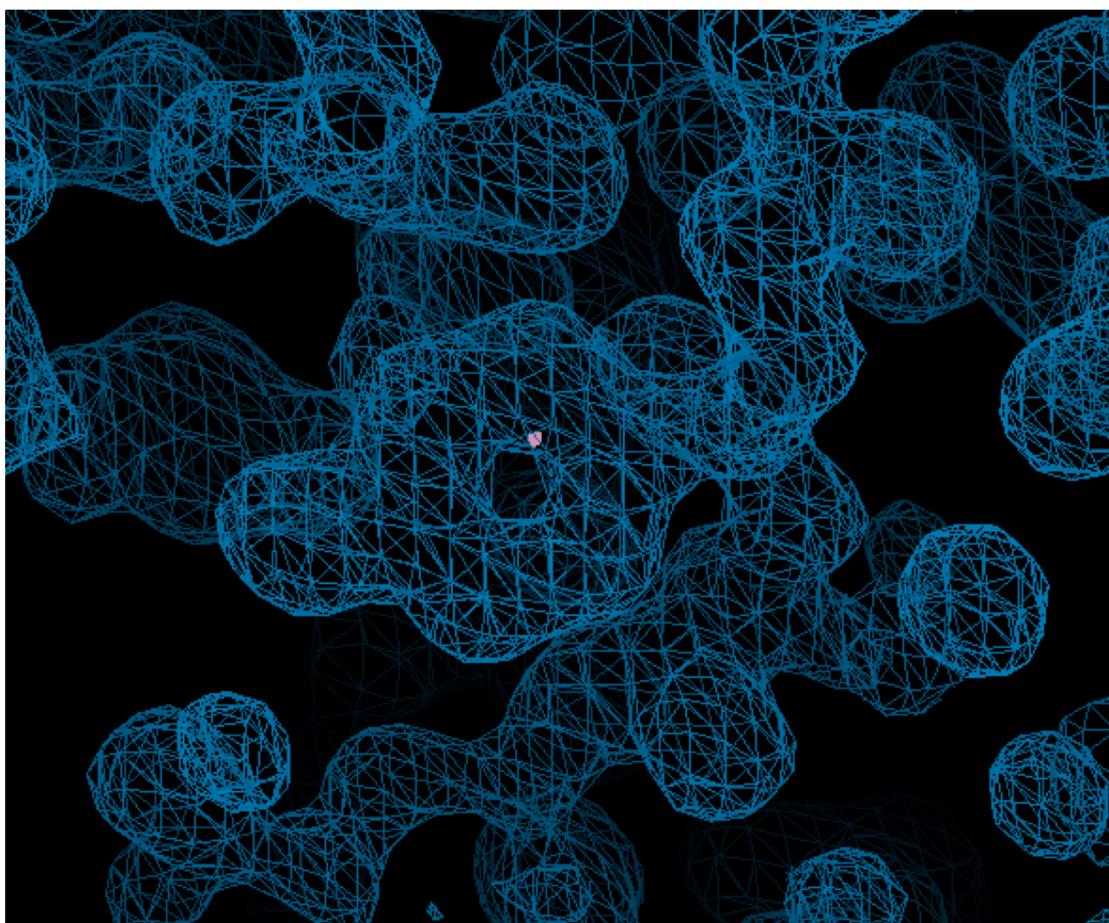


Figure 5.10: Electron density prior to automated model building at a map sigma level of 0.98. All the atoms of a tyrosine residue are easily seen without fitting of a model, providing a good candidate electron density map for automated model building by ARP/wARP.

viable hydrogen bonding was taking place with the main-chain oxygen of Asp⁸¹ and the oxygen of MPD¹ (Figure 5.11). This was in addition to the manual fitting of the remaining 37 residues unaccounted for by the automated modelling and the correct alignment of all residues centrally within the lumen of connected electron density. Manipulation of the model built into the calculated electron density was carried out using the graphics program Coot 0.5.2 (Emsley & Cowtan, 2004) and refined using Refmac5.

5.10.2 Preventing Bias in the Model

ARP/wARP built a model on the basis of placing 'dummy' atoms into regions of density and then attempted to match them up into side-chains and polypeptide-backbones. Therefore certain errors were found upon manual inspection, commonly concerning aromatic side-chains, where density from the automated, refined density did not match the obvious aromatic-ring density that had been seen in the experimentally observed electron density prior to automated modelling and refinement, such as Trp⁷⁷ and Arg³⁴ shown in Figure 5.12. Therefore as part of the refinement of the model, these regions of imposed data-bias were manually corrected in Coot.

5.10.3 Fitting of Ligands and Water Molecules

After the process of correcting and refining regions of ambiguity in the ARP/wARP model and the manual modelling of regions of unmodelled, connected electron density, the modelling of other regions of unexplained density was carried out. Using the difference map after refining with the experimental model, Coot identified 'blobs' of true electron density. These were then modelled as either water, or networks of water, if blobs were present as single spheres with sensible hydrogen-bonding distances of 1.9-2.5Å, or as potential ligands. Ligands expected were the contents of the mother-liquor (MPD or sodium acetate) and the glycerol soaked in during cryo-protection.

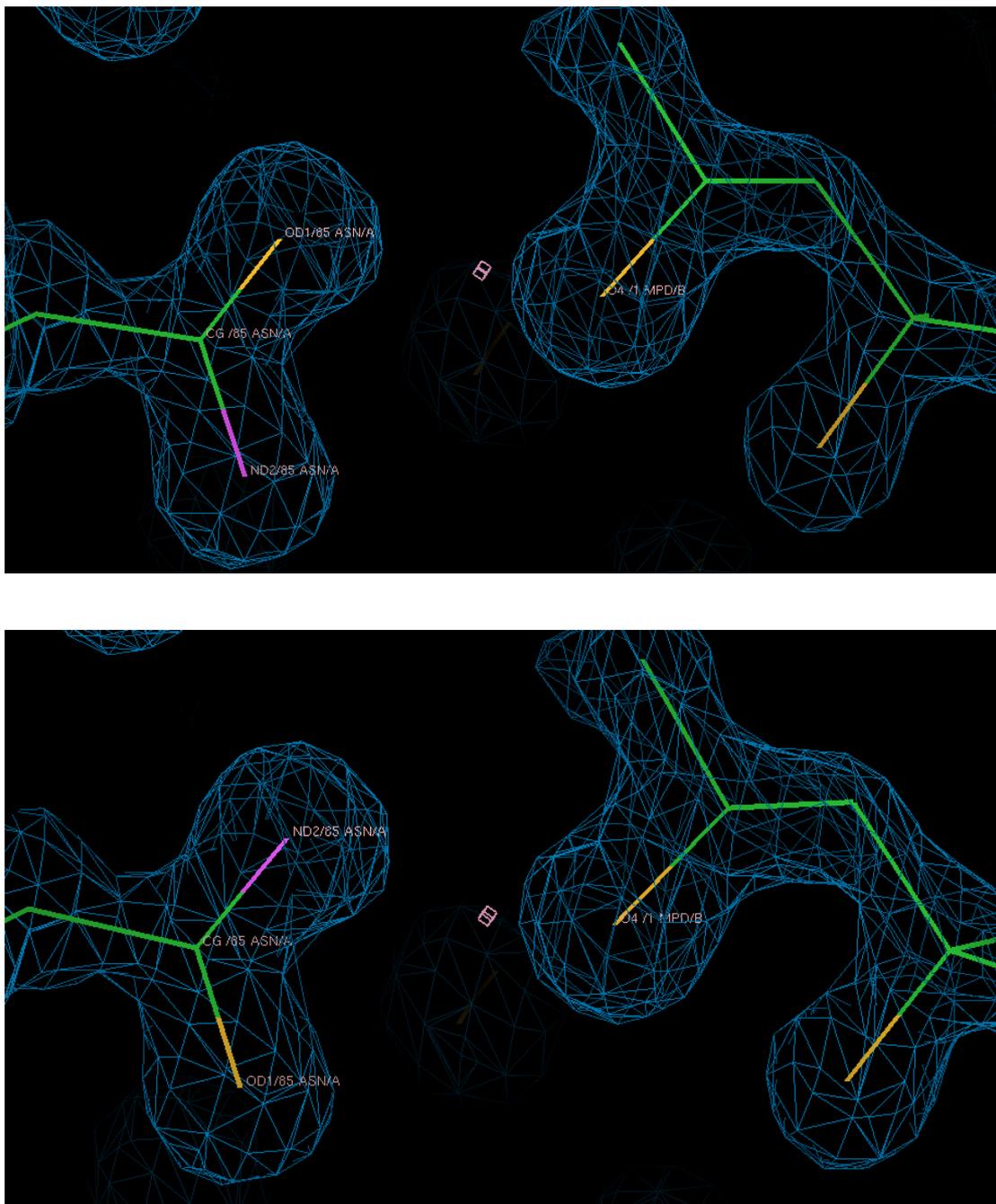


Figure 5.11: Correction of ARP/wARP model building. Top: original fitting. Bottom: Corrected for chemical viability. Residues with high rotational ability (such as Asn⁸⁵) pictured were manually investigated to correct hydrogen bonding.

In this example, a 180° rotation around C_γ was required to align the nitrogen (pink) of Asn⁸⁵ with the oxygen (yellow) of MPD¹ to fulfil correct hydrogen bond potential.

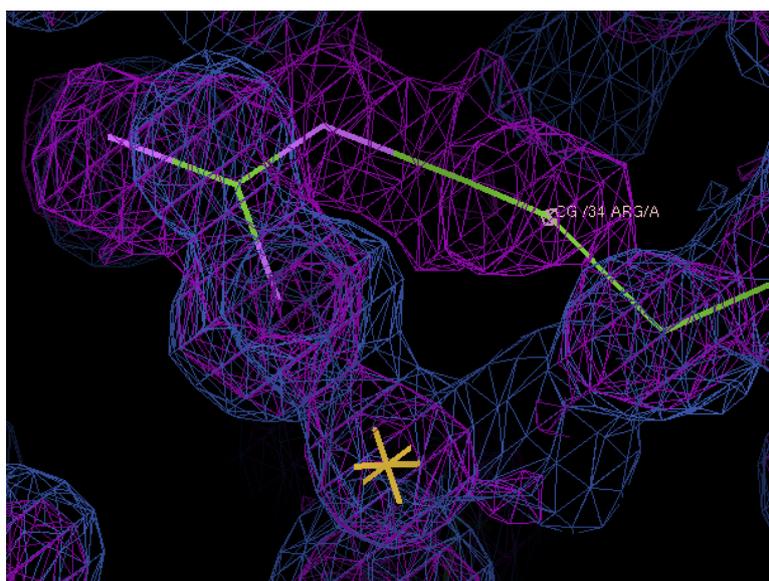
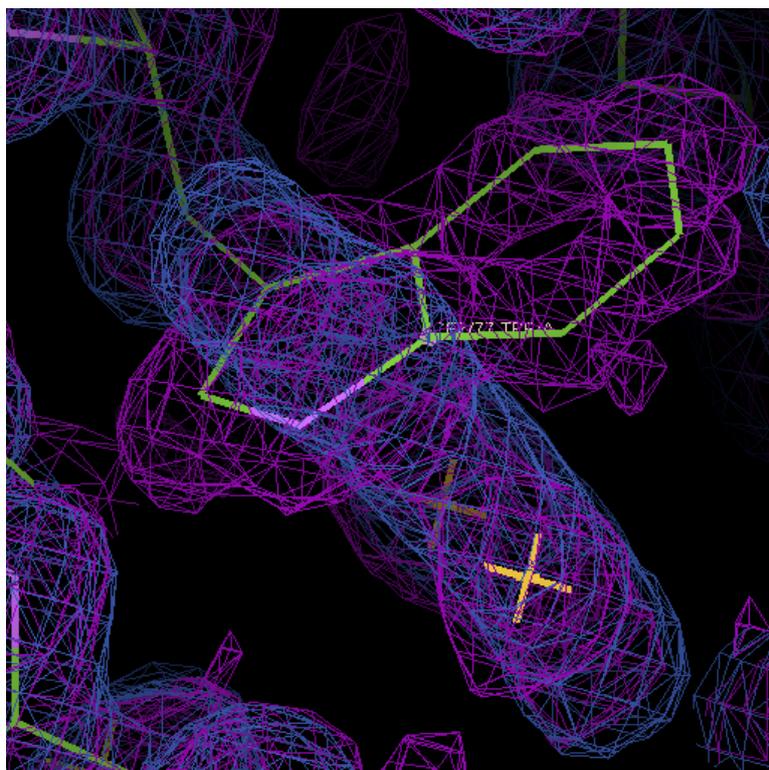


Figure 5.12: Model bias induced by automated ARP/wARP building. *Top:* Trp⁷⁷. *Bottom:* Arg³⁴. The density altered by refinement against the automated model (purple) occasionally induced incorrect orientations of side-chains when compared to the experimentally observed density (blue). These were corrected before further refinement. The yellow 'star' represents an unconnected 'dummy' atom placed into density by ARP/wARP.

During this process, unmodelled density was also investigated in case it revealed possible biological ligands that might help with the elucidation of BPSL1204 function. This however was not the case. Alongside the auto-fitting of 100 now corrected residues and the 37 residues modelled in manually, 1 molecule of MPD per protein molecule was modelled into 'blobs' alongside 179 water molecules. This was quite a low number of solvent molecules for high resolution data, with a residue:water ratio of 1:1.3. On average, a ratio of 1:1 is common for data to 2 Å resolution, however when using data to approximately 1 Å resolution, a 1:1.6 ratio is more common (Carugo & Bordo, 1999).

5.10.4 Phase Extension to 1.05 Å Resolution

At a later date from the collection of MAD data, native crystals were re-grown and data to 1.05 Å resolution were collected at 80% synchrotron beam-transmission at 12,658 keV on the Diamond I02 beamline. Cell dimensions of $a=99.76$ Å, $b=32.82$ Å, $c=48.28$ Å, $\alpha=90^\circ$, $\beta=116.11^\circ$ and $\gamma=90^\circ$ were calculated, which were very similar to those of the MAD and original native data. Using Convert/Modify/ExtendMTZ in CCP4i, phases from the 1.3 Å resolution data were extended to 1.05 Å resolution after data indexing and scaling, with the FreeR dataset from the original 2.4 Å resolution MAD data applied.

5.10.5 Refinement and R-factors

Refinement of the built model was carried out in Refmac5 (Murshudov et al, 1997). Each round of refinement carried out 10 cycles of maximum-likelihood-restrained refinement whilst allowing for unweighted, individual anisotropic B-factors. Isotropic B-factors assume that vibrational or thermal motion of an atom in a molecule can occur in every direction about a mean position (Tronrud, 2004). Anisotropic B-factors are a more realistic interpretation of the vibrational movement of a molecule at an atomic scale, when covalent bonds and torsion angles restrict movement to only a few directions. Anisotropic Displacement Parameters (ADPs) are only enforceable when the data used are of a sufficiently high resolution (commonly >1.2 Å) (Eyal et al, 2007), as was the case with the 1.05 Å resolution BPSL1204 data. B-factors were unweighted to reduce

restraints on movement and B-factor occupancy of individual atoms. This allowed for refinement of B-factors to a level best approximating reality, as the model was fitted to well defined, near-atomic-resolution experimental density.

A B-factor cut-off monitoring limit of 10.00σ was imposed, in order to identify in the output log-file any parts of the model exhibiting potentially excessive deviation in thermal motion between consecutive atoms, which could then be corrected in the model if considered invalid before further refinement. In addition, refinement of the model was carried out with 5% of the experimental data (free R-factor) not used in refinement as a control against model-bias (Brunger, 1993), which can lead to misleading density features. This would have biased the model away from the true structure.

An assessment of increasing, unbiased refinement was the reduction of *Rwork* and *Rfree* (Brunger, 1993). *Rwork* is a value that describes how well the built model predicts the actual experimental data, by comparing the structure factors calculated from the model (F_{calc}) with those from the original measured diffraction intensities (F_{obs}). *Rfree* consists of a subset (5% for the BPSL1204-MA structure described here) of data excluded from the refinement. Therefore the refinement of the model was judged to be effective as long as after each round of refinement both of the R-factors decreased in value as the model approached the best fit to the experimental density. Over ten rounds of refinement, *Rfree* and *Rwork* decreased from 0.3496 and 0.3458 to 0.1617 and 0.1390 respectively. Table 5.2 summarises all the crystal and refinement parameters calculated for the datasets used in the solution of the BPSL1204-MA crystal structure.

5.11 Structure Validation

5.11.1 Molprobity

As part of refinement and validation, a late-stage *Rfree*=0.1908 and *Rwork* 0.1667 .pdb file was input to the Molprobity server (Duke University) (Davis et al, 2007) for analysis of atomic contacts and clash-scores, with analysis of

Dataset	Native (1)	Native (2)	Seleno-Met (peak)
Space Group	C2	C2	C2
<i>Cell Dimensions</i>			
A (Å)	99.96	99.76	99.71
B (Å)	32.91	32.82	32.89
C (Å)	48.35	48.28	48.34
α (°)	90.00	90.00	90.00
β (°)	116.04	116.11	115.89
γ (°)	90.00	90.00	90.00
Resolution (Å)	20.8-1.29	30.8-1.05	45-2.4
Beamline	Diamond I02	Diamond I02	Diamond I04
Wavelength (Å)	0.932	0.979	0.98
Observed/unique refls	125464/35052	195595/65407	35268/9842
R _{merge} (outer shell)	3.5 (34.1)	4.7 (46.1)	3.3 (9.7)
Completeness (%) (outer shell)	90.4 (99.9)	99.6 (99.5)	99.8 (100.0)
Redundancy (outer shell)	3.6 (3.5)	3.0 (2.9)	3.6 (3.6)
I/ σ I (outer shell)	22.3 (3.5)	22 (1.6)	18.3 (6)
Heavy atom	-	-	Se
No. of sites	-	-	3
Solvent content (%)	-	40.82	-
Matthews coefficient (Å ³ /Da)	-	2.08	-
Final Refinement Statistics			
No. of reflections (working/free)	-	62118/3106	-
No. of atoms (protein/water/MPD)	-	2168/179/8	-
R _{work} /R _{free}	-	0.1390/0.1617	-
<i>RMS deviations from ideality:</i>			
Bond length (Å)	-	0.0330	-
Bond angles (°)	-	2.2163	-
<i>Average B-factors:</i>			
Overall (Å ²)	-	12.9	-
Main chain (Å ²)	-	10.9	-
Side chain (Å ²)	-	13.6	-
Water (Å ²)	-	22.2	-
<i>Ramachandran plot proportions:</i>			
Favoured/Allowed/Disallowed(%)	-	98.5/1.5/0.0	-

Table 5.2: Data collection, phasing and refinement statistics. Hydrogens are not included in these statistics.

geometry on a residue-to-residue basis highlighting bad rotamers, Ramachandran outliers and anomalous bond-lengths. This allowed further refinement of the structure to ensure accurate stereochemical representation of the experimental density in the model, including incorporation of 'riding' hydrogen atoms into the structure refinement.

Hydrogen positions were not directly refined into the model due to little observed density in the expected regions for hydrogens in protein, even at high resolution when electron density maps were contoured as low as 0.5σ . However, their relative riding positions were used in order to assess and refine bond lengths and angles. As in reality proteins do have hydrogens, the use of riding hydrogens during refinement brought the model closer to reality, measured as a decrease in R-factors. The Reduce program, run using Molprobity, added 1069 hydrogens to the model for validation (Chen et al, 2010). This was for stereochemical analysis of the structure using the logic that only a structure that leaves room for hydrogens can be accurate. Any collisions that occurred between hydrogens were incorporated into the clash-score.

Several further rounds of refinement were carried out to achieve the final structure. Clash-scores were reduced from 16.21 to 1.86 clashes with $>0.4\text{\AA}$ overlaps per 1000 atoms; poor rotamers were reduced from 5.08% to 0.87%; bad angles and bad bond-lengths were reduced from 1.46% and 3.65% respectively to 0% and the Molprobity score, where lower is better, was reduced from 2.30 to 0.95. This improved the structure enough to raise it from the 25th percentile to the 97th percentile of all structures assessed by Clash-score and the 99th percentile of Molprobity scores from structures of similar resolution (Figure 5.13). A multi-criterion chart was created for Coot, to facilitate the inspection of each residue that may have contributed to penalties. Study of those residues that added to scores such as poor rotamers and $C\beta$ deviations revealed that the model fit the density very well, indicative of true but chemically uncommon angles within the protein structure.

All-Atom Contacts	Clashscore, all atoms:	1.86	97 th percentile* (N=208, 1.05Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Poor rotamers	0.87%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	98.47%	Goal: >98%
	Cβ deviations >0.25Å	1	Goal: 0
	MolProbity score [^]	0.95	99 th percentile* (N=830, 1.05Å ± 0.25Å)
	Residues with bad bonds:	0.00%	Goal: 0%
	Residues with bad angles:	0.00%	Goal: <0.1%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

[^] MolProbity score is defined as the following: $0.42574 \cdot \log(1 + \text{clashscore}) + 0.32996 \cdot \log(1 + \max(0, \text{pctRotOut} - 1)) + 0.24979 \cdot \log(1 + \max(0, 100 - \text{pctRamFavored} - 2)) + 0.5$

Figure 5.13: Output from the Molprobity analysis (Davis et al, 2007) of the final BPSL1204 structure

5.11.2 PROCHECK

The structure was further validated by the PROCHECK program (Laskowski et al, 1993) run through CCP4i. This program also analysed bond-angles and bond-lengths of chi-angles, main-chain bonds, planar-groups and overall side-chain stereochemistry presented in the form of graphs and plots against standard, accepted values (Figure 5.14). A comprehensive Ramachandran plot to illustrate the validity of the overall structure was also created (Figure 5.15).

The Ramachandran plot (Ramachandran et al, 1963) generated revealed, as with the Ramchandran analysis of Molprobit, that 98.5% of residues occupied favoured steric angles, with only 1.5% occupying φ and ψ angles that were allowed but not ideal. There were assessed to be no residues occupying disallowed angles (Figure 5.15). Investigation of individual residue plots of χ -angles, main-chain and side-chain angles showed that the model fit to the density conformed to established stereochemistry. The bar chart output analysing deviation of χ -angles from the mean was used to locate possible problem residues and further refine them to improve the geometrical statistics.

The program SFCheck (Vaguine et al, 1999) was also run as part of the PROCHECK program, which analysed the relationship between the created atomic model and the original experimental data. In addition, an overall B-factor was calculated of 18 Å², suggesting that the model built represents a protein of reasonably high overall order, with low levels of disorder.

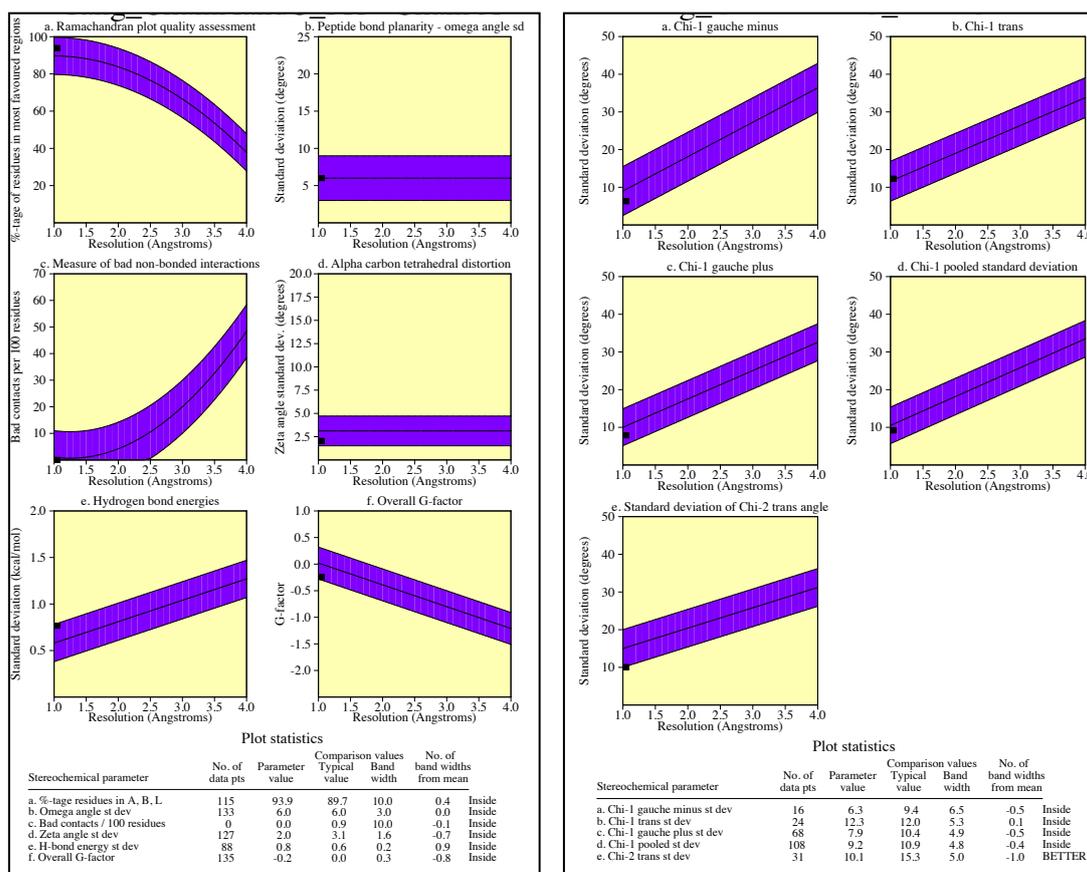


Figure 5.14: Procheck analyses (Laskowski et al, 1993). *Left:* Plots of main-chain features. *Right:* Plots of side-chain features. For the main-chain, all aspects (including peptide-bond planarity and hydrogen bond energies) fall within the accepted purple boundaries, which represent well-fit structures of similar resolution.

Side chain features such as chi-1 angles fall within the accepted purple boundaries. The measurement of chi-2 angle outliers are due to the uncommon but experimentally true rotamers

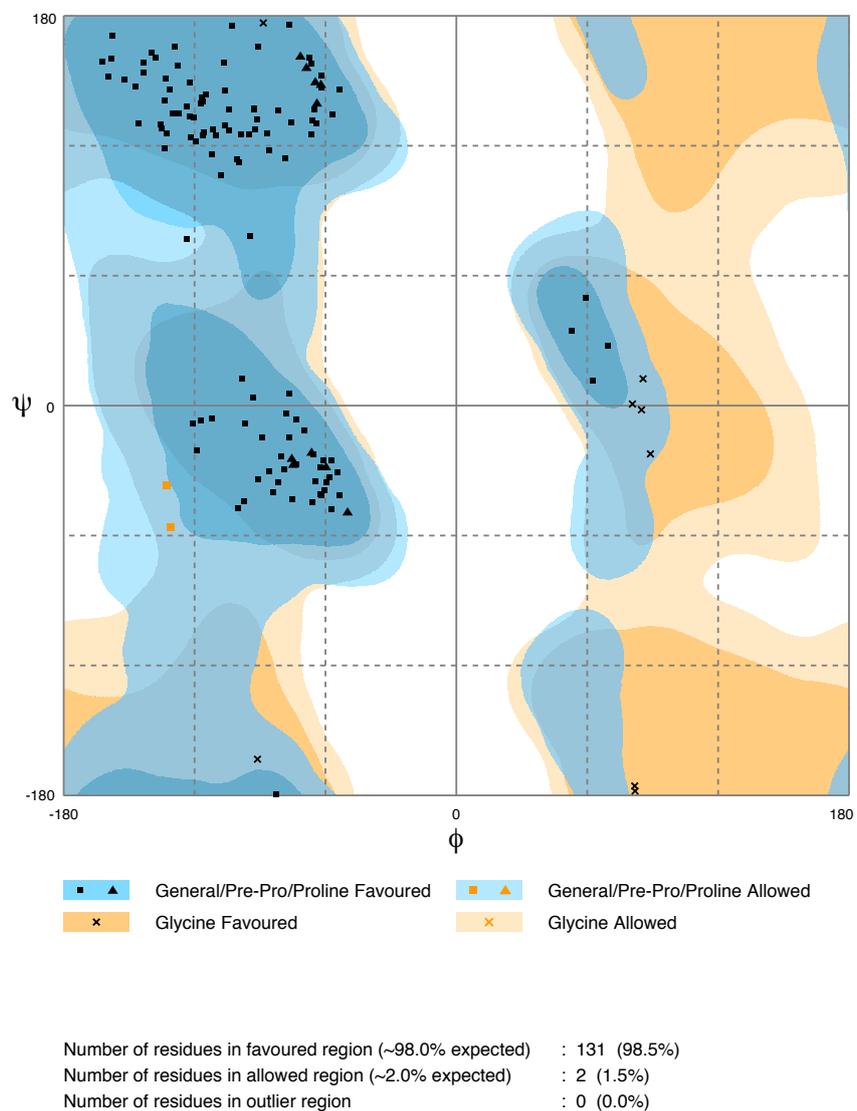


Figure 5.15: PROCHECK Ramachandran analysis (Laskowski et al, 1993). 98.5% of residues were shown to occupy favoured regions, with 1.5% in allowed regions. No residues occupied forbidden regions of steric hindrance.

Chapter 6: Structural Investigation of BPSL1204 Function

6.1 Structural Features of BPSL1204

The BPSL1204-MA crystal structure revealed a protein that consists of two domains connected by a short loop, an N-terminal domain consisting of one α -helix and a 4-stranded β -sheet and a C-terminal domain comprising two α -helices and a 4-stranded β -sheet (Figure 6.1). These domains appear structurally very similar to each other and are arranged stacked together so that the N-terminal domain's β -sheet is backed onto by the C-terminal domain's α -helices. Figure 6.2 shows the secondary structure arrangement and topology of BPSL1204. An 11-residue N-terminal 'linker' region (of which only 6 residues are sufficiently ordered to be modelled and refined) connects the first α -helix in the N-terminal domain to the point at which the lipid anchor was truncated (Ala²⁰), immediately after the LLAGCA glycerylcysteine lipoprotein motif.

During the building of the BPSL1204 structure, Met⁴² was seen to occupy multiple conformations when a model was built into the experimental electron density, in a roughly 7:3 ratio, as shown in Figure 6.3a. This residue was found to be on the extreme reach of the β 2- β 3 loop. Though the loop itself appears well ordered with low B-factors, the significant sulphur atom (or selenium in the original MAD data) was seen to have reasonably high B-factors (Figure 6.3b). This conformational flexibility without uniformity throughout the crystal is likely the reason for phase calculations being based upon only 3/5 of the possible methionines in the BPSL1204 sequence. The other un-phased methionine was the N-terminal Met¹, shown to be present via MS experiments (Section 4.5.2), and was thus disordered in the crystal lattice.

Investigation of the molecular interface between the two domains revealed several key interactions. Figure 6.4 shows these interaction networks in the graphical representation in Pymol. Several hydrogen bonds create the region of the interface furthest away from the connective loop L1, namely Tyr²⁶OH-Glu²⁴OE1, Glu²⁴OE1-Arg⁸⁸N and Arg³⁴NH1-Val⁸⁴O. Several networks of water molecules also support these interactions. These types of interactions are typical

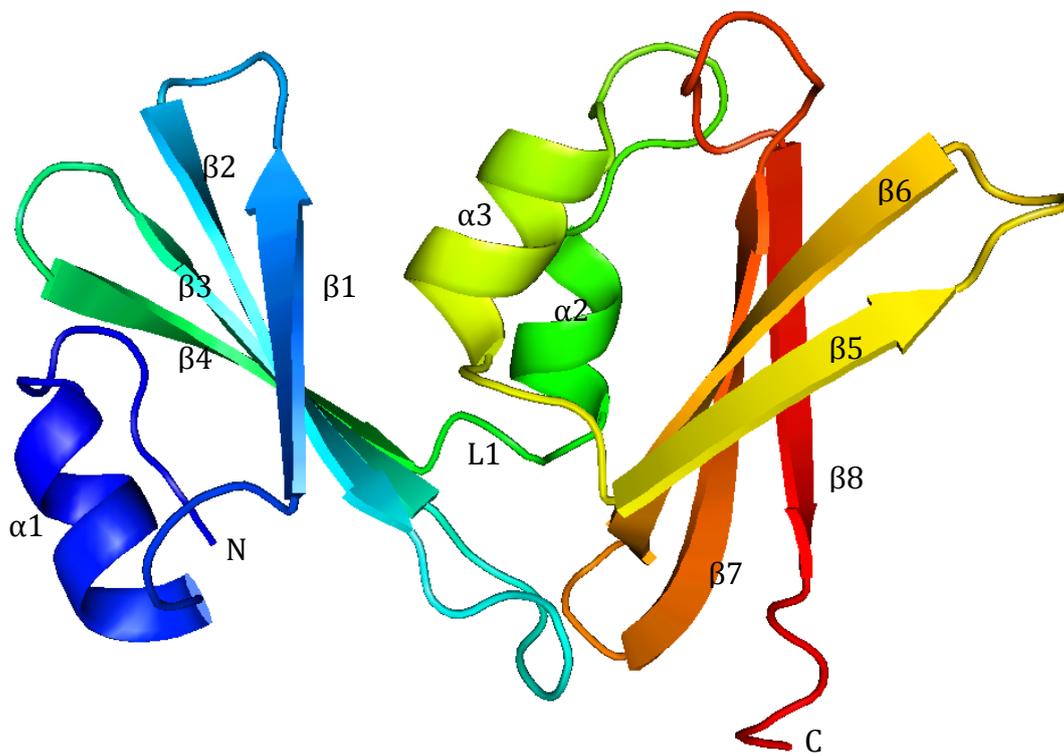
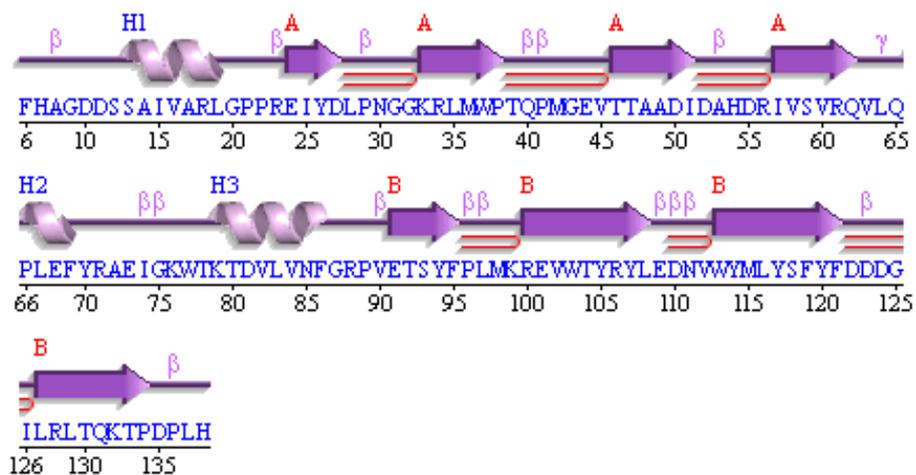


Figure 6.1: The 1.05 Å resolution structure of BPSL1204. The cartoon model was created using PyMol (Schrödinger, LLC), with N- to C-terminus coloured as blue-red, to represent the overall folding and structural arrangement of BPSL1204.

Two domains of $\alpha 1\beta 4$ and $\alpha 2\beta 4$ are clearly seen, separated by a very short loop (shown here as L1 in green).



Key:

Sec. struc: Helices labelled H1, H2, ... and strands by their sheets A, B, ...
 Helix Strand
 Motifs: β beta turn γ gamma turn beta hairpin

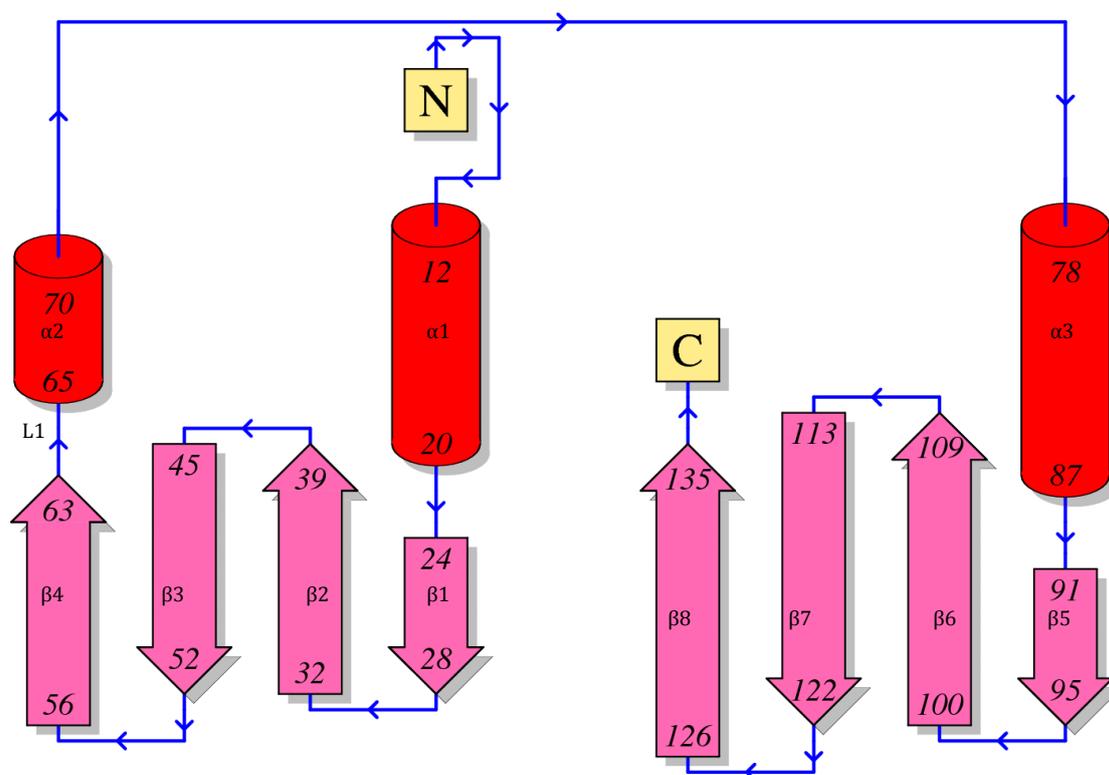


Figure 6.2: *Top:* Secondary structure arrangement of α -helical, β -strand and β -sheet elements of BPSL1204. *Bottom:* Topology schematic of BPSL domain arrangements, with α -helices represented by cylinders and β -strands represented as arrows.

The above illustrations allow for the identification of the specific secondary structure a particular residue is involved in, in addition to aiding with the location of one secondary structure element in regards to the overall protein structure and sequence. The diagrams were created using the PDBsum application made available by the PDB (Laskowski, 2009).

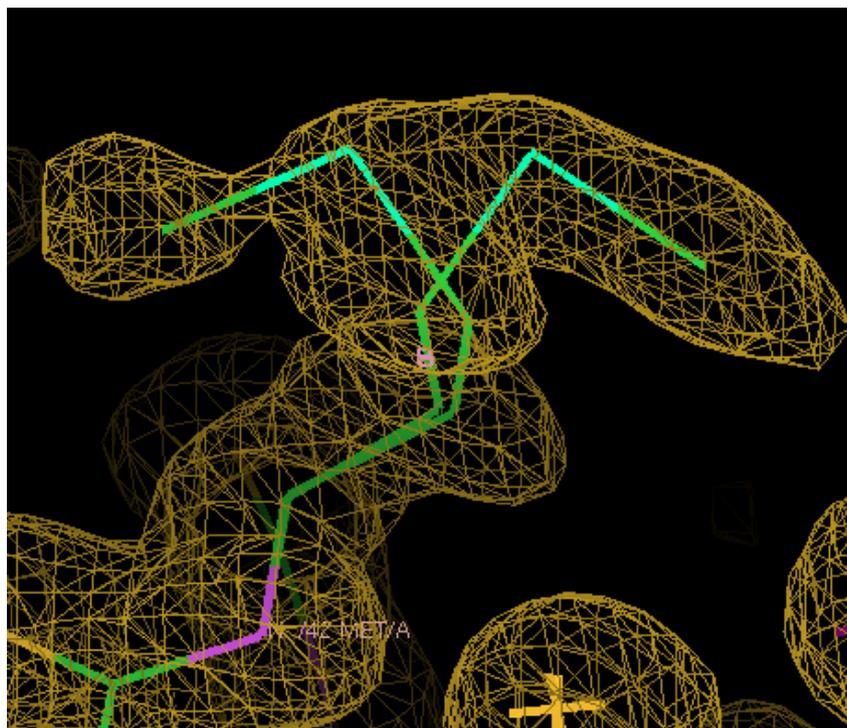


Figure 6.3a: Met⁴² was seen to occupy two conformations. From left-to-right in the above image a roughly 3:7 ratio was modelled into the electron density in Coot.

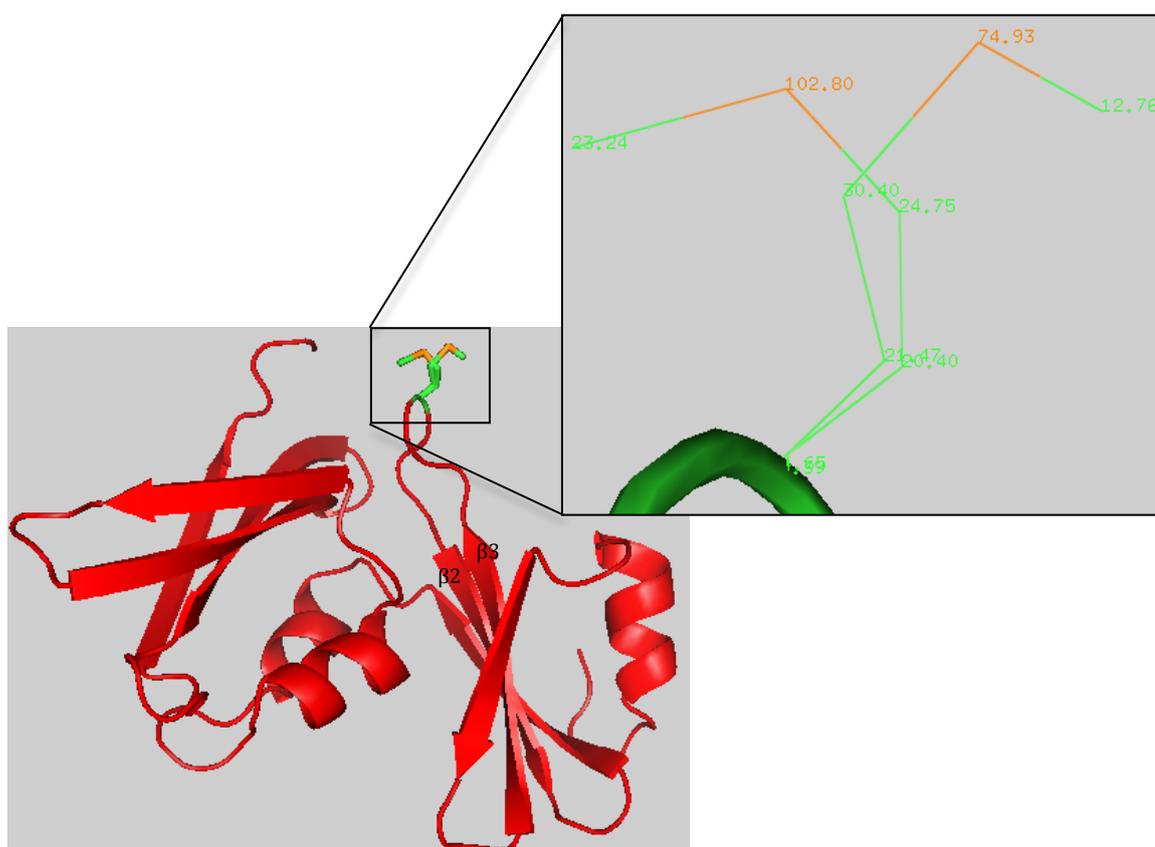


Figure 6.3b: Location and flexibility of Met⁴² in BPSL1204. Met⁴² is found at the ‘tip’ of the β -turn between β 2 and β 3. The inset image shows the B-factors of the atoms in the side-chain of Met⁴². There appears to be a reasonably high degree of positional movement around the C γ and S δ of the residue, the root of the two different conformations seen.

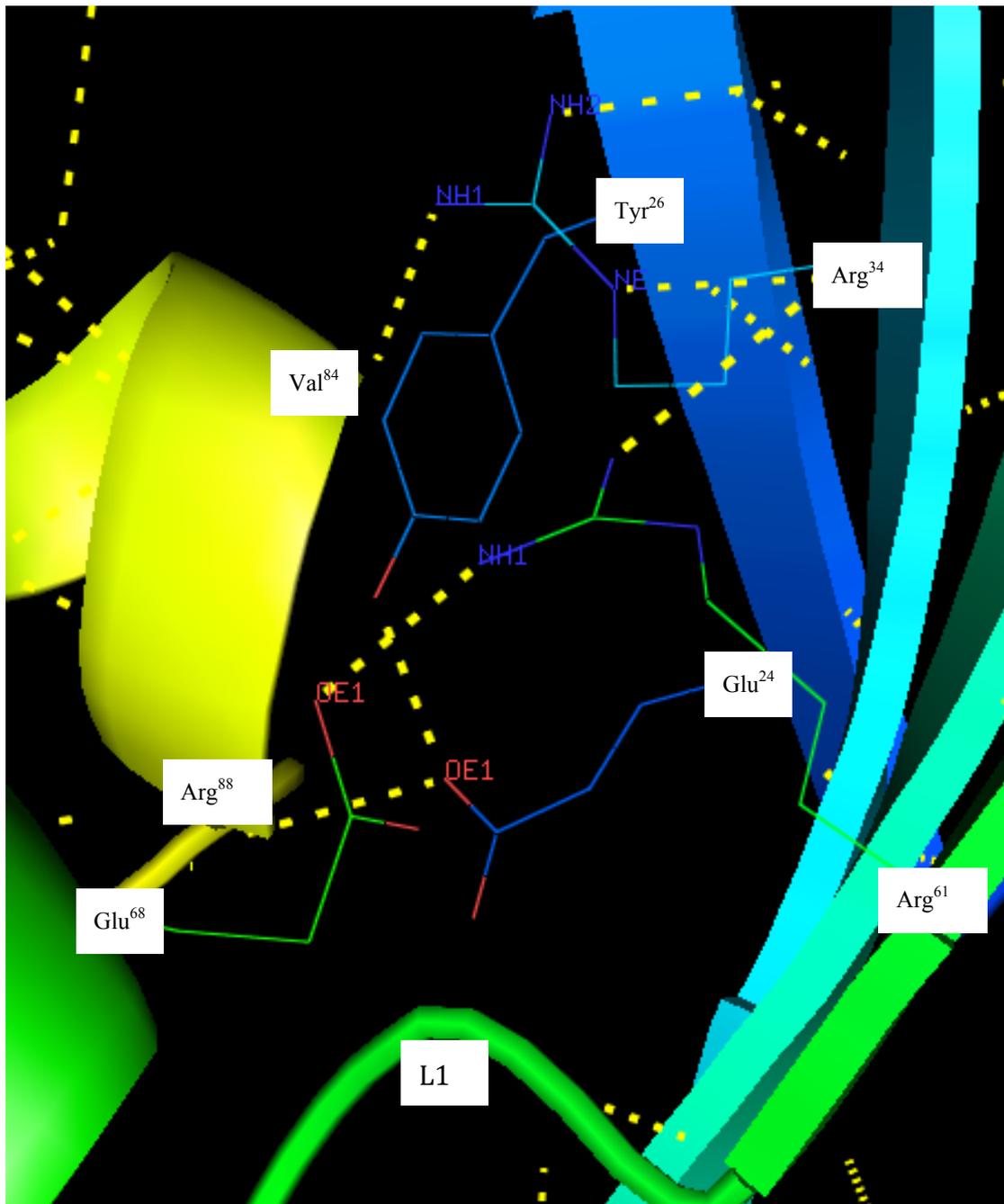


Figure 6.4: Investigation of the domain interface in Pymol of the BPSL1204 crystal structure.

Hydrogen bonds are coordinated between residues Tyr²⁶OH-Glu²⁴OE1, Glu²⁴OE1-Arg⁸⁸N and Arg³⁴NH1-Val⁸⁴O. In addition an ion-pair is seen to bridge over the interdomain loop L1 between Arg⁶¹ and Glu⁶⁸.

of those expected at a protein interface, biologically or in a crystal. However, the presence of hydrophilic residues such as glutamate and arginine that create the hydrogen bonds, within a reasonably buried interfacial region, raises doubt as to the biological relevance of such an interface.

This is compounded when the interdomain region surrounding the small L1 loop between the two domains is probed. As shown in both Figure 6.4 and Figure 6.5, the loop is bridged by a potential ion-pair of Arg⁶¹NH1 and Glu⁶⁸OE1. A water bridge crossing the domains between Arg⁶¹NH2-H₂O³-Asn⁸⁵OD1 further supports this strong interaction. In addition, a potential ion-pair is seen between Arg⁵⁴NH2 and Asp⁵⁰OD2. Although these residues both belong to the N-terminal domain, their strong interaction leads to an orientation of Arg³⁴NH1 that allows hydrogen bonding with Val⁸⁴O in the C-terminal domain.

When the B-factors of the residues of L1 are examined as in Figure 6.6, it can be seen that the loop has very little vibrational movement with low B-factors, highly suggestive of a rigid structure. However when one considers that the loop is flanked by very strong ionic interdomain interactions, one may well expect to see very little movement in this small loop as it becomes almost clamped into place. It is also unusual to find such charged, hydrophilic interactions buried between two protein domains with hydrophobic interactions much more common (Barlow & Thornton, 1983). Therefore it may be possible that the low B-factors seen in L1 are a non-biological representation of the region's flexibility with the interdomain interactions formed as a product of packing into a crystal lattice. The biological conformation of BPSL1204 domains may be different to that witnessed in the crystal structure, perhaps with an increase in flexibility between domains.

With the crystal structure now solved to a resolution of 1.05 Å, bioinformatic investigation was required to establish potential structural similarities and to attempt to predict binding partners in order to obtain insights into the unknown function of BPSL1204 in *Burkholderia pseudomallei*.

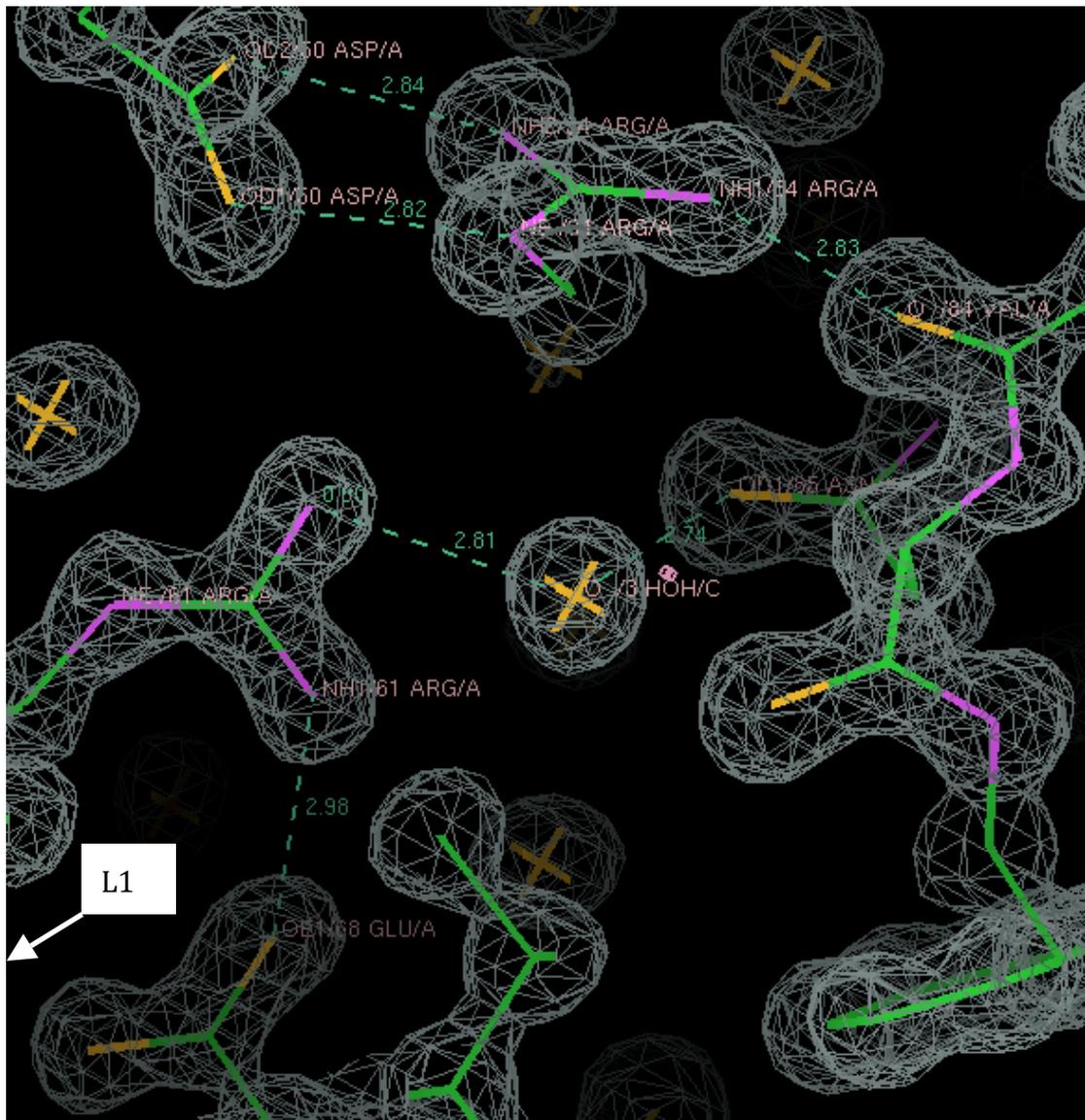


Figure 6.5: Investigation using Coot of the major bonds formed in the interface between BPSL1204 domains. Electron density is contoured to 2σ at 1.05 Å resolution.

The ionic interaction between Asp⁵⁰ and Arg³⁴ holds an interdomain hydrogen bond between Arg³⁴ and the main chain of Val⁸⁴.

A further ion pair is seen to flank L1, between Arg⁶¹ and Glu⁶⁸. This acts to strongly hold either end of the interdomain loop L1 in an inflexible conformation. In addition to determining the interface surrounding L1, Arg⁶¹ is also then held in coordination with H₂O³, creating a cross-domain water-bridge to Asn⁸⁵, further structuring the quaternary arrangement of the protein.

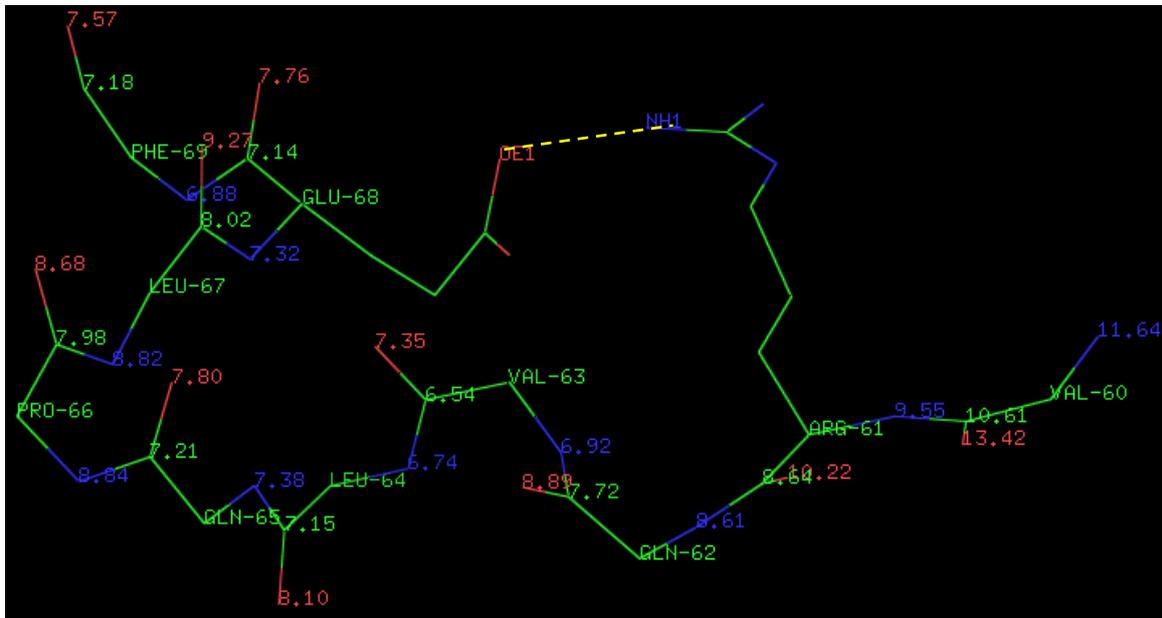


Figure 6.6: B-factors within the short L1 interdomain loop. The figure shows each main-chain atom within the L1-loop for analysis of potential flexibility. B-factor values are shown as Å². Val⁶⁰ and Phe⁶⁹ are shown as the next residues either side of L1.

No main-chain atom from Arg⁶¹ to Glu⁶⁸ breaches the >10.00 Å² cut-off included in refinement, reflecting a very rigid structure with little flexibility. The Arg⁶¹ and Glu⁶⁸ side chains are shown to illustrate the ion-pair (shown by the yellow dotted line) that is potentially responsible for the low vibrational movement in L1.

6.2 DALI analysis of BPSL1204-MA

6.2.1 The DALI server

The structural coordinates of BPSL1204-MA in the format of a .pdb file were uploaded to the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server) (Holm & Rosenstrom, 2010). DALI performed a comparison of the protein with all the protein structural coordinates present in the PDB (Protein Data Bank) (Berman et al, 2000), to find those coordinates that shared most 3D similarity to BPSL1204. Each potential hit was given a Z-score (Holm et al, 2008) to identify the level of 3D similarity. The higher in value the Z-score was the greater the structural similarity and Z-scores of less than 8 were counted as probably not significant. Other information like the level of sequence identity, the number of residues involved and the RMSD values of the C α atoms superimposed were also given by Dali.

6.2.2 Results of Dali Search

The structure in the PDB with the strongest structural similarity to BPSL1204-MA was that of the 161-residue β -lactamase Inhibitory Protein I (BLIP-I) from *Streptococcus mutans*. This model (PDB entry 3d4e, Joint Center for Structural Genomics, <http://www.jcsg.org>) had a Z-score of 9.2 in comparison to BPSL1204-MA, an RMSD of 1.7 Å for C α atoms but only 9% sequence identity. BLIP-I is also a two-domain protein, with both domains consisting of $\alpha 2\beta 4$, which appeared structurally similar to each other and to the two BPSL1204 domains.

Structural similarity to the 87-residue *E. coli* OsmE osmotically-inducible lipoprotein of unknown function was also seen (PDB: 4dm5, Mamelli et al), with a Z-score of 8.9 and a C α RMSD of 1.8 Å. This protein consists of a single $\alpha 3\beta 4\alpha 1$ domain, which when superimposed upon BPSL1204 showed very high structural similarity with the BPSL1204 C-terminal $\alpha 2\beta 4$ domain (Figure 6.7). However, due to the presence of extra helices in the single OsmE domain and the existence of two domains in BLIP-I that both exhibited similarity to both of the two

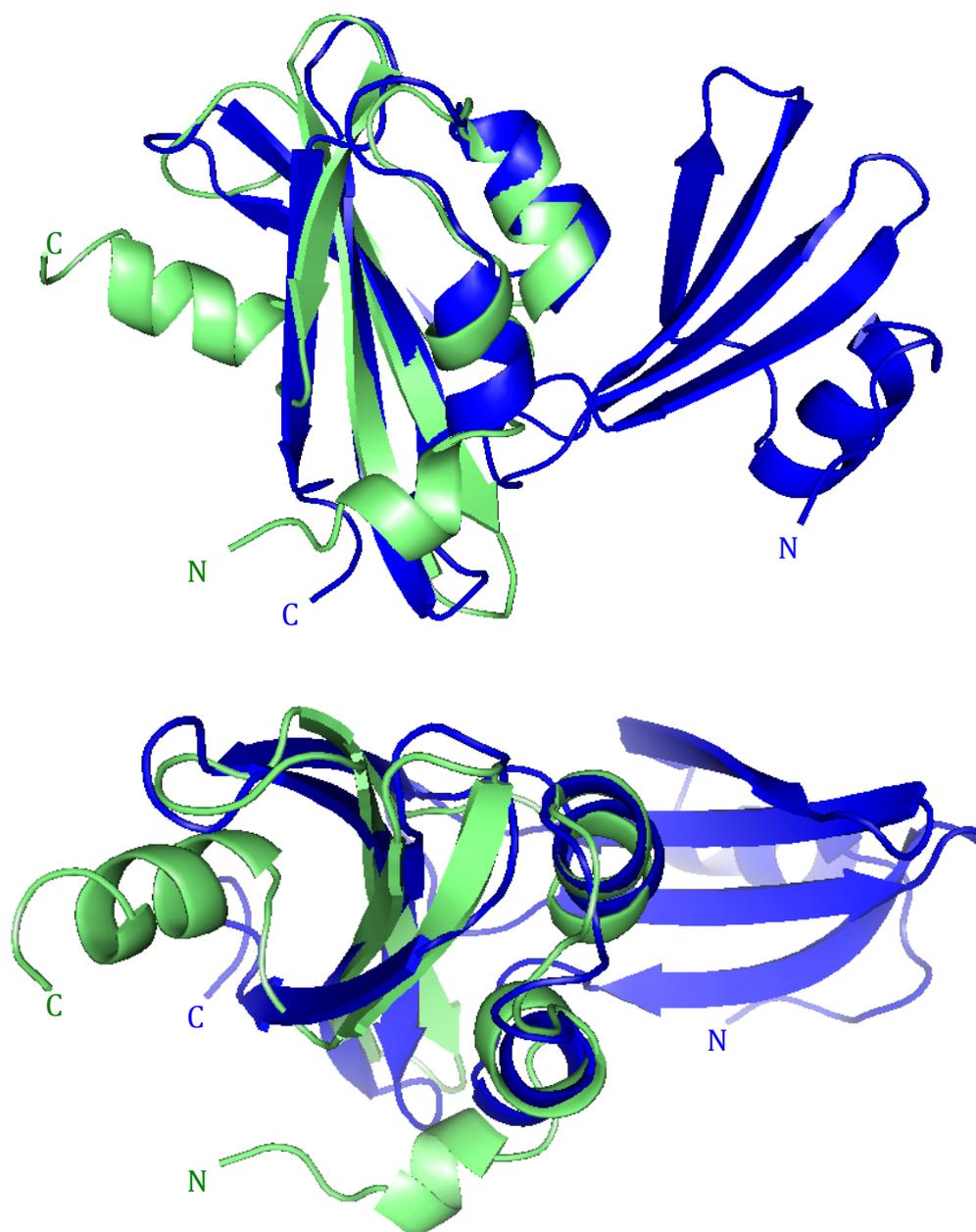


Figure 6.7: Superimposition of BPSL1204 (blue) with OsmE (green) in PyMol. *Top:* Side view of the $\alpha 3\beta 4\alpha 1$ OsmE on the $\alpha 1\beta 4/\alpha 2\beta 4$ BPSL1204. *Bottom:* Above view of the OsmE:BPSL1204 superimposition.

The side view (top) clearly shows OsmE matching well to the $\alpha 2\beta 4$ C-terminal domain of BPSL1204, however a small extra α -helix (bottom front) and an extra C-terminal α -helix in OsmE do not agree with the BPSL1204 structure.

The above view (bottom) shows how the superimposition contains α -helices that overlay well in both structures and maintains the curvature of the β -sheet. However OsmE contains only one domain, unlike BPSL1204 and BLIP-I.

domains of BPSL1204, coupled with the lack of assigned OsmE function, it was decided that BLIP-I provided the best candidate for further structural comparisons.

Aside from BLIP-I and OsmE, among the top-scoring structures found by the Dali search were complexes of BLIP-I bound to a β -lactamase (PDBs: 2g2w, 1jtg, 3c7u, 1s0w and 2b5r), with homology found with the BLIP-I component of each complex. The results of the Dali search are summarised below:

PDB code	Z-score	C α RMSD	N $^{\circ}$ of Res	Description
3d4e	9.0	1.7	162	Putative BLIP-I
4dm5	8.9	1.8	87	OsmE
2g2w	7.8	3.4	164	β -lactamase SHV-1 with BLIP-I
1jtg	7.0	3.3	165	β -lactamase TEM-1 with BLIP-I
3c7u	6.9	3.4	165	β -lactamase TEM-1 with BLIP-I
1s0w	6.9	3.3	165	β -lactamase TEM-1 with BLIP-I
2b5r	6.9	3.3	165	β -lactamase TEM-1 with BLIP-I

The BPSL1204 Dali search summary above (Holm et al, 2008) shows the top 7 results from a list of 1000, which rapidly decreased in Z-score and increased in RMSD after BLIP-I and OsmE results. As can be seen, BLIP-I and OsmE gave the lowest RMSD suggestive of the best matching coordinates, with Z-scores >8.0 significance level. All of the structures identified as being potentially structurally similar to BPSL1204 held low sequence identity. Therefore preliminary structural analysis of BPSL1204 suggested structural similarities to BLIP-I, and therefore offered a potential biological function.

6.3 Comparisons of BPSL1204-MA and BLIP-I

6.3.1 Sequence Homology

When the sequence of BPSL1204 and BLIP-I from *Streptococcus mutans* are aligned using the ClustalW server (www.ebi.ac.uk/Tools/msa/clustalw2) (Larkin et al, 2007), only 16 of BLIP-I's 179 residues are identical to the entire 166-residue sequence of BPSL1204 (Figure 6.8). As a 20-residue truncated BPSL1204-MA was actually used for the structural analysis in DALI, this removed a single aligned residue from BPSL1204 and left only 8.4 % identity for BLIP-I (15/179) or 10.3% identity (15/146) for BPSL1204-MA between the two protein sequences. 30 non-conserved residues in the truncated alignment also possessed similar chemical properties (16.8% of BLIP-I and 20.5% BPSL1204-MA), but in general it is clear that the sequences shared very little significant homology to each other.

In addition if the *Burkholderia* family of bacteria, which reports only homologs of unknown function BPSL1204 lipoprotein, are excluded from a search of the NCBI BLAST database then even with the database search loosened to report the top 1000 matches from all other recorded species there is no homology found with a BLIP or BLIP-related protein. Indeed, even a simple BLASTp (Altschul et al, 1990) against all the translated sequences of just *S. mutans* strains did not find a sequence homolog of BPSL1204, with proteins such as a replicative DNA helicase with Evalue = 1.2 being found before BLIP-I. Therefore with no detectable sequence homology found between BPSL1204 and BLIP-I, the structural arrangements of each macromolecule were investigated and compared.

6.3.2 Similarities of Individual Domains

When the superimposed BLIP-I coordinates (PDB: 3d4e) from the DALI server (Holm et al, 2008) are overlaid with the BPSL1204-MA structure (Figure 6.9), it is clear that there is a high degree of 3D similarity between the N-terminal $\alpha 2\beta 4$ domain of BLIP-I and the C-terminal $\alpha 2\beta 4$ domain of BPSL1204-MA. In addition

```

CLUSTAL 2.1 multiple sequence alignment

gi|53718840|ref|YP_107826.1|          -----MFARLPVLFACAALLAGCAQFWQQFHAGDDSSAIV 36
gi|193885428|pdb|3D4E|A          GEDKKGGTKPSNEAALTKTENLDFRLSFNKKIKVTTDQNHFSGGTSIEQLK 50
                                     ::: * : : ** * . . . :
                                     * * * * * * * * * * * * * * * * * * *
gi|53718840|ref|YP_107826.1|          ARLGPPREIYDLENGGKRLMWPQPMGEVTTAAIDAHDRIVSVRQVLOP 86
gi|193885428|pdb|3D4E|A          QWFGDPNKSEQRNAGNITLDSYTWVKDGAVINAOQLYKNSTVARSISNFSF 100
                                     :* *.: : * * * * * * * * * * * * * * * * * * *
                                     * * * * * * * * * * * * * * * * * * *
gi|53718840|ref|YP_107826.1|          LEFYRAEIGKWKTKDVLVNFGRPVETSFFPLMKREVWYRYLEDNWWYML 136
gi|193885428|pdb|3D4E|A          SREAKIGKEDYDELKIGESYKKIIVEIKLGEFVLSQSSSDKEXKQTVWSS 150
                                     . : .: : : : : : : : * * * * * * * * * * * * * * *
                                     * * * * * * * * * * * * * * * * * * *
gi|53718840|ref|YP_107826.1|          YSFYFDDDGILRLTQKTPDPLHDPDRRTLF 166
gi|193885428|pdb|3D4E|A          GIKTKSSSATIELYFEN-GLLKKNKTQKDL 179
                                     ..... :* * . . * : : * *

```

Figure 6.8: ClustalW-alignment of full-length BPSL1204 (top row) with full-length BLIP-I (bottom row). Colours indicate residues with similar chemical properties (Larkin et al, 2007).

When aligned, only 16 residues out of 166 BPSL1204 residues match those in BLIP-I (*). When truncated only 15 align, leaving a low % sequence identity. Some residues exist as conservative substitutions (:), but in general no significant sequence homology exists.

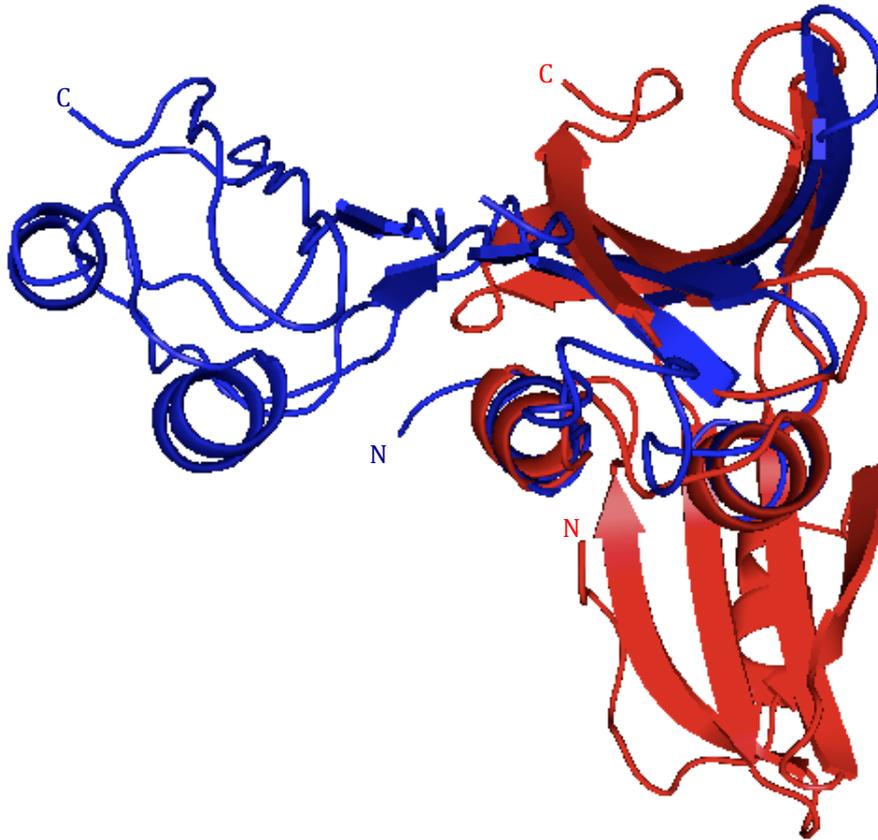


Figure 6.9: Dali superimposition of BLIP-I (blue) on BPSL1204 (red), made in PyMol. Both proteins consist of two domains, $\alpha 2\beta 4/\alpha 2\beta 4$ and $\alpha 1\beta 4/\alpha 2\beta 4$ respectively, but the best structural superimposition places the N-terminal domain of BLIP-I over the C-terminal domain of BPSL1204.

to this, as illustrated by Figure 6.10, each of the two domains from each protein can be superimposed upon each other in all possible combinations (1204-N:BLIP-C, 1204-C:BLIP-C, 1204-N:BLIP-N and 1204-C:BLIP-N). These domains, which exhibit no significant sequence homology to each other, all possess strikingly similar tertiary structure. It is therefore clear that the N-terminal and C-terminal domains of both proteins all have essentially the same fold (Figure 6.11).

Both the BLIP-I N-terminal and the BPSL1204-MA C-terminal domain share the same $\alpha 2\beta 4$ arrangement, whereas the BPSL1204-MA N-terminal domain exists as $\alpha 1\beta 4$, perhaps due to the solubilising truncation. Figure 6.12 shows the predicted secondary sequence of full-length BPSL1204, which holds the potential for another N-terminal α -helix. However, there does not appear to be enough N-terminal polypeptide upstream of $\alpha 1$ when BPSL1204 is truncated after the LLXGCX motif to allow the formation of an $\alpha 1$ -loop- $\alpha 2$, as seen at the BLIP-I N-terminus. Sequence analysis by RADAR (Rapid Automatic Detection and Alignment of Repeats) (Heger & Holm, 2000) detected no significant repeated sequences within either protein. Therefore the observed structural similarities in both the inter- and intramolecular superimpositions may be due to chance, although there is also the possibility that they have related functions but diverged evolutionarily a long time ago, perhaps through some gene doubling event.

6.3.3 Differential Arrangement of Tertiary Structure

Upon further investigation of the BLIP-I:BPSL1204 superimposition, it was clear that, although the domains individually share very similar 3D structures between each protein, the spatial arrangement of domains in each crystal structure differ considerably. The two domains of BLIP-I lie reasonably flat side-by-side connected by a looped region that crosses behind the second $2\alpha 4\beta$ to begin the C-terminal domain, whereas the domains of BPSL1204-MA stack on top of one another connected by a considerably shorter looped region that directly joins domains and gives a slight twist between them (Figure 6.13).

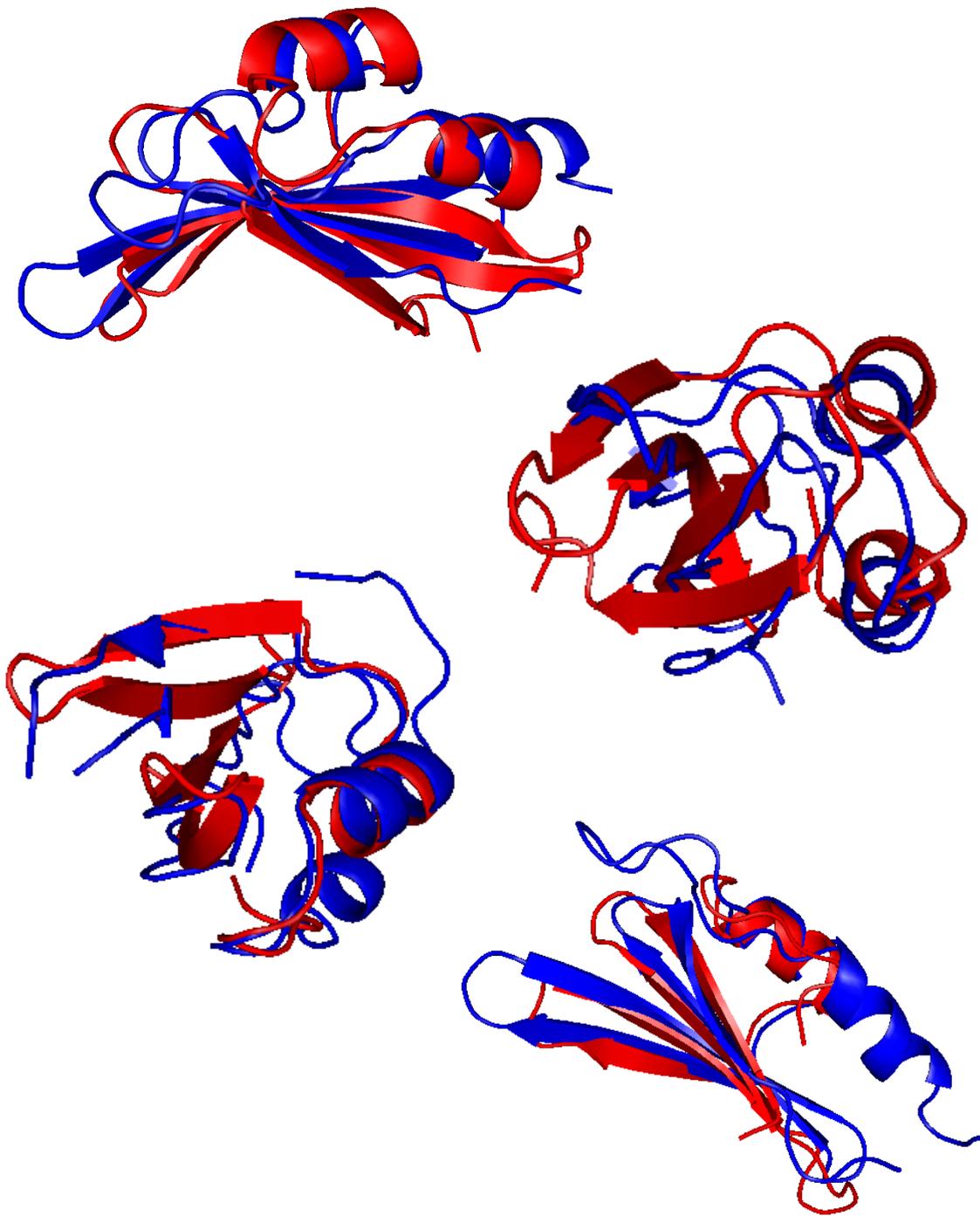


Figure 6.10: Superimpositions of individual BLIP-I (blue) and BPSL1204 (red) domains in PyMol. *Top left:* C-dom BPSL1204, N-dom BLIP. *Top right:* C-dom BPSL1204, C-dom BLIP. *Bottom left:* N-dom BPSL1204, C-dom BLIP. *Bottom right:* N-dom BPSL1204, N-dom BLIP.

Superimposition of the individual domains of each protein overlay with a high degree of structural similarity, suggestive of both proteins fulfilling a structurally similar function.



Figure 6.11: Self-superimpositions of BPSL1204 and BLIP N- and C-domains (red/orange and blue/cyan respectively).

Both domains of both proteins superimpose very well upon themselves (above) and each other (Figure 6.10). All domains have essentially the same fold.

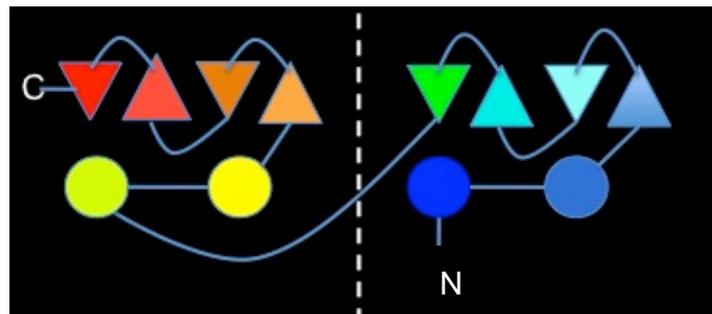
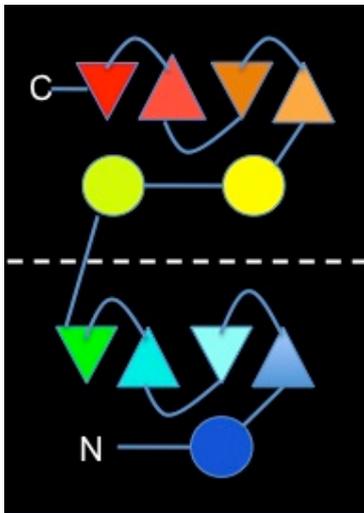
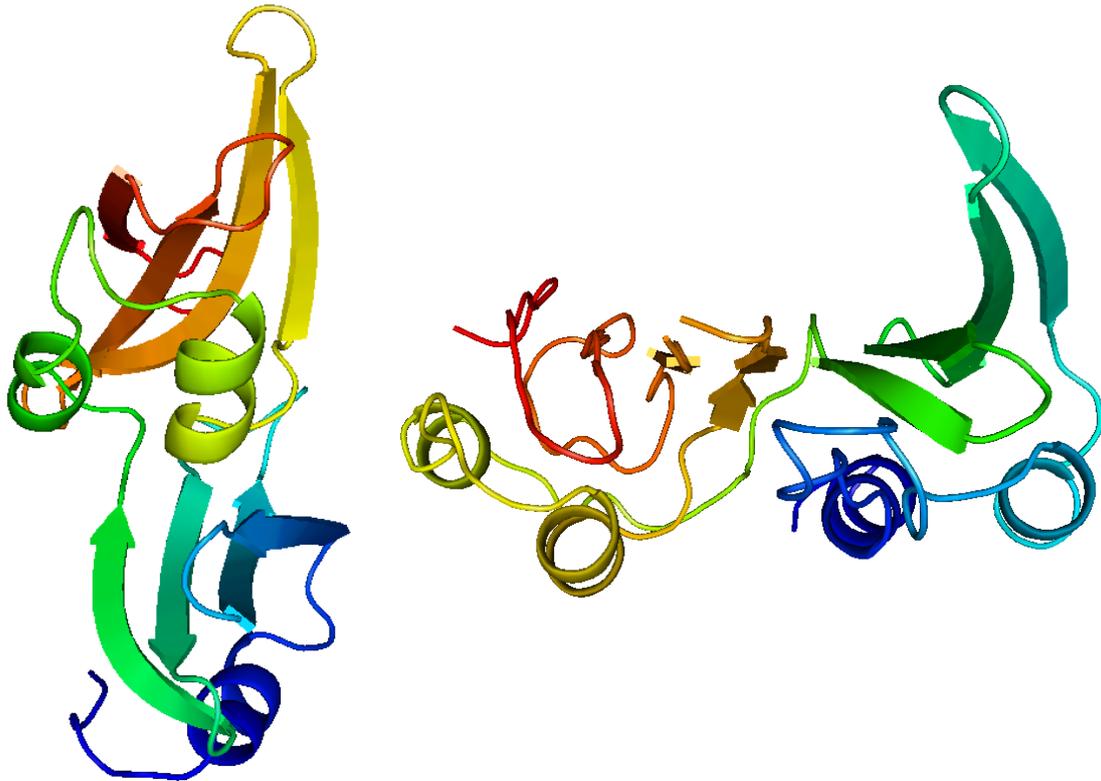


Figure 6.13: Comparison of BPSL1204 and BLIP-I domain arrangements. *Left top:* BPSL1204 above view. *Right top:* BLIP-I above view. *Left bottom:* Schematic diagram of BPSL1204 domain arrangement. *Right bottom:* Schematic diagram of BLIP-I domain arrangement. Schematic diagrams represent α -helices as circles and β -strands as triangles. All images are coloured blue to red, N- to C-terminus.

As can be seen, BPSL1204 domains stacked on top of each other, whilst BLIP-I domains exist in a side-by-side conformation.

This was clearly demonstrated by the DALI superimposition of BPSL1204 upon the TEM-1:BLIP-I inhibition complex, which placed the best two complementing domains over one another, whilst the remaining domain for BPSL1204 extended into 3D space in a totally different orientation from the key inhibiting BLIP-I domain (Figure 6.14).

Further investigation of the looped regions between domains showed that B-factors of residues in the BLIP-I loop appeared to create a region of slight flexibility (higher B-factors), whereas the connecting loop in the BPSL1204 structure showed residues occupying space with relatively low B-factors, indicative of a more rigid 3D structure that may not have led to a similar structural flexibility between domains.

6.3.4 Crystal Packing Versus Biological Conformation.

With the high structural similarity between BLIP-I and BPSL1204 domains, but with significant difference in domain arrangements, it was possible that the process of crystallising BPSL1204 might have caused the packing of monomers into the crystal lattice in a non-biological conformation. This is clearly not the case for BLIP-I because, as seen in Section 6.3.4, a number of independent structures of BLIP-I as a monomer and in complex with its inhibition target β -lactamase show a consistent side-by-side domain orientation. However with only one BPSL1204 structure it was possible that in a different crystal lattice, BPSL1204 might have the same domain arrangement as BLIP-I.

Extensive initial crystallisation trials of BPSL1204 (see Section 5.2) had suggested that the 9% MPD/0.1M sodium acetate pH5.5 condition from which the crystals that led to the 1.05 Å resolution structure was the only condition to generate crystals of useful diffraction (see Section 5.3.3). Therefore it was proposed a structure solution of a high sequence homologue from a closely related species to *B. pseudomallei* should be carried out to see if that had a different domain arrangement from BPSL1204 and resembled that of BLIP-I.



Figure 6.14: Superimposition of BPSL1204 (red) on a complex of BLIP-I (blue) with its TEM-1 β -lactamase target protein (green).

The best alignment placed the C-terminal domain of BPSL1204 over the C-terminal domain of BLIP-I, but due to differential domain arrangements the N-terminal BPSL1204 domain extended out into non-interacting space.

In the BLIP-I inhibition of TEM-1, the N-terminal domain of BLIP-I is responsible for forming the inhibition complex, via the protruding loop designated by \blacktriangledown . If the domain arrangement seen in BPSL1204 was biologically accurate, then it seemed unlikely that BLIP-I-like activity was possible due to a different quaternary arrangement of domains.

If so, then it would allow the inference that BPSL1204 and its homologue are both functional homologs of BLIP-I.

To this end the 92% amino-acid sequence identical protein BCAL2351 from *B. cenocepacia* was identified by a BLAST search using BPSL1204 (see Section 7.1.2). BCAL2351 was restriction-cloned into pET14b and pBAD/hisB (genome of *B. cenocepacia* H111 supplied by the lab of Dr Mark Thomas, University of Sheffield Medical School) to include an identical truncation to that of BPSL1204-MA (following the general protocol described in Section 3.4.2). Unfortunately, even after extensive over-expression trials in BL21 and Tuner (DE3) and Top10 cells it was not possible to achieve soluble protein (Figure 6.15), even though the two protein sequences differed by only 11 residues, most of which were conservative substitutions (Figure 6.16).

However due to the lack of flexibility in loop L1, as judged by low B-factors and the presence of a strong ionic interaction in the interdomain looped region of BPSL1204, it was considered unlikely that BPSL1204 and all near identical sequence homologues would occupy the exact same domain arrangement as BLIP-I. Nevertheless the high structural similarity between domains made it possible that BPSL1204 performs a similar biological function to BLIP-I.

6.3.5 Potential Interactions With TEM-1 Target

The best Dali structural match was with the N-terminal of BLIP-I over the C-terminal of BPSL1204. If the $\alpha 1\beta 4$ N-terminal domain of BPSL1204-MA alone is superposed on the $\alpha 2\beta 4$ N-terminal domain of BLIP-I (from *Streptomyces exfoliatus*) in complex with its inhibition target TEM1 (from *Escherichia coli*) using the DALI alignment, then it is clearly seen that both N-terminal domains present a short, protruding loop that forms a concave surface that occupies the convex lip of the TEM1 active site (PDB 3gmw)(Gretes et al, 2009). This is the accepted mode of inhibition of TEM1 β -lactamase activity by BLIP-I (Figure 6.17).

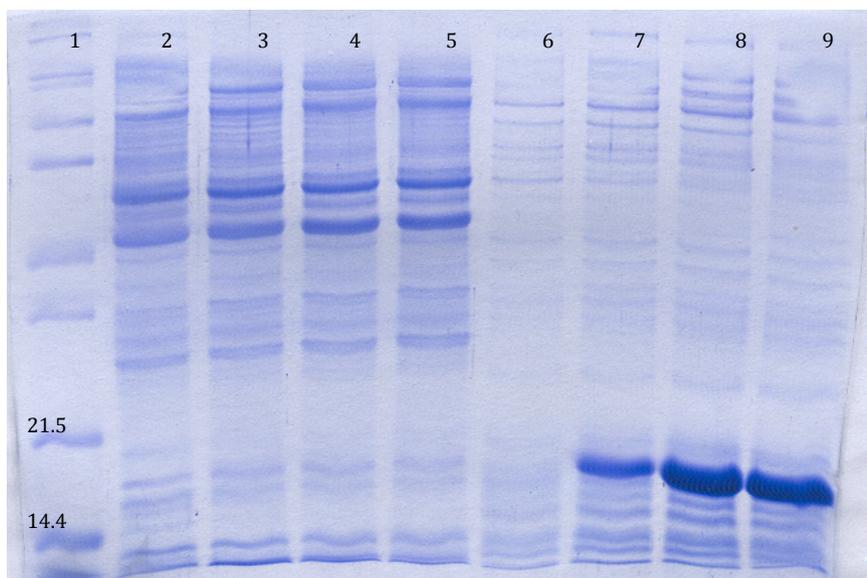


Figure 6.15: Example of SDS-PAGE analysis of BCAL2351-MA from BL21:pET14b over-expression for 3 hours in 1 mM IPTG. Lane 1= Mark12. Lane 2-5: No induction, 25, 30 and 37°C soluble fractions. Lane 6-9: No induction, 25, 30 and 37°C insoluble fractions.

Heavy insoluble over-expression of BCAL2351-MA is seen, with no soluble protein observed. Variations in [IPTG], expression temperature from 4-37°C and expression time from 3-28 hours were investigated, but no BCAL2351-MA was produced.

The same experiments were performed in TOP10 cells transformed with pBAD/hisB:BCAL2351-MA, using arabinose ($0.1-1 \times 10^{-5}$ %) as the inducing agent, but again no soluble protein was produced resulting in SDS-PAGE gels like the one shown above.

```

Query 3  ARLPVLFACAAALLAGCAQPWQQFHAGDDSSAIVARLGPPREIYDLPNGGKRLMWPTQPM 62
Sbjct 7  ARLPVL CA ALLAGCAQPWQQ+ AG D SAIVAR+GPPREIYDLP GGKRLMWPTQPM 66
          ████████████████████
Query 63 GEVTTAADIDAHDRIVSVRQVLQPLEFYRAEIGKWTKTDLVNFGRPVETSYPFLMKREV 122
Sbjct 67 GEITVAADVDAHRIVNVRQVLQPSFYRAEIGKWTKTDLVNFGRPVETSYPFLMKREV 126

Query 123 WTYRYLEDNVWYMLYSFYFDDGILRLTQKTPDPLHDPDRRTLF 166
Sbjct 127 WTYRYLEDNVWYMMYSFYFDPOGILRTTQKTPDPLHDPDRRTLF 170

```

Figure 6.16: Sequence alignment of BPSL1204 (Query) with BCAL2351 (Subject). Both sequences contain the LLAGCA glycery-cysteine lipid-anchoring motif. As can be seen, the two proteins are sequence homologs of each other.

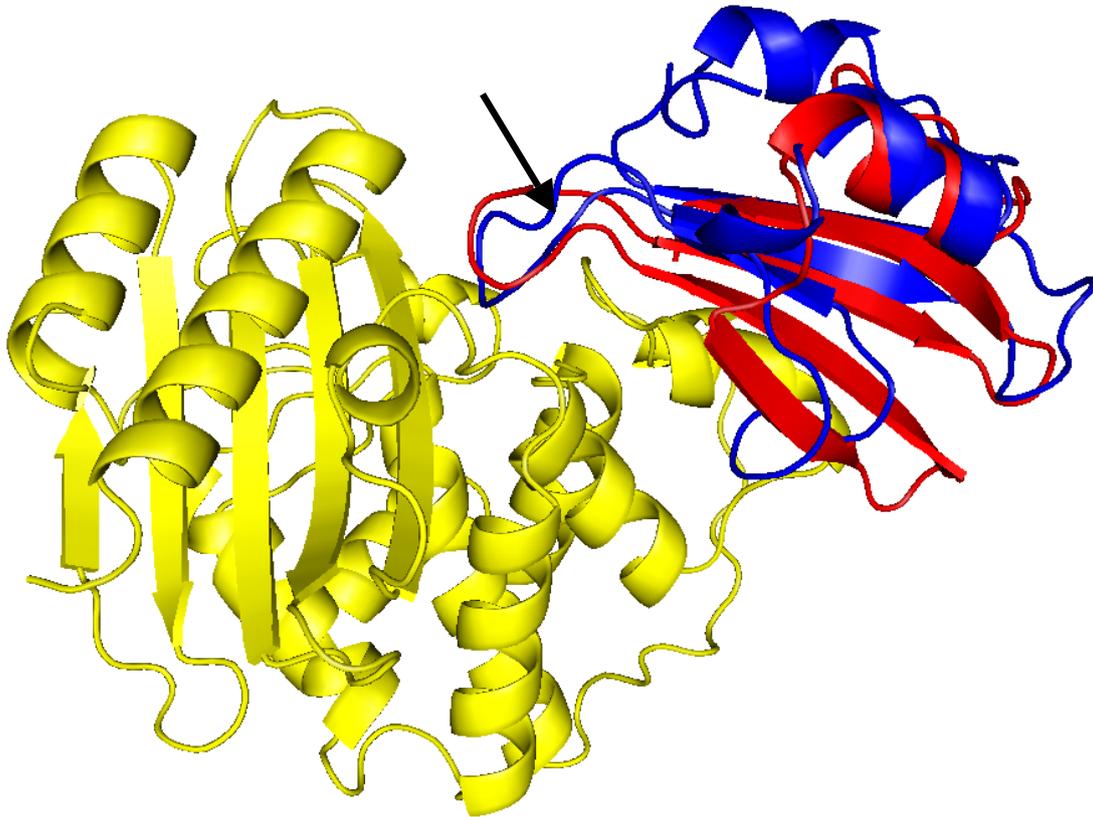


Figure 6.17: Inhibition-complex of TEM-1 (yellow) with the N-terminal domain of BLIP-I (blue) (Strynadka et al, 1996), superimposed with the N-terminal domain of BPSL1204 (red) in PyMol.

The concave surface offered by the BLIP-I loop is approximately replicated in structure by BPSL1204 (indicated by the **black** arrow) and rests over the small loop-helix-loop in the active site entrance, though the loop does not rest in the opening of the active site in BPSL1204 in the same conformation as BLIP-I.

This could be due to the unbound BPSL1204, with a loop not in the conformation held when in complex, being superimposed to the best possible fit upon a rigidly bound BLIP-I held in conformation by the specific contacts that allow inhibition-complex formation.

This potential 'inhibitory' loop from BPSL1204 did not superimpose perfectly upon that of BLIP-I (Figure 6.17) but it was concluded that the slight variation could potentially be a difference between possible complexed and uncomplexed conformations of the loop, with more rigidity expected in the inhibition-complex structure. Figure 6.18 shows the structural alignment of uncomplexed BLIP-I from *S. exfoliatus* with the same BLIP-I complexed with its TEM-1 target. The structural alignment had great confidence with a Z-score of 27.8, in addition to a very low C α RMSD of 0.5. It can be seen that there is a slight conformational change in the inhibitory loop upon binding to TEM-1, supporting a potential BPSL1204 binding-event. Furthermore the loop regions from complexed BLIP-I, uncomplexed BLIP-I and BPSL1204 were isolated from their independent .pdb files and aligned for comparison using Dali-lite (Holm et al, 2008). The results are summarised in Figure 6.19. As expected, the uncomplexed and complexed BLIP-I loops aligned in 3-D very well, whilst the BPSL1204 loop appeared to align with uncomplexed BLIP-I slightly better than complexed BLIP-I.

However, upon comparison of the amino-acid sequences of the short inhibition loops there is very little homology. When the loop sequences of BLIP-I from *S. mutans*, *S. clavigulus* and *S. exfoliatus* are all compared against the loop in BPSL1204, a single glycine residue was the only residue to align between the 4 proteins (Figure 6.20). It had been previously discovered within several TEM1:BLIP-I complexes, for example PDB 3gmw and 2g2w (containing the TEM1 homolog SHV1 from *Klebsiella pneumoniae* and the *Streptomyces clavuligerus* BLIP-I), that residues Gly⁴⁸ and Asp⁴⁹ played an integral part of the inhibition mechanism (Reynolds et al, 2006). Gly⁴⁸ offers increased Ramachandran rotation angles to facilitate the sharp bend in the loop and a hydrogen bond with the TEM1 Tyr¹⁰⁵, which in turn offered multiple hydrogen bonds between Asp⁴⁹ and several TEM1 active site residues (Ser¹³⁰, Lys²³⁴ and Arg²⁴⁴), essential for BLIP-I binding (Figure 6.21) (Strynadka et al, 1996).

Investigation of the equivalent loop in BPSL1204 with the knowledge that the conserved glycine functions to provide increased rotational flexibility means it is possible that the BPSL1204 loop could twist to align the adjacent glutamate with

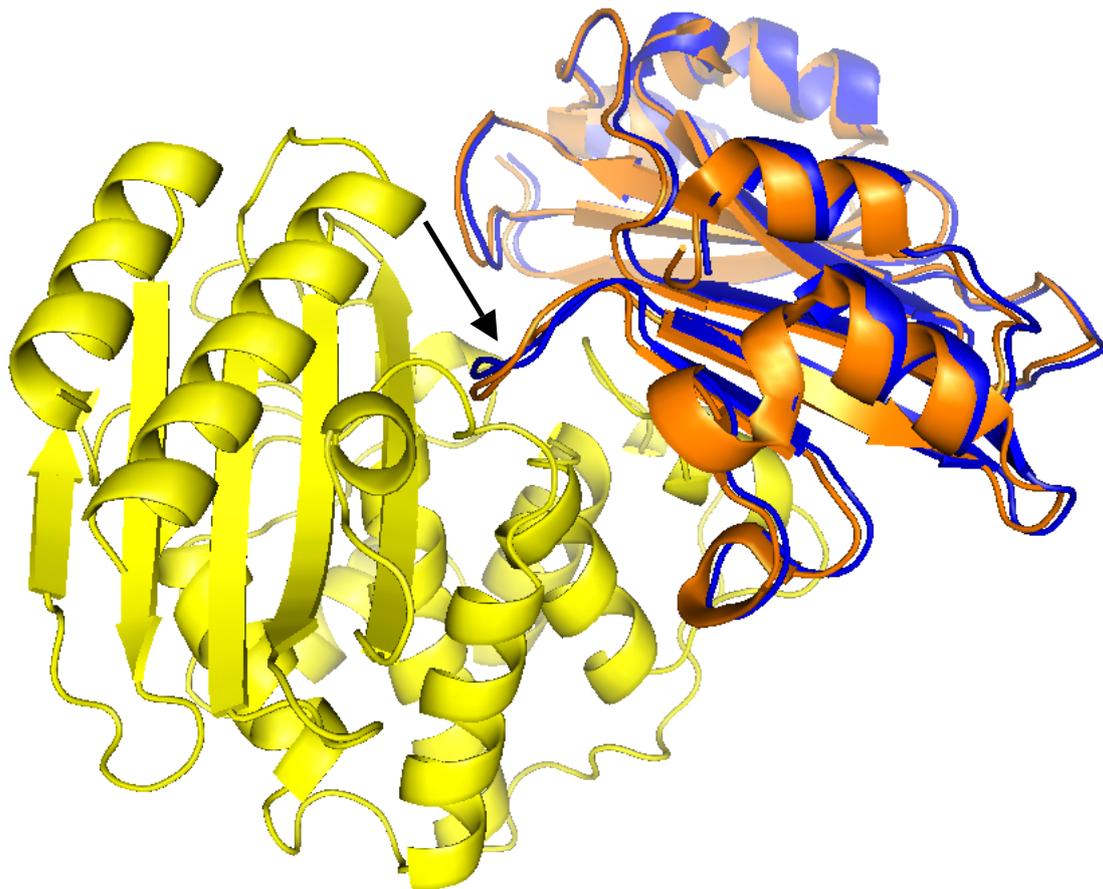
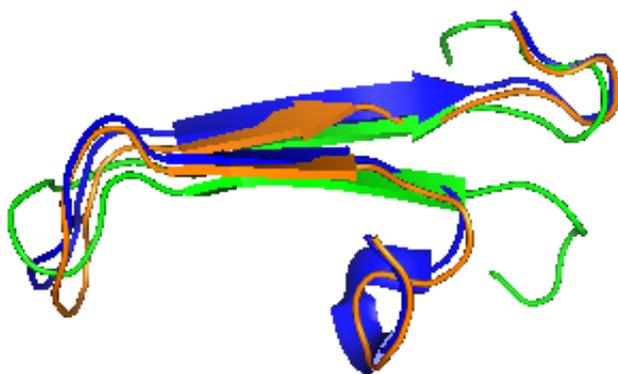


Figure 6.18: Superimposition of **uncomplexed BLIP-I** upon **BLIP-I in complex** with **TEM1**.

The N-terminal inhibitory loop of BLIP-I occupies similar coordinates in both the apo- and halo-structures, however there is a slight change in conformation upon binding to TEM-1.



Alignment	RMSD	Z-score	%ID
BLIP-<i>lu</i> to BLIP-<i>lc</i>	0.4	5.3	100
BPSL1204 to BLIP-<i>lc</i>	2.1	1.2	4
BPSL1204 to BLIP-<i>lu</i>	1.8	1.3	4

Figure 6.19: Dali-lite structural alignment of the inhibitory loop region of **uncomplexed BLIP-I**, **complexed BLIP-I** and the relative looped region in **BPSL1204**. The 15 residues either side of a conserved and essential glycine (see Figure 6.20) are shown, as the Dali server requires a minimum of 30 residues for 3-D alignment.

The uncomplexed BPSL1204 appears to resemble the loop of an uncomplexed BLIP-I slightly more than a complexed BLIP-I, with slightly lower C α RMSD and slightly higher (but still poor) Z-score.

```

a_S_clavigulus_BLIPlloop      -SIHCRGHAAGDYAYATFGFT- 21
b_S_exfoliatus_BLIPlloop      DSILCF-TESGDYAPYGGFSFT- 21
c_S_mutans_BLIPlloop          ITLDSYTWVKGDAVINAQL-YK- 21
d_BPSL1204_loop                -RLMWPTQPMGEVTTAADIDAHD 22
                                :          *:          . :

```

Figure 6.20: Alignment of inhibiting loops from **multi-species** BLIP-Is with crystal structures (essential Asp ± 10 residues) with the corresponding region in BPSL1204. Alignment was performed using ClustalW, with colour reflecting residues with similar properties.

All species lack homology throughout the 21 residues in the loop, but several conservative substitutions are seen (:), with a totally conserved (*) Gly/G residue -1 from the essential Asp/D (:).

A Glu/E in BPSL1204 replaces the essential, conserved Asp in the BLIPs. This conveys similar chemical properties but presents an extra δC , which may induce steric hindrance preventing BPSL1204 binding a β -lactamase.

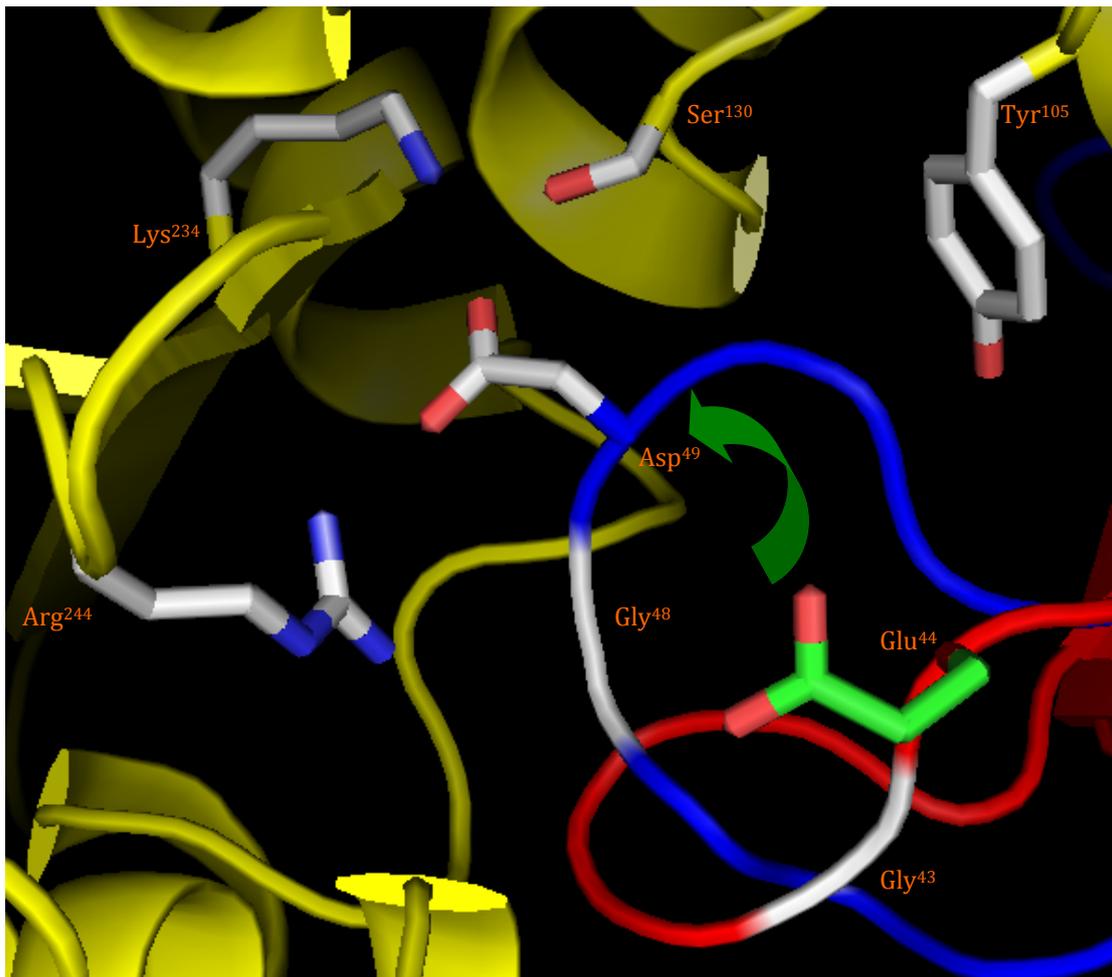


Figure 6.21: Investigation of TEM-1 (yellow) active site binding by BLIP-I (blue), and corresponding BPSL1204 (red) binding potential, in PyMol.

In BLIP-I binding, Gly⁴⁸ rotation twists the loop to align Asp⁴⁹ to occupy the active site mouth, where electrostatic interaction with TEM-1 Arg²⁴⁴ and Lys²³⁴ facilitates binding. The loop is further stabilised by TEM-1 Tyr¹⁰⁵ and Ser¹³⁰ hydrogen-bonding to the twisted loop backbone.

By superimposing unbound BPSL1204 superimposed on TEM-1:BLIP-I, it is possible to visualise a rotation around Gly⁴³ (bottom of the picture) that might bring Glu⁴⁴ (green) into alignment with BLIP-I Asp⁴⁹, as shown by the green arrow.

the essential aspartate of BLIP-I (Figure 6.21). This would offer the same potential for hydrogen bonding, although the side chain would extend further into the binding site in the opening of the TEM1 active site. Slight differences elsewhere in the BPSL1204 loop might 'lift' the glutamate away from the binding site enough to account for the extra atom. Alternatively BPSL1204 could bind to a protein similar to TEM1 but with a deeper active site entrance.

6.3.6 Potential Explanations for Structural Similarity of BPSL1204 to BLIP-I

Given the striking structural similarity between the individual domains of BPSL1204 and BLIP-I, and the discovery of the structural if not sequential potential for the N-terminal domain of BPSL1204 to provide a BLIP-I like inhibition-loop, it was proposed that BPSL1204 might act upon a protein with a function similar to that of a β -lactamase. This theory was supported by the fact that proteins such as cell-wall transpeptidases and other penicillin-binding proteins have very similar structures to those of β -lactamases (Knox et al, 1996)(Figure 6.22). On the basis of structural similarities in fold and in active site residues and the fact that they all bind β -lactam antibiotics, it has been proposed that β -lactam hydrolysing enzymes such as β -lactamases evolved from proteins that manipulated cell-wall peptidoglycan structure (Samraoui et al, 1986).

Such an explanation could explain the different arrangement of domains with high structural similarity, as they may be involved in an analogous function to that of a BLIP that requires an alteration in binding orientation. In addition, it was considered that the BPSL1204 structure used in comparisons was that of an N-terminal 20-residue truncate, with the truncation in aid of releasing solvent-interacting domains from a membrane-associating glycerylcysteine lipoprotein anchor. β -lactamases are traditionally thought of as secreted proteins (Koshland & Botstein, 1980), therefore questions were raised as to the biological purpose of covalently anchoring an effector of a secreted process. Therefore BPSL1204 might have the potential to function in a membrane-associated process with some analogy to β -lactamase activity or cell wall biosynthesis.

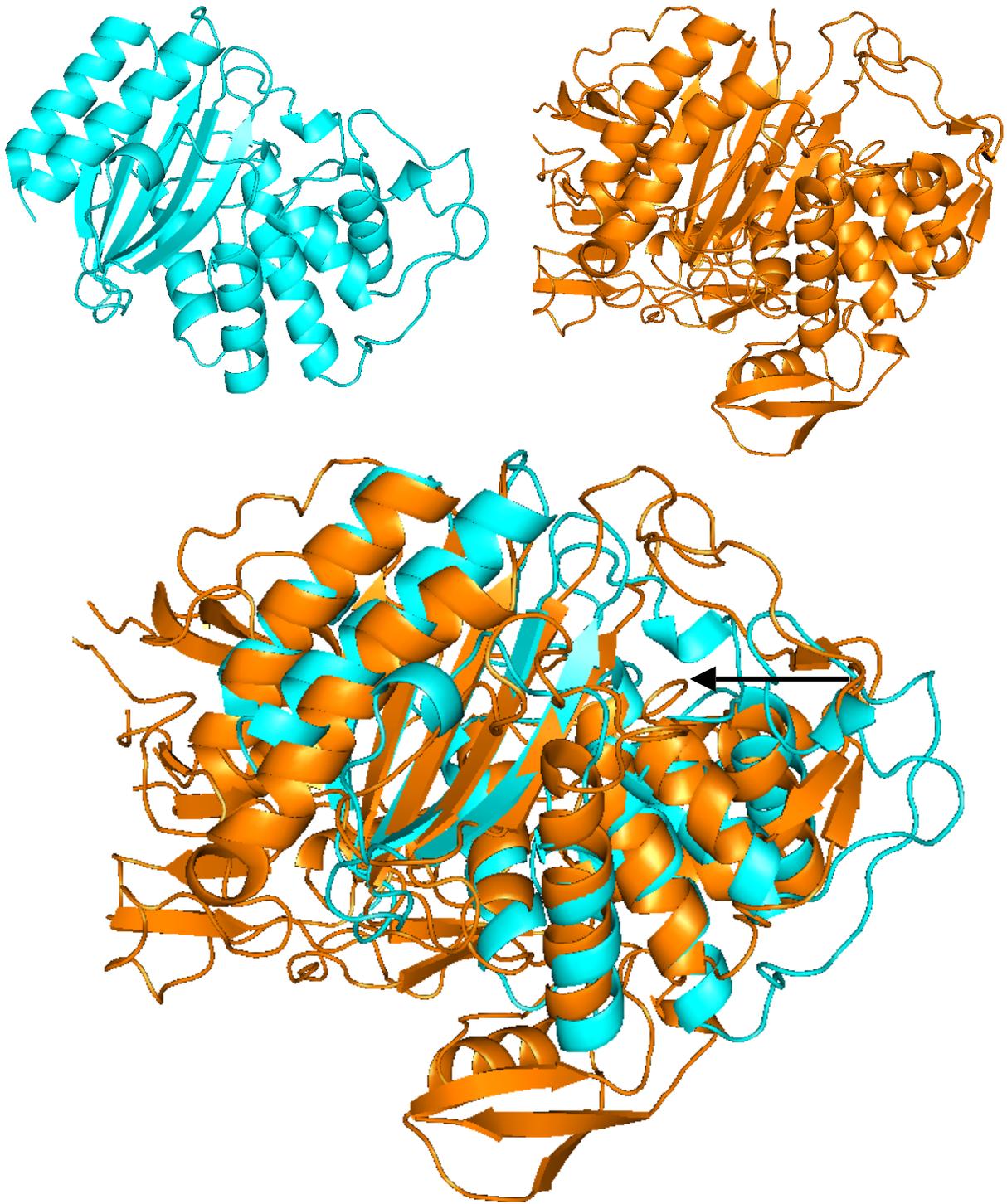


Figure 6.22: Structural comparison of a β -lactamase with a transpeptidase in Pymol. *Top left:* *E. coli* TEM1 β -lactamase (PDB: 1BTL)(Jelsch et al, 1993). *Top right:* *S. pneumoniae* PBPx2 transpeptidase (PDB: 1PYY)(Chesnel et al, 2003). *Bottom center:* Dali-lite pairwise super-imposition of 1PYY and 1BTL (Holm et al, 2008).

The two proteins share high structural similarity, with most α -helices and β -sheets superimposing to a high degree. Both also present the same-shaped opening to the active site, designated by .

However it is also possible a structural similarity in the absence of sequence similarity may be purely due to chance. It is believed that there may only be a limited number of protein folds (Govindarajan et al, 1999) and therefore there may be structural resemblances that do not have any relationship to a common function. Nevertheless in the case of BPSL1204, particularly because like BLIP-I the domains are duplicated, the resemblance seemed worth investigating further.

6.4 *E. coli* TEM-1 Inhibition Assay

6.4.1 Assay Design and Data Collection

Given the inter-species TEM1:BLIP-I complexes that had been previously solved (PDB 3gmw and 2g2w) and shown to exhibit inhibition, it was proposed that an assay to analyse possible inhibition of the *E. coli* TEM1 β -lactamase by BPSL1204-MA would be carried out, in order to see if it is a general β -lactamase inhibitory protein like BLIP-I. Following the procedure described in Section 3.8.1, data were collected across a range of BPSL1204-MA:TEM1 concentrations up to 4000:1, using a control assay of TEM1 and substrate without BPSL1204 and a control assay of substrate alone. Any increase in absorbance at a wavelength of 486 nm resulting from the hydrolysis of the β -lactam ring of the chromogenic β -lactamase substrate Nitrocefin (Figure 6.23) would be measured.

6.4.2 Inhibition Assay Results and Observations

The graph in Figure 6.24 represents the data collected during the inhibition assays. No absorbance was detected at λ 486nm in the absence of TEM1, whereas absorbance increased over time in the presence of only TEM1. The addition of BPSL1204-MA even to ratios as high as 4000:1 BPSL1204-MA:TEM1 resulted in no inhibition of TEM1, represented by a line that actually increased at a greater rate than that of the TEM1 only control. This slight increase is probably due to minute differences between experimentally applied substrate:enzyme ratios, perhaps due to pipette inaccuracies and the high sensitivity of the assay.

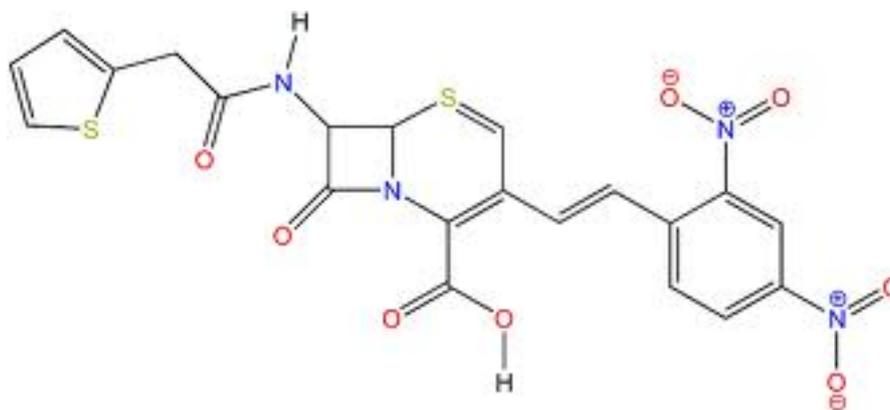


Figure 6.23: The chemical structure of Nitrocefin (picture from Calbiochem). Cleavage of the square β -lactam ring results in a change in absorbance from 380 nm to 486 nm.

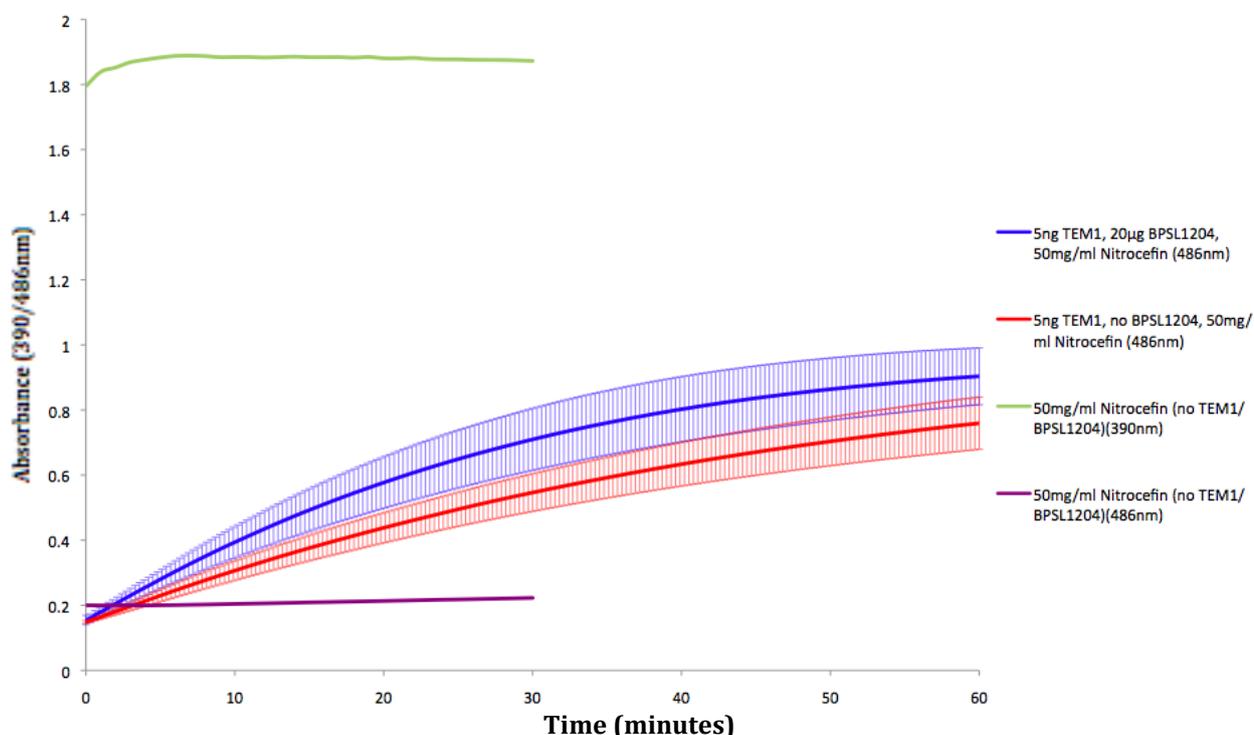


Figure 6.24: Line graph of Nitrocefin hydrolysis plotted over time in Microsoft Excel.

The graph showed that there was no hydrolysis of Nitrocefin in the absence of TEM1 (green and purple). Control time courses were only carried out for 30 minutes and repeated only once, due to a constant plateau being measured.

Upon introduction of TEM1 (red) hydrolysis was measured as an increase in absorbance at 486 nm, with a reading taken every 30 seconds for 60 minutes. However upon the addition of BPSL1204 in saturating concentrations (blue) no decrease in absorbance was seen when a reading was taken every 30 seconds for 60 minutes, indicative of lack of inhibition of TEM1. The error bars represent the standard deviation from 3 repeats of each experiment. Standard deviations regularly overlap in the above plot, and therefore the results of the experiment were considered as not being significantly different between experiments.

No further enzymatic measurements or calculations of TEM1 inhibition were carried out as it was considered apparent that BPSL1204 did not act in a BLIP-I-like fashion upon *E. coli* TEM1.

The conclusion that can be made is that BPSL1204-MA shows no sign of inhibiting the activity of the *E. coli* general β -lactamase TEM1.

6.4.3 Conclusions of the TEM1 Inhibition Assay

Regretfully, the assay was limited by the fact that the only conclusive result would be if inhibition did occur, otherwise the lack of inhibition allowed only further speculation. As it was seen that BPSL1204-MA does not inhibit *E. coli* TEM1, this led to several possible conclusions. Lack of inhibition suggested that BPSL1204-MA did not interact with this *E. coli* β -lactamase, but still may have the potential to act as an inhibitor of *B. pseudomallei*-specific β -lactamases. Additionally, the truncated BPSL1204-MA was used in the assay. Perhaps a full length BPL1204 may confer inhibition onto a β -lactamase target, but with the conserved glycerylcysteine lipoprotein LLAGCA motif immediately upstream of the truncation site, this seemed unlikely. Moreover, we had not successfully over-expressed the full-length BPSL1204 in a soluble form. Thirdly, BPSL1204 may interact with other penicillin-binding proteins such as cell-wall transpeptidases. Finally, it could be the case that BPSL1204 functions with a totally different, potentially novel biological function.

6.5 Implications of Structural Bioinformatics

Investigation of the crystallographic model of BPSL1204 revealed new information about this putative lipoprotein from *B. pseudomallei*. It was confirmed that BPSL1204 has two structurally similar domains that stack on top of one another and are connected by a seemingly rigid, short loop. A reasonably long 11-residue linker region exists between the first domain and the point of covalent linkage to a cell membrane, presumably conferring relatively flexible orientations whilst protruding from the membrane. However, although structural comparisons generated many observations that allow some speculation about the function of this previously unknown structure, no solid evidence for a function has been obtained. It was therefore decided that further biological experiments were required in order to establish the cellular function of BPSL1204.

Chapter 7: Biological Investigation of BPSL1204 Function

7.1 Design of a Pull Down Assay

7.1.1 *Burkholderia cenocepacia*

Due to the dangerous nature of *B. pseudomallei* (see Section 1.1) as a pathogenic Class III biohazard organism, it was not possible to perform experiments involving live cultures in any of the Class I and II laboratories available in Sheffield. Therefore experiments involving the elucidation of BPSL1204's biological function were carried out using the Class II paraphyletic species *Burkholderia cenocepacia* (strains HIII and J2315, provided by Dr Mark Thomas of the Sheffield Medical School). This close relative of *B. pseudomallei* is a plant pathogen (Mahenthiralingam et al, 2008), although it does also infect humans with cystic fibrosis (CF) to a high mortality rate (Jones et al, 2004). Of the seven Gram-negative *B. cepacia complex* species over 90% of the fatalities in *B. cepacia*-infected CF patients are caused by *B. cenocepacia* and *B. multivorans* (Mahenthiralingam et al, 2008). As a very close genetic relative of *B. pseudomallei*, *B. cenocepacia* also possesses a proteome in which many polypeptide amino acid sequences are identical or very similar. Therefore it was proposed that any interactions observed in *B. cenocepacia* homologs of *B. pseudomallei* proteins could allow for the inference of those proteins' biological functions within *B. pseudomallei*.

7.1.2 BPSL1204 and BCAL2351

By performing a BLAST search using the full-length amino acid sequence of BPSL1204 against the entire proteome of non-redundant sequences from *B. cenocepacia* (Strain J2315, isolated from a CF patient in Edinburgh), it was discovered that one protein, BCAL2351, had an Evalue of 3e-109 and 88% identity with BPSL1204. In addition to this, the lipoprotein glycerylcysteine membrane-anchoring motif LLAGCA was also present further showing that BCAL2351 was confidently a homolog of BPSL1204. An alignment of BPSL1204-MA was then carried out using BLAST against a similarly N-terminal truncated

BCAL2351 sequence in order to compare the soluble domains to be used in the experiment. This required the removal of 23 residues from the BCAL2351 N-terminus compared to the 20 residues removed from BPSL1204 (Figure 7.1). This resulted in a sequence with 92% identity to BPSL1204-MA. This meant that BCAL2351 is a clear protein homolog of BPSL1204, and is therefore extremely likely to have a soluble region with domains of identical structure to BPSL1204-MA.

7.1.3 Aims of the Assay

With the results of structural comparisons being inconclusive as to the function of BPSL1204, it was proposed that performing a pull-down assay using a purified and immobilised *B. cenocepacia* homolog of BPSL1204 washed with whole-cell *B. cenocepacia* lysate would lead to the isolation of proteins that biologically interact with the BCAL2351 homolog. The cellular processes that isolated binding-partners were involved with could then also allow for the discovery of the BPSL1204 function in *B. pseudomallei* through inference.

In addition to the discovery of biological binding partners, it was proposed that the results of a pull-down assay might aid in the confirmation or rejection of structural conclusions regarding BLIP-I. The use of whole-cell lysate would provide soluble target proteins from the cytoplasm and periplasm of the bacterial cell. In Gram-negative bacteria, β -lactamases are commonly found in the periplasm, therefore the assay would allow for the possibility of BCAL2351 binding to enzymes that are related to β -lactamases.

7.2 N-terminal His-tagged BCAL2351

7.2.1 pET28a Cloning and Transformation

To facilitate immobilisation of BCAL2351-MA, it was originally decided to clone a 6xHis-tag to the N-terminus of the truncated protein for ionic attachment to a Ni-NTA column. Primers were designed to introduce a 5' 6xHis-tag nucleotide

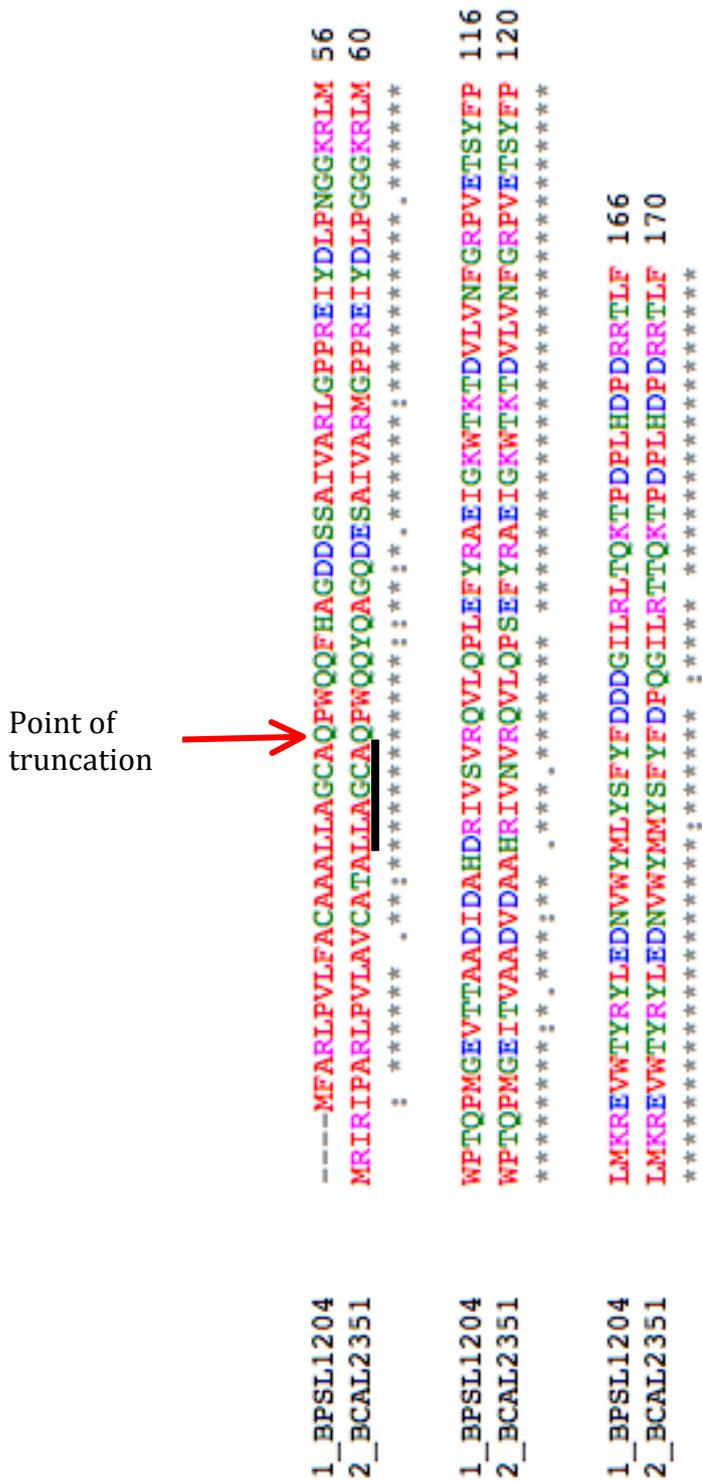


Figure 7.1: Sequence alignment of BPSL1204 and BCAL2351 in ClustalW. Colours represent residues with similar chemical properties.

Most residues are identical (*), with most differences existing as conservative substitutions (:). Both proteins share the LLAGCA motif.

sequence with a 10-residue linker region when cloned into pET28a, generating a construct with a 51 bp addition to the 444 bp BCAL2351-MA. PCR was carried out following the protocol described in Section 3.2.1, using an annealing temperature of 55°C and 0.5 µl of a single *B. cenocepacia* colony boiled in 30 µl sterile water (Figure 7.2). The PCR product was then restriction cloned into pET28a using NdeI and BamHI, using the protocol described in Section 3.4.2. A ligation was carried out and then sub-cloned into DH5α, with successful clones isolated using colony PCR, confirmed by plasmid DNA sequencing and transformed into the BL21(DE3) over-expression strain, following the methods described in Sections 3.4.2 and 3.4.3 respectively.

7.2.2 Insolubility in BL21 (DE3) Cells

A BL21(DE3) pET28a:bcal2351-MA colony was cultured overnight and subjected to 5-50 ml over-expression trials, optimised around the standard method described in Section 3.5.3. Extensive variations of conditions such as inducing temperature (4-37°C), length of over-expression time (2-24 hours), concentration of inducing agent (0.1-2 mM IPTG in 0.5 mM steps) and method of cell lysis (sonication and BugBuster fractionation) were employed, but all analysis by SDS-PAGE (Figure 7.3) showed that any protein that was being produced existed entirely in an insoluble state. This suggested that as BCAL2351-MA and BPSL1204-MA were 92% identical in sequence, then the resulting insolubility might have been due to the modified N-terminal His-tag contributing an extra 16 residues (an additional 11%) of 'alien' polypeptide. It was decided to attempt to refold the protein.

7.2.3 Inclusion Body Preparation and Refolding Assays.

Cell-paste containing insoluble His-tagged BCAL2351-MA was produced using the media and method described in Section 3.5. Protein over-expression was induced for 4 hours at 37°C using 1 mM IPTG. 5 g of cells were resuspended in 50 ml of 50 mM Tris pH 8.0, sonicated on ice for 3x20 seconds at 16 microns, and the insoluble fraction isolated into a pellet by centrifugation in a JA25.50 rotor

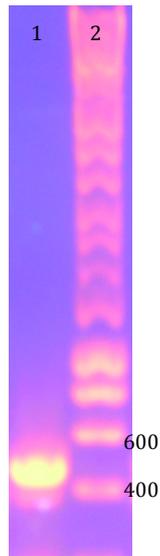


Figure 7.2: 1% agarose gel electrophoresis of N-his BCAL2351-MA PCR product. Lane 1 = PCR product. Lane 2 = Hyperladder I (bp). A pure band is seen at the expected molecular weight of 495 bp (444 + 51 bp).

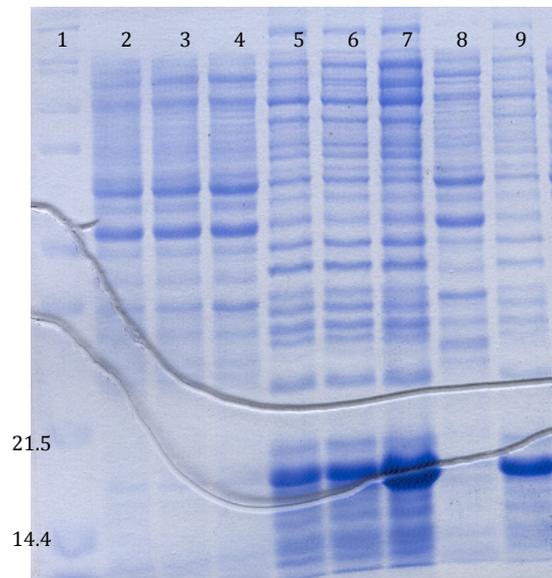


Figure 7.3: SDS-PAGE analysis of N-hisBCAL2351-MA over-expression, at 1mM IPTG for 4 hours. Lane 1 = Mark12 (kDa), Lane 2-4 = No induction, 15 and 37°C soluble Bugbuster fractionation. Lane 5-7 = No induction, 15 and 37°C insoluble Bugbuster fractionation. Lane 8+9 = 25°C, fractionation by sonication, soluble and insoluble respectively.

Lane 7 represents the typical amount of insoluble N-hisBCAL2351-MA produced from many variations of time, [IPTG], temperature and fractionation method. No soluble protein was ever produced.

for 15 minutes at 24500 rpm at 4°C. The pellet was resuspended with sonication in 50 mM Tris pH 8.0, 0.1 M NaCl and 0.5% Triton X-100, and then re-pelleted at 10000 rpm for 20 minutes in a JA25.50 rotor to collect the inclusion bodies. This step was repeated 2 further times, in the absence of Triton X-100, to wash away all cell-debris and purify inclusion bodies into a pellet. The final inclusion body pellet was resuspended in 2 ml of 8 M urea, with the protein concentration then established by a Bradford Assay to allow for the addition of 8 M urea to give a final protein concentration of 1 mg/ml.

This sample was then applied to the QuickFold kit, available from Athenaes Enzyme Systems, and the screening protocol in Application Manual V.2 supplied with the kit was followed exactly for all of the 15 conditions provided. Analysis of the refolding screen was carried out by SDS-PAGE of samples of the soluble supernatants after the last microcentrifugation step of the manufacturers protocol. As can be seen in Figure 7.4, results of the screen suggest refolding Condition 11 (50 mM Tris.HCl pH 8.5, 9.6 mM NaCl, 0.4 mM KCl, 1 mM EDTA, 0.5% Triton X-100, 1 mM DTT) resulted in at least some soluble, refolded protein. However, upon optimisation to remove EDTA (to improve binding to a Ni-NTA column by reducing the Ni²⁺-chelating effect of EDTA presence) and Triton X-100 (a strong detergent that may have been difficult to remove in wash steps and may have perturbed biological interactions), it was not possible to obtain soluble protein as protein visibly precipitated, and the supernatant post-microcentrifuge contained no protein as assessed by SDS-PAGE.

The remaining majority of the 8 M urea-denatured N-His BCAL2351-MA was scaled up and treated to a 1:20 dilution using Condition 11, as applied in the QuickFold protocol. This was then spun down at 20000 rpm at 4°C for 20 minutes in a JA25.50 rotor to remove permanently insoluble material, and then the soluble, N-His BCAL351-MA-containing supernatant was concentrated to a final volume of 2 ml of 2 mg/ml protein. This was then run on a gel-filtration column, following the method described in Section 3.6.4, to assess stoichiometric state and to confirm BPSL1204-MA-like soluble behaviour in the absence of an activity assay. As can be seen in Figure 7.5, the gel filtration profile generated



Figure 7.4: SDS-PAGE analysis of QuickFold refolding experiment. Lane 1 = Mark12 (kDa). Lane 2-8 = Condition 9-15 soluble fraction. Lane 9 = Soluble N-hisBCAL2351 in 8 M urea.

The gel analysis suggested that N-hisBCAL2351 could be refolded using QuickFold Condition 11 (Lane 4), compared to the 8 M solubilised protein in Lane 9. Conditions 1-8 gave no soluble protein on SDS-PAGE.

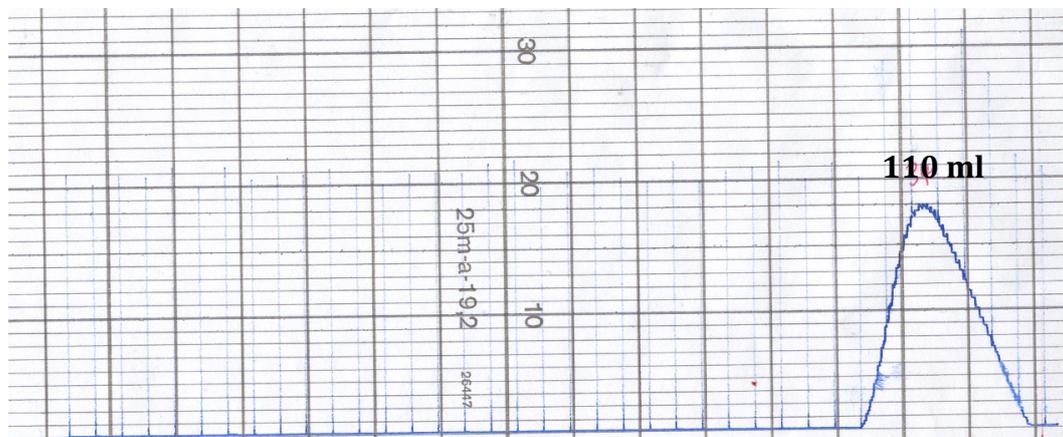


Figure 7.5: Gel-filtration trace of N-hisBCAL2351 post-refolding optimisation. The peak runs off the column after 110ml in Fraction 34 (BPSL1204 ran off after 100ml) indicating a folded monomer of N-hisBCAL2351, however the sample did not bind a HisTrap column.

was very similar to that of pure, monomeric BPSL1204-MA as documented in Section 4.4.4 (Figure 4.8). However, when the 6 ml fractions under the gel-filtration peak, now buffer-exchanged into 50 mM Tris pH 8.0 and 0.5 M NaCl via the gel-filtration process, were applied to a fully charged 1 ml His-Trap Ni-NTA column there was no binding of N-His BCAL2351-MA to the His-Trap column. It was therefore concluded that although N-His BCAL2351-MA seemed to have refolded post-urea denaturing, it had mis-folded to adopt a soluble, monomeric conformation that buried the 6xHis-tag and therefore could not be a biological conformation similar to that of BPSL1204, negating the use of N-His BCAL2351-MA in any possible pull-down assay.

7.3 Affi-Gel15 Pull Down Assay

7.3.1 Design of a Second Pull-down Method

Giving the insolubility of BCAL2351-MA (N-His construct described in Section 7.2.2 and without a tag described in Section 6.3.3), a different approach was decided upon. With the aim of the pull-down to model BPSL1204 interactions using *B. cenocepacia* homologs, it was proposed that BPSL1204 would interact with the same proteins as BCAL2351 from *B. cenocepacia* cell-free extract. Therefore with soluble, untagged BPSL1204-MA already obtainable, a method for immobilisation was required. It was decided that immobilisation of the acidic BPSL1204-MA, with a theoretical pI of 5.32 as calculated by the online ExPASy ProtParam tool, to Affi-Gel 15 cross-linked agarose beads available from Bio-Rad would allow the purification of binding partner(s) from cell-free *B. cenocepacia* extract via bead affinity chromatography.

7.3.2 Culturing *B. cenocepacia* and Safe Practice

The Class II Thomas laboratory routinely used *B. cenocepacia* to microbiologically and biochemically investigate aspects of infection of CF patients. Therefore all cultures were carried out under their supervision using that lab's equipment. Following the LB media preparation and culturing methods described in Section 3.5, the protocol applied to *E. coli* cultures was used for

growing *B. cenocepacia* with the exception of using a lower shaking speed of 150 rpm to reduce aerosol production. LB media was supplemented with 100 mg/ml ampicillin, to which *B. cenocepacia* is highly resistant. All work with whole cells, in culture or in a pellet, was carried out in a BioMat2 Class II Microbiology Safety Cabinet with laminar airflow, sealable protective screen and UV sterilisation capabilities, with all waste media and equipment used autoclave-sterilised before leaving the Class II lab environment. For the initial assays, *B. cenocepacia* was cultured overnight (16-18 hours) at 37°C.

7.3.3 First Pull-Down Assay

Pure BPSL1204-MA was immobilised to the Affi-Gel 15 cross-linked agarose beads via N-hydroxysuccinimide esterification of free primary amines, such as lysine side-chains, upon combination in an aqueous environment. 50 mM HEPES pH 7.5 was used rather than Tris, to prevent swamping of the beads with primary amines. After bait-protein immobilisation, excess hydroxysuccinimide moieties were blocked by the addition of ethanolamine. Cell-free extract was prepared by sonication of *B. cenocepacia* cell-paste in 50 mM HEPES pH 7.5, followed by centrifugation to remove insoluble proteins and cell-debris. Cell-free extract was then used to wash the Affi-Gel 15:BPSL1204-MA beads, which were then washed in low salt concentration to remove non-binding proteins before elution of binding partners was achieved by high salt concentration elution or boiling in LDS. All steps were applied to a non-*Burkholderia* control protein, the 34 kDa YloQ from *Bacillus subtilis*, (Cladiere et al, 2006; Ruzheinikov et al, 2004) to account for random Affi-Gel 15:immobilised protein background. The full details of the protocol used can be found in Section 3.8.2.

7.3.4 Analysis of Isolated Fractions By SDS-PAGE

Samples of bait proteins, cell-free extract before the assay and all fractions taken during the assay were run on a 4-12% NuPAGE gradient SDS-PAGE gel, available from Invitrogen, to identify the molecular weights of proteins present in each step of the experiment. This was stained in colloidal Coomassie G-250 InstantBlue protein stain/destain, to allow for 'from-gel' mass spectrometry at a

later stage, as standard non-colloidal Coomassie R-250 is not compatible with mass spectrometric methods. As can be seen in Figure 7.6, in both LDS-boiled beads and both NaCl elutions there was a band present at ~80 kDa that was not present in the YloQ control lanes. This band represented a potential protein binding-partner for BPSL1204-MA. A massive background of contamination in all samples was witnessed in Figure 7.6, perhaps due to non-specific interactions with the agarose beads regardless of the ethanolamine blocking reaction. However, this contamination was present in all control lanes too, thus still allowing comparisons for potential binding-partners of BPSL1204.

Inspection of YloQ control lanes shows that YloQ was also present in the elutions (due to *in vitro* dimeric nature leading to salt-induced dissociation of subunits) that may mask other potential binding partners that have mass of ~34 kDa. It was decided in future assays to use the monomeric BPSL1549 (Cruz-Migoni et al, 2011) as a control protein.

7.3.5 Second Pull-Down Assay

In order to corroborate the results of the original assay, the pull-down was repeated again using the *B. pseudomallei* BPSL1549 protein as a control, chosen as it was monomeric and also of interest to see if BPSL1549 would pull-down *B. cenocepacia* proteins different to those of BPSL1204. Empty beads blocked only with ethanolamine were used as a second control, to account for the background attributable to non-specific interactions with the beads. The method described in Section 3.8.2 was employed and boiling in LDS or ionic perturbation by 2 M NaCl was used to elute final fractions.

NuPAGE 4-12% SDS-PAGE analysis of the resulting fractions and controls (Figure 7.7) revealed that the ~80 kDa band discovered in the original Affi-Gel 15:BPSL1204-MA pull-down (Section 7.3.4) was once again isolated from the cell-free extract, along with the large quantity of background non-specific/bead-

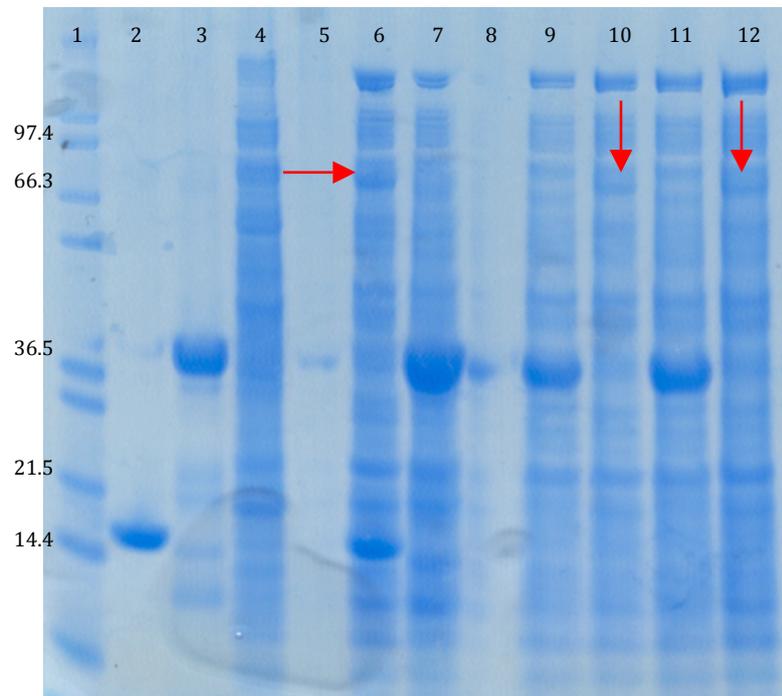


Figure 7.6: SDS-PAGE analysis of first pull-down.

- Lane 1 = Mark12 (kDa).
- Lane 2 = Purified BPSL1204.
- Lane 3 = Purified YloQ
- Lane 4 = *B. cenoepecia* cell-free extract
- Lane 6 = BPSL1204 LDS-boiled eluate
- Lane 7 = YloQ LDS-boiled eluate
- Lane 9= YloQ 1.5 M NaCl eluate
- Lane 10 = BPSL1204 1.5 M NaCl eluate
- Lane 11 = YloQ 2 M NaCl eluate
- Lane 12 = BPSL1204 2 M NaCl eluate
- Lane 5 + 8 = Empty

Many contaminating proteins were pulled out in all experiments. A single band was pulled out of the cell-extract in the BPSL1204 pull-down only, at ~80 kDa, using either LDS or NaCl elution (→). The presence of YloQ in control lanes is due to the dimeric nature of the control protein.

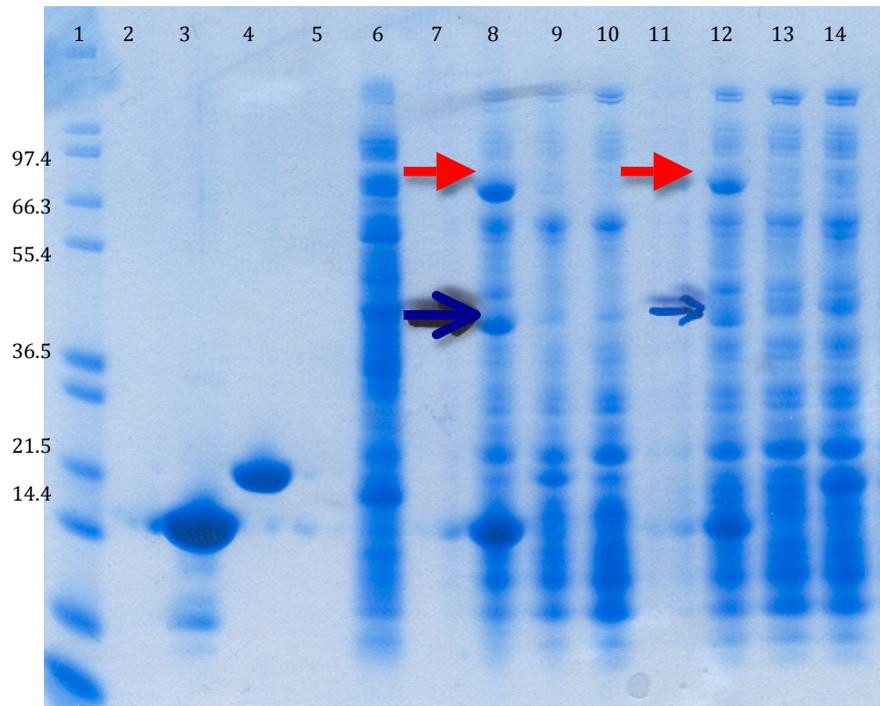


Figure 7.7: SDS-PAGE analysis of second pull-down.

Lane 1 = Mark12 (kDa)
 Lane 3 = Purified BPSL1204
 Lane 4 = Purified BPSL1549
 Lane 6 = *B. cenocepacia* cell-free extract
 Lane 8 = BPSL1204 LDS eluate
 Lane 9 = BPSL1549 LDS eluate
 Lane 10 = Empty beads LDS eluate
 Lane 12 = BPSL1204 2 M NaCl eluate
 Lane 13 = Empty beads 2 M NaCl eluate
 Lane 14 = BPSL1549 2 M NaCl eluate
 Lanes 2, 5, 7 and 11 = empty lanes

The ~80 kDa band seen in the first pull down assay using BPSL1204 was replicated in the second pull-down assay (lane 8 and 12, ). In addition, a further band was isolated at around ~45 kDa in the BPSL1204 pull-down only (.

The BPSL1549 control beads did not pull down any bands that did not appear in the empty beads control.

interacting contaminants present in the control samples. However, in these experiments an additional band of ~45 kDa was seen to be present only in the BPSL1204 lanes. This secondary band however appeared to contribute to significantly less of the mass of isolated proteins in the BPSL1204 pull-down samples than the twice isolated ~80 kDa band, and was not witnessed in the first pull-down assay that had used exactly the same cell-free extract sample and same elution methods as the second assay.

7.3.6 Identification of Proteins 'Pulled-Out' of Cell Extract

The colloidal-Coomassie-stained gels were subjected to band excision, reduction via DTT and alkylation via iodoacetamide of the protein-gel slice followed by overnight tryptic digestion. Separation of protein from the gel was achieved by systematic addition of ammonium bicarbonate with increasing concentrations of acetonitrile, followed by treatment with formic acid (Shevchenko et al, 2006). Finally MALDI-TOF mass spectrometry of the resulting trypsin-digested peptide fragments was performed (Bonk & Humeny, 2001). These procedures were performed by Dr Mark Dickman of the University of Sheffield ChELSI (Chemical Engineering at the Life Science Interface) Department of Chemical and Biological Engineering. Both of the isolated bands unique to the Affi-Gel 15:BPSL1204 lanes from the second pull-down assay were investigated, alongside the corresponding region in the adjacent control lanes to account for non-specific background components. The masses of the ionised trypsin-fragments from each of the 2 identified bands were independently input into the MASCOT server (Cottrell, 2012) (www.matrixscience.com/server.html), which calculated which translated protein sequences from the *B. cenocepacia* J2315 genome were most likely to generate the observed trypsin-fragment masses.

The mass spectrometry analysis revealed that the ~80 kDa band in only BPSL1204 pull-downs was a Methionyl tRNA Synthetase with 97 matching fragments and a score of 2159 (Figure 7.8). However the experiment performed on the ~45 kDa band gave a less confident result, with three putative protein hits of Glutamate-1-semialdehyde 2,1-aminomutase with 12 fragments scoring 293,

[SYM_BURCJ](#) Methionyl-tRNA synthetase OS=Burkholderia cepacia (strain J2315 / LMG 16656) GN=metG PE=3 SV=1
[SYM_BURXL](#) Methionyl-tRNA synthetase OS=Burkholderia xenovorans (strain LB400) GN=metG PE=3 SV=1
[MNMG_BURCJ](#) tRNA uridine 5-carboxymethylaminomethyl modification enzyme mnmG OS=Burkholderia cepacia (strain J2315 / LMG 16656) GN=mnmG PE=3 SV=1
[SYM_BURMA](#) Methionyl-tRNA synthetase OS=Burkholderia mallei GN=metG PE=3 SV=1
[PNP_BURCA](#) Polyribonucleotide nucleotidyltransferase OS=Burkholderia cenocepacia (strain AU 1054) GN=pnp PE=3 SV=1
[K2C1_HUMAN](#) Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6
[K1C10_HUMAN](#) Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6
[MNMG_BORPD](#) tRNA uridine 5-carboxymethylaminomethyl modification enzyme mnmG OS=Bordetella petrii (strain ATCC 35061) GN=mnmG PE=3 SV=1
[PNP_HERAR](#) Polyribonucleotide nucleotidyltransferase OS=Herminiimonas arsenicoxydans GN=pnp PE=3 SV=1
[K2M2_SHEEP](#) Keratin, type II microfibrillar, component 7C OS=Ovis aries PE=1 SV=1
[SYM_ACIAD](#) Methionyl-tRNA synthetase OS=Acinetobacter sp. (strain ADP1) GN=metG PE=3 SV=1
[K1C9_HUMAN](#) Keratin, type I cytoskeletal 9 OS=Homo sapiens GN=KRT9 PE=1 SV=3
[MNMG_POLNS](#) tRNA uridine 5-carboxymethylaminomethyl modification enzyme mnmG OS=Polynucleobacter necessarius

SYM_BURCJ **Mass:** 79143 **Score:** 2159 **Queries matched:** 97 **emPAI:** 3.89
 Methionyl-tRNA synthetase OS=Burkholderia cepacia (strain J2315 / LMG 16656)

[K2C1_HUMAN](#) Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6
[GSA_BURCM](#) Glutamate-1-semialdehyde 2,1-aminomutase OS=Burkholderia ambifaria (strain J2315 / LMG 16656) GN=gsmA PE=3 SV=1
[TOLB_BURCA](#) Protein tolB OS=Burkholderia cenocepacia (strain AU 1054) GN=tolB PE=3 SV=1
[TRYP_PIG](#) Trypsin OS=Sus scrofa PE=1 SV=1
[ENO_BURCA](#) Enolase OS=Burkholderia cenocepacia (strain AU 1054) GN=eno PE=3 SV=1
[K1C10_HUMAN](#) Keratin, type I cytoskeletal 10 OS=Homo sapiens GN=KRT10 PE=1 SV=6
[RS19_METS5](#) 30S ribosomal protein S19P OS=Metallosphaera sedula (strain ATCC 5136) GN=RS19 PE=3 SV=1
[Y1082_PASMU](#) Uncharacterized protein PM1082 OS=Pasteurella multocida GN=PM1082 PE=4 SV=1
[UTP20_MOUSE](#) Small subunit processome component 20 homolog OS=Mus musculus GN=Utp20 PE=3 SV=1
[K2C6B_HUMAN](#) Keratin, type II cytoskeletal 6B OS=Homo sapiens GN=KRT6B PE=1 SV=5
[K2C8_MOUSE](#) Keratin, type II cytoskeletal 8 OS=Mus musculus GN=Krt8 PE=1 SV=4
[GSA_SHEAM](#) Glutamate-1-semialdehyde 2,1-aminomutase OS=Shewanella amazonensis (strain ATCC 35061) GN=gsmA PE=3 SV=1
[MNME_FINM2](#) tRNA modification GTPase mnme OS=Fingoldia magna (strain ATCC 29328) GN=mnme PE=3 SV=1
[LPXB_BURCC](#) Lipid-A-disaccharide synthase OS=Burkholderia cenocepacia (strain AU 1054) GN=lpxB PE=3 SV=1
[LPXB_BURP8](#) Lipid-A-disaccharide synthase OS=Burkholderia phymatum (strain DSM 171) GN=lpxB PE=3 SV=1
[SFSA_PYRFU](#) Sugar fermentation stimulation protein homolog OS=Pyrococcus furiosus GN=sfsA PE=3 SV=1
[L_HENDV](#) Large structural protein OS=Hendra virus GN=L PE=3 SV=2

GSA_BURCM **Mass:** 44931 **Score:** 293 **Queries matched:** 12 **emPAI:** 0.34
 Glutamate-1-semialdehyde 2,1-aminomutase OS=Burkholderia ambifaria (strain J2315 / LMG 16656)

TOLB_BURCA **Mass:** 45757 **Score:** 158 **Queries matched:** 7 **emPAI:** 0.24
 Protein tolB OS=Burkholderia cenocepacia (strain AU 1054) GN=tolB PE=3 SV=1

ENO_BURCA **Mass:** 45658 **Score:** 76 **Queries matched:** 3 **emPAI:** 0.15
 Enolase OS=Burkholderia cenocepacia (strain AU 1054) GN=eno PE=3 SV=1

Figure 7.8: Results of MASCOT server analysis of mass spectroscopic trypsin-digested pull-down targets. *Top:* List of proteins detected in the ~80 kDa band. *Bottom:* List of proteins in the ~45 kDa band.

The most convincing hit of all detected was the 79.1 kDa Met-tRNA Synthetase. Gel slices taken from the corresponding regions of the control lanes revealed only contaminating human proteins

TolB with 7 fragments scoring 158 and Enolase with 3 fragments scoring 76. Various keratins discovered in most experiments were presumably a product of human skin or hair contamination either during the gel staining/destaining process or the gel-excision procedure.

7.3.7 Choice of Target Protein

Given the large number of fragments generated, increasing the likelihood of locating the correct protein sequence, and that any score under 75 is considered insignificant, the 79.1 kDa methionyl-tRNA synthetase was the most confident result from all the mass spectrometry outputs. In addition to this, this protein was the only one present in both pull-down assays and in the greatest abundance in the samples that were loaded onto the original SDS-PAGE gels. Enolase was discounted immediately due to its very low score and fragment count. The glutamate aminomutase was also discounted as an artifact of random interaction due to its low score and low fragment count.

TolB scored low also, but is a periplasmic protein involved in uptake of colicin A by the Tol/Pal system (in *E. coli*) (Walburger et al, 2002). As β -lactamases and cell-wall transpeptidases are thought to exist in the periplasm, this may be more relevant both in relation to the BLIP similarity and also in relation to the subcellular location expected of BPSL1204. However, the apparently low binding capability to BPSL1204, as seen by only appearing in 1 out of 2 pull-down assays in lower abundance than the methionyl-tRNA synthetase, combined with a low MASCOT score and fragment count made the odds of a stable biological interaction improbable. In addition to experimental results, the Tol/Pal system is well established (Vianney et al, 1996) with no reported relation to proteins with domain-folds similar to BLIP-I.

It was decided to investigate the seemingly strong interaction between methionyl-tRNA synthetase and BPSL1204. It was considered whether BPSL1204 might act as a regulatory protein of methionyl-tRNA synthetase function, or aid in the delivery of one or both of the enzymes' tRNA and

methionine substrates. This did however seem unlikely as tRNA synthetases are cytoplasmic but BPSL1204 is likely to be covalently anchored to a membrane. Alternatively BPSL1204 may not be a true lipoprotein, or that the interactions discovered by the Affi-Gel 15 pull-down assays may have been strong but non-specific and non-biological. It was decided that experiments investigating the ability of pure BPSL1204-MA and pure methionyl-tRNA synthetase to form a stable, soluble complex were required, in tandem with assays for tRNA synthetase activity. This would also allow for crystallographic investigation of any potential biological interactions.

7.4 BPSL0998

7.4.1 Novel Features of BPSL0998

Investigation of the genetic origin of the methionyl-tRNA synthetase using the BLAST databases revealed it to be the gene product of BCAL2646 in *B. cenocepacia*. As the source of the BPSL1204 bait-protein under investigation was from *B. pseudomallei*, the genome of *B. pseudomallei* was searched for BCAL2646 homologs, uncovering the 88% identical BPSL0998 with a match of Evalue = 0.0. Figure 7.9 shows the sequence of BCAL2646 aligned with BPSL0998. Upon BLAST investigation of the BPSL0998 sequence compared to the *E. coli* equivalent in the PDB (PDB: 3H9C, used as a potential homology-model for Molecular Replacement) it was discovered that the N-terminal and central region (residues 1-514) of BPSL0998 has 55% identity with *E. coli* Met-tRNA synthetase, however the C-terminal region of the BPSL0998 sequence (residues 540-637) lacked any homology to the *E. coli* protein (Figure 7.10).

After a 26-residue 'linker' region, the sequence of this region has 48% identity to the Trbp111 structure-specific tRNA-binding protein from *Aquifex aeolicus* (Swairjo et al, 2000)(Figure 7.10). Individual Trbp111 protein is proposed to dimerise, bind and stabilise the right-angled T-loop common to all tRNA macromolecules (Li et al, 1996). Apart from any BPSL1204 interest, this meant that BPSL0998 was interesting, as the presence of a single structure-specific tRNA-binding domain added to a protein such as a met-tRNA synthetase that

already binds tRNA as a catalytic substrate was an interesting situation, which might be investigated by X-ray crystallographic studies.

7.4.2 pBAD/hisB Cloning of BPSL0998

In order to produce and purify BPSL0998, the gene was PCR amplified using the primers described in Section 3.4.1 and a self-designed touchdown method based upon the general protocol in Section 3.2.1, due to difficulties in purity and amount of amplification. This required a 10-fold dilution of *B. pseudomallei* genome, twice the concentration of primers described in the standard method and a final 10% DMSO concentration. Then 20 cycles of 5x standard PCR loops were employed, with each cycle dropping the annealing temperature by 1°C from 70-50°C. This allowed for swamping of the genomic template with primers, and by starting the reaction above the annealing temperature and touching down the annealing it allowed the very specific primers to bind either side of the large 2178 bp gene instead of the partially specific, smaller-fragment binding sites that were leading to unfavourable contaminants in the original PCR reaction (Figure 7.11).

Amplified genes were then purified by gel-extraction and subjected to restriction cloning into the pBAD/hisB vector (available from Invitrogen) using NcoI and BamHI following the method described in Section 3.4.2. After selection of a successful DH5α sub-clone by colony PCR (Figure 7.11), the pBAD/hisB:BPSL0998 construct was transformed into the arabinose-inducible Top10 *E. coli* cell-line.

7.4.3 Insolubility in Top10 Cells

Extensive over-expression tests were performed upon the Top10:pBAD/hisB:BPSL0998 cell-line in LB media and 100 µg/ml carbenicillin, including variations in temperature (4-37°C), length of over-expression (3-30 hours), method of cell-lysis (sonication and BugBuster fractionation) and concentration of arabinose inducing-agent (1×10^{-5} – 0.1 %). All tests were

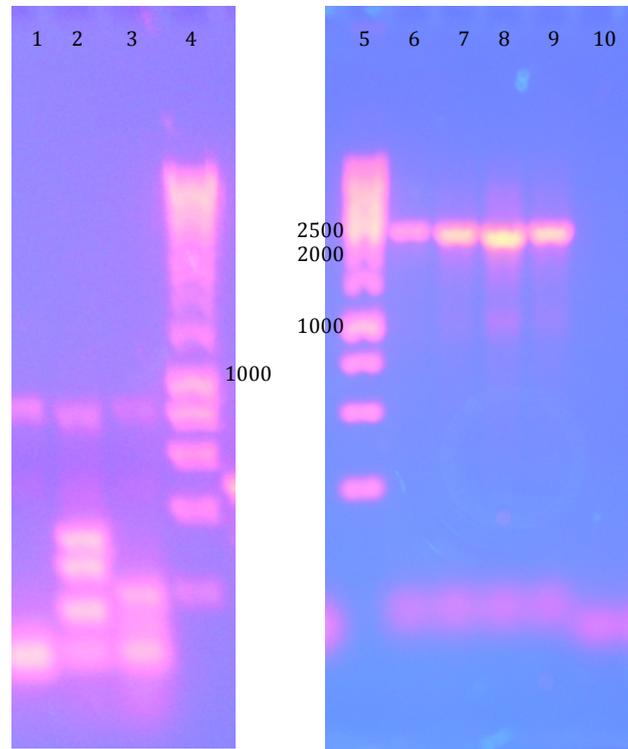


Figure 7.11: 1% gel electrophoresis of BPSL0998 cloning stages.

Lane 1-3 = Standard PCR method, 50, 55 and 58°C annealing.

Lane 4 = Hyperladder I (bp).

Lane 5 = 1Kb ladder (bp).

Lane 6 = Pure BPSL0998 (2178 bp) PCR product using touchdown.

Lane 7-10 = Colony PCR products.

High contamination by non-specific low MW fragments was seen using standard PCR methods, so more intricate techniques were applied to achieve the purity seen in Lane 6. Of only four colonies isolated after pBAD/hisB:BPSL0998 transformation into DH5, three generated positive BPSL0998 fragments.

analysed by SDS-PAGE, and as can be seen in Figure 7.12 BPSL0998 was heavily over-expressed, but was only ever produced into the insoluble fraction of *E. coli* Top10 cells regardless of over-expression optimisation attempts.

7.5 Attempts to Purify BCAL2646 by Bulk Pull Down

7.5.1 BCAL2646 and BPSL0998 are Homologs

Following the discovery that over-expressed BPSL0998 was insoluble in *E. coli* Top10, it was decided that as *B. pseudomallei* BPSL0998 and *B. cenocepacia* BCAL2646 are homologs (Section 7.4.1) and BCAL2646 had already been isolated as a soluble protein from *B. cenocepacia* cell-extract in small-scale pull-down assays, then it might be possible to purify BCAL2646 via a large-scale pull-down assay. Therefore interactions between a cross-species complex were investigated.

7.5.2 Log Phase and Overnight Culturing of *B. cenocepacia*

Given the biological role of BPSL0998 and BCAL2646 as methionyl-tRNA synthetases, it was proposed that BCAL2646 would be more abundant in *B. cenocepacia* cultures that were currently in a log-phase of growth. As at this stage of rapid cell proliferation protein synthesis in each individual cell would be at the highest rate in the cell cycle, it was anticipated that this would allow for a greater BCAL2646 concentration per cell lysed. However, in order to replicate the results already witnessed in the original assays, overnight cultures were also prepared as in Section 7.3.2.

7.5.3 Scaling up of the Assay

Based upon the protocol described in Section 3.8.2, the 50 µl assays were scaled up 20-fold in all respects to a final Affi-Gel15 bead slurry volume of 1 ml. The large-scale pull-down assay was carried out twice, using either log-phase *B. cenocepacia* or cells from an overnight culture. The only alteration to the

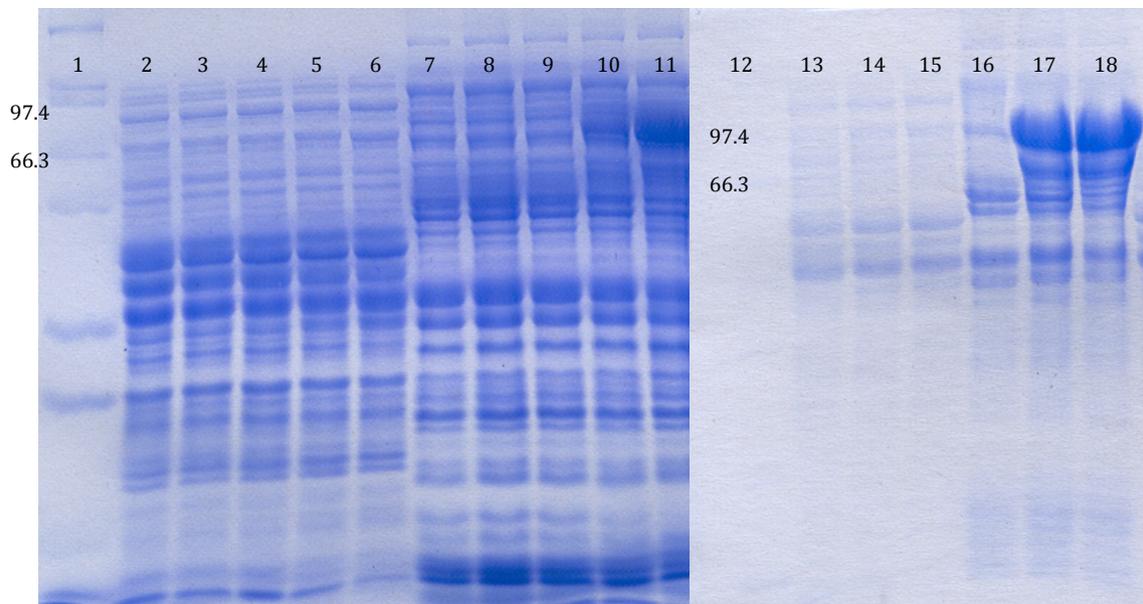


Figure 7.12: SDS-PAGE analysis of BPSL0998 over-expression in the pBAD system.

Lane 1 + 12 = Mark12 (kDa)

Lane 2-6 = 0, 0.00001, 0.0001, 0.001, 0.01 % arabinose soluble, 3 hrs at 37°C

Lane 7-11 = 0, 0.00001, 0.0001, 0.001, 0.01 % arabinose insoluble, 3 hrs at 37°C

Lane 13-15 = 4, 25, 37°C, 6 hrs at 0.1 % arabinose, soluble

Lane 16-18 = 4, 25, 37°C, 6 hrs at 0.1 % arabinose, insoluble

All variations in over-expression attempts resulted in heavy over-expression of the 79.1 kDa BPSL0998, but only in the insoluble fraction. No soluble protein was ever obtained.

previously used procedure was that 20 mg of pure BPSL1204 in 50 mM HEPES pH 8.0 were left to bind at 4°C on a slowly rotating wheel for 3 days. Of this, only approximately 8.5 mg of BPSL1204 bound, judged by the post-binding supernatant containing 7.8 mg protein, and each of the two wash-step supernatants containing 3.1 and 0.5 mg of protein respectively. This left a protein concentration of approximately 20 mg/ml of AffiGel15:BPSL1204 beads.

Variation of the protocol in Section 3.8.2 was carried out to optimise the bulk collection of BCAL2646. 6.5 g of *B. cenocepacia* cell-paste from the log-phase culture was resuspended in 13 ml of 0.1 M NaCl and 50 mM HEPES pH 7.5, sonicated and the insoluble fraction pelleted by centrifugation, leaving approximately 360 mg of soluble, *B. cenocepacia* cell-free protein extract. The 13 ml was used to resuspend Affi-Gel15:BPSL1204 1 ml of beads in a 15 ml Falcon tube, and was allowed to bind to the beads for three hours on a rotating wheel at 4°C. After washing of unbound material by 2 x 15ml of 50 mM HEPES pH 7.5 in 0.2 M NaCl, 100 µl of 2 M NaCl in 50 mM HEPES pH 7.5 was used to elute bound protein following a 1 minute incubation and a gentle centrifugation step to leave BCAL2646-containing supernatant. This elution step was performed three times to totally elute bound protein, and the entire process was then repeated using the overnight-cultured *B. cenocepacia* cells. However very little protein was actually obtained (Table 7.1), with little difference witnessed between log-phase and overnight cultures, although SDS-PAGE analysis of 25 µl of each eluate (Figure 7.13) showed that BCAL2646 constituted the major component of each wash step.

7.5.4 Limitations of Purifying by Bulk Assay

Even though the component of each eluted fraction contained BCAL2646 (Figure 7.13), there was very little mass of protein actually eluted. With 8.5 mg of BPSL1204 bait-protein originally bound, the greatest mass of protein eluted was in the 1st eluate with a total mass of 0.12 mg protein, of which an estimated 40% was BCL2646. Therefore there was an elution ratio of approximately 1:0.0056 mg BPSL1204:BCAL2646 $[(0.12/100 \times 40)/8.5]$. It was concluded that either

Pull-down Experiment	Eluate (post SDS-PAGE 25µl)	µg Total Protein	Approx µg BCAL2646 (40% total)
Log-phase culture	1 st 105 µl 2 M NaCl	120	48
	2 nd 90 µl 2 M NaCl	70	28
	3 rd 85 µl 2 M NaCl	30	12
Overnight culture	1 st 85 µl 2 M NaCl	120	48
	2 nd 60 µl 2 M NaCl	55	22
	3 rd 65 µl 2 M NaCl	20	8
1 st Test Pull-down	1 st 25µl 1.5 M NaCl	25	10
	2 nd 25 µl 2 M NaCl	10	4

Table 7.1: Retrieval of BCAL2646 from bulk pull-downs using AffiGel15:BPSL1204 and *B. cenocepacia* cell-extract.

Combination of all the above elutions gave a final sample of 450 mg in 500 µl (0.9 mg/ml), of which approximately 40% was BCAL2646, therefore a very low yield of ~180 mg target protein from ~20 g of combined starting cell-mass. There was very little difference observed between using log-phase or overnight cultures for BCAL2646 isolation.

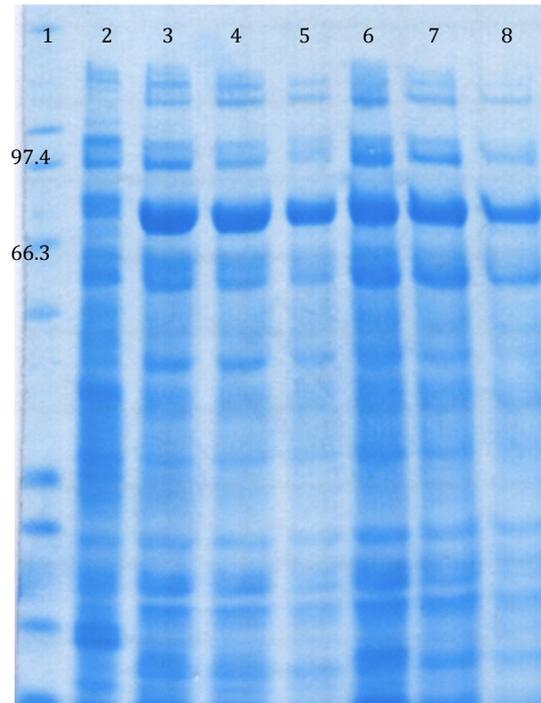


Figure 7.13: SDS-PAGE analysis of bulk pull-down assays.

Lane 1 = Mark12 (kDa)

Lane 2 = Cell-free extract

Lane 3-5 = Eluates 1-3 in 2 M NaCl from log-phase culture

Lane 6-8 = Eluates 1-3 in 2 M NaCl from overnight culture

Each eluate (25 μ l loaded) shows that the 79.1 kDa BCAL2646 is the major component of each elution, but only to a maximum of approximately 40%.

only 0.013% [$100/(360 \text{ mg cfe}/8.5 \text{ BPSL1204 available}) \times 0.0056 \text{ mg BCAL2646 eluted}$] of the soluble protein from *B. cenocepacia* was BCAL2646 in the first place or that the interaction, though always occurring, was very transient.

As *B. cenocepacia* is predicted to contain approximately 5700 proteins (including membrane and secreted proteins, which were very unlikely to be in the cell-free extract), even if only 1/5700 of proteins per-cell were BCAL2646 at any given point in time, there should still have been at least approximately 0.065 mg of BCAL2646 in the 360 mg cell-free extract applied in the assay, therefore only 74% of the theoretically minimum mass of BCAL2646 possible was extracted. This was considered an unlikely situation, with at least hundreds of copies per cell of the methionyl-tRNA synthetase BCAL2646 likely to be present in a population of growing cells, therefore reflecting a very small yield from the bulk assay. It was therefore concluded that although at any point in time some BCAL2646 macromolecules were interacting with BPSL1204, as seen constantly in gels such as Figure 7.13, the vast majority of BCAL2646 remained in an unbound state, either reflecting an above-average occurrence of non-specific, non-biological interactions or a transient and unstable complex.

7.5.5 Heparin Column Purification of BPSL2646

To investigate the potential formation of a BPSL1204:BCAL2646 complex, all the 6 eluates from the bulk pull-down assay (3 x log-phase culture and 3 x overnight culture) and the 2 M NaCl eluates from the original pull-down assays (Figure 7.13 Lane 3-8 and Figure 7.6 Lanes 10 and 12) were pooled together to give approximately 0.45 mg in 500 μl of 0.9 mg/ml protein (Table 7.1) in 2 M NaCl and 50 mM HEPES pH 7.5, which was washed and Viva-spin buffer-transferred into 50 mM Tris pH 7.5 with minimal residual NaCl as in Section 3.6.5. This was then applied to a 1 ml Heparin HiTrap column, available from GE Healthcare, on the prediction that due to the RNA-binding capabilities of a methionyl-tRNA synthetase, such as BCAL2646, the pooled fractions could be purified further using affinity chromatography. A gradient of 0-1.5 M NaCl was used to elute protein from the heparin column in 0.5 ml fractions, with the UV trace set to high

sensitivity (Figure 7.14), before the fractions under the peak were analysed by running 50 µl of each fraction on SDS-PAGE (Figure 7.15). Fractions 14-16 contained the purest BCAL2646, and were combined to give a resulting 1.4 ml of 0.05 mg/ml, or 70 µg of approximately 85% pure BCAL2646.

7.5.6 Analysis of BPSL1204: BCAL2646 Complex Formation

Although 70 µg was a very small yield of BCAL2646, it was considered detectable as seen by the trace in Figure 7.14. The 1.4 ml heparin column fractions were Viva-spin buffer-exchanged and concentrated into 90 µl of 50mM Tris pH 7.5 and 0.15 M NaCl at approximately 6.3 µg/ml protein. This was combined with 10 µl of 1 mg/ml BPSL1204 in the same buffer, to give a final complex ratio of ~1:0.7 BPSL1204:BCAL2646. 90 µl of the complex mix was allowed to properly combine for 15 minutes at room temperature, before being loaded onto a Superdex200 GL 11 gel-filtration column, available from GE Healthcare, pre-equilibrated with the complex buffer. The column was run at 0.7 ml/min and 0.5 ml fractions were collected, with the UV trace (Figure 7.16) set to the highest sensitivity. No high molecular weight peaks were seen, corresponding to no BCAL2646:BPSL1204 complex or monomeric BCAL2646, with a peak witnessed at Fraction 23 that corresponded with the classic monomeric BPSL1204-MA peak seen in purification processes (Section 4.4.4). The peak seen at the buffer front however was an anomaly. Fractions under the peaks and samples of the pre-gel-filtration samples were analysed by SDS-PAGE (Figure 7.17). Assessment of Fractions 31 and 32 revealed very low levels of BCAL2646, but at the running front of the buffer. Therefore it was proposed that BCAL2646 perhaps interacted with the dextran-agarose beads and definitely not BPSL1204.

7.6 Conclusions

The initially apparently reproducible interaction between BPSL0998/BCAL2646 and AffiGel15:BPSL1204 seemed to contradict the structural analysis of BPSL1204 consisting of a BLIP-like fold. However, upon attempts to purify BCAL2646 by affinity to BPSL1204 in bulk pull-down, it was seen that the

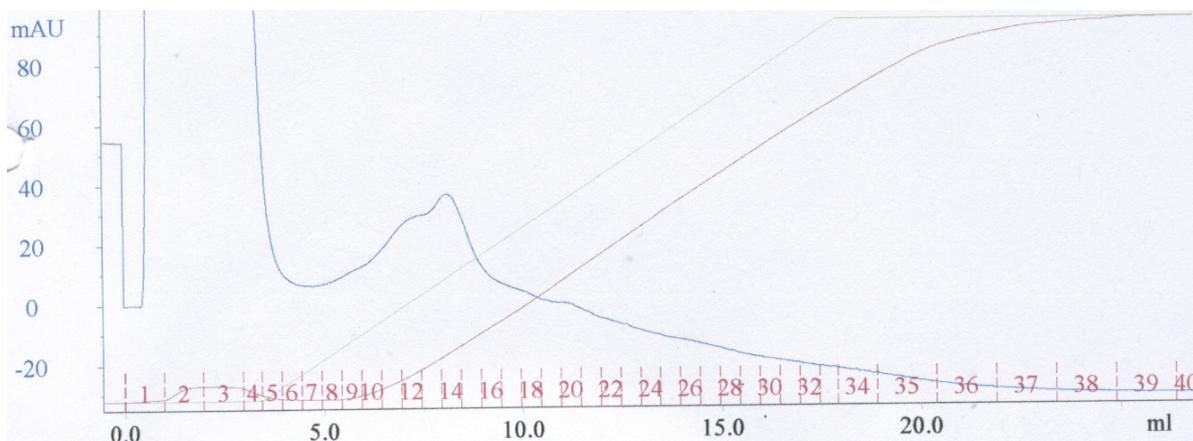


Figure 7.14: Heparin column trace of BCAL2646 purification. The 0-1.5 M NaCl gradient is seen as the red conductivity line. The blue line represents the high-sensitivity UV trace detecting protein presence. A peak is eluted between Fractions 12 and 16.

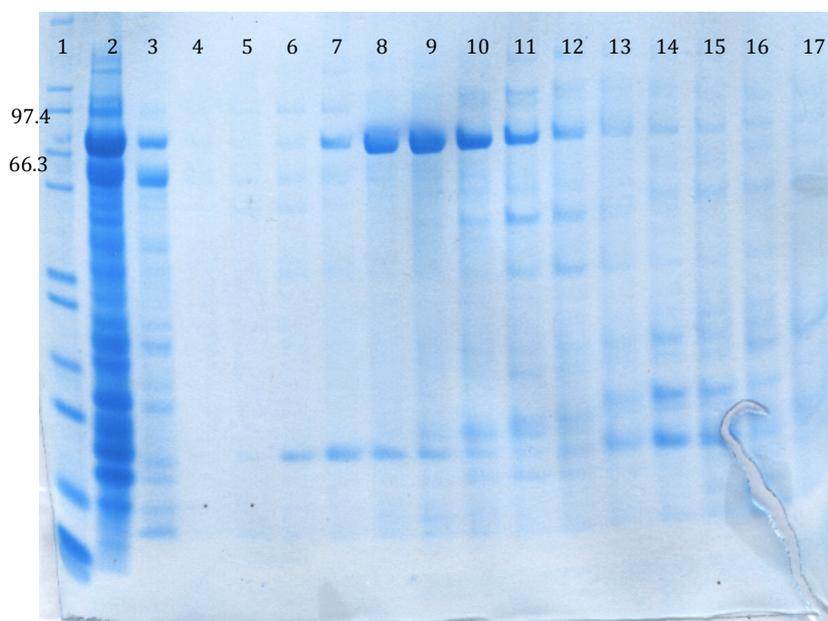


Figure 7.15: SDS-PAGE analysis of BCAL2646 heparin column purification.

Lane 1 = Mark12 (kDa) Lane 3 = Fraction 2 (non-binding wash)
 Lane 2 = Sample loaded Lanes 4-17 = Fractions 10-23

Combining Fractions 14-16 created a BCAL2646 sample of 1.4 ml of 50 $\mu\text{g/ml}$. This was estimated to be 85% pure.

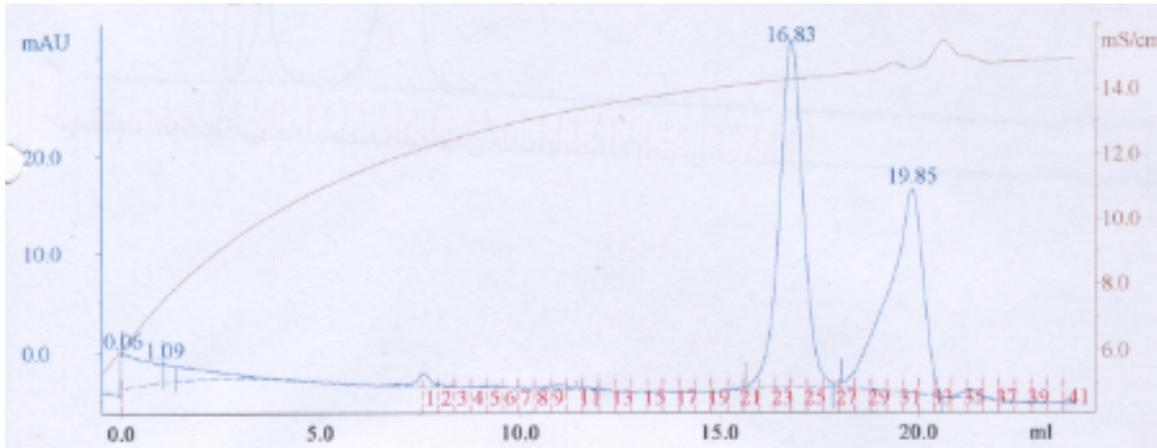


Figure 7.16: Gel filtration trace assessing BPSL1204:BCAL2646 complex formation.

The ‘classic’ low molecular weight BPSL1204 monomeric peak is seen in Fraction 24, with no high molecular weight peak seen for BCAL2646 or BPSL1204:BCAL2646 complex formation (excluded volume or Fraction 1-19 of included volume). An unexpected peak in Fraction 33, at the buffer running front, is seen.

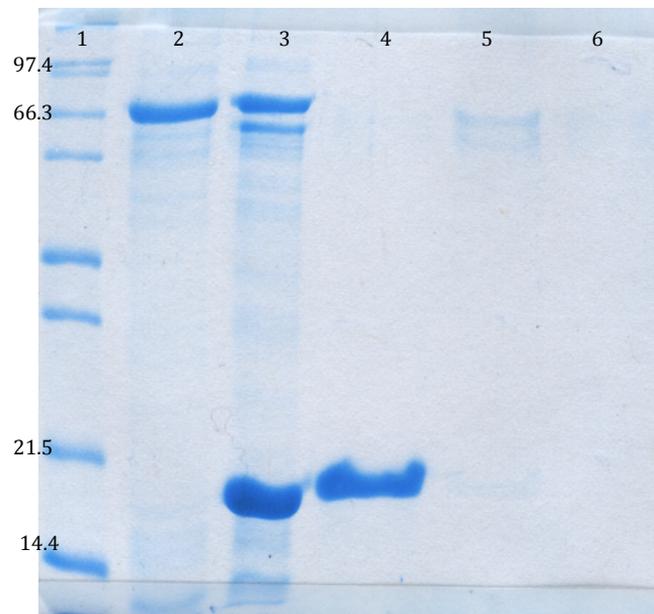


Figure 7.17: SDS-PAGE analysis of gel-filtration peaks from attempts to complex BPSL1204 with BCAL2646.

Lane 1 = Mark12 (kDa)	Lane 4 = Fraction 22
Lane 2 = Heparin-purified BCAL2646	Lane 5 = Fraction 31
Lane 3 = Sample loaded on GF column	Lane 6 = Fraction 32

Fraction 22 contains monomeric BPSL1204. Some BCAL2646 appears to be anomalously present in Fraction 31, at the buffer running front

interaction between BCAL2646 and BPSL1204 was a very weak and transient one and as such was unlikely to be of biological significance. This was supported by the total lack of evidence for complex formation by gel-filtration analysis. This therefore implied that BPSL1204 did not interact with BCAL2646/BPSL0998 or if it did, it required further complex stabilisation by co-factors unknown. Therefore it seemed that the pull-down result was a product of strong non-specific interactions.

Since BPSL1204 carries a conserved glyceryl-cysteine lipid-anchoring motif, it was proposed that BPSL1204 may interact with proteins associated with a cell membrane or may exist as a secreted protein. Therefore further experimentation yet was required to try to ascertain the cellular function and location of BPSL1204.

Chapter 8: Cellular Localisation of BPSL1204 and the Discovery of Potential Immunogenic Targets

8.1 Preparation For Localisation Study

8.1.1 Supposition Behind Study

To avoid safety issues associated with culturing *B. pseudomallei*, *B. cenocepacia* and BCAL2351 were again used as a model for BPSL1204 biochemical investigation. It was proposed that polyclonal antibodies raised in rats against epitopes present on BPSL1204 would elicit a binding event when extracted in sera and used to challenge BCAL2351. This was due to the likelihood that one or more BPSL1204 epitopes against which antibodies were raised by a host would be present on the 92% identical soluble domain of BCAL2351.

Therefore it was hypothesised that upon the fractionation of *B. cenocepacia* into cytoplasmic, periplasmic, secreted, inner membrane and outer membrane fractions, the said cellular compartments could be Western-blotted to nitrocellulose (Burnette, 1981) and subjected to binding by a primary rat IgG-antibody against BPSL1204. Primary antibodies bound to BCAL2351 could then be bound by a secondary anti-rat-IgG antibody conjugated to a horseradish peroxidase (HRP) reporter-enzyme for detection by X-ray film chemiluminescence (Kricka, 1991). Analysis of the developed X-ray film would then allow for the identification of the cellular fraction of the bacterium in which BCAL2351, and therefore by implication BPSL1204, could be found biologically. This might then add extra information to the structural and/or pull-down results.

8.1.2 Generation of Polyclonal Antibodies and Rat Hyper-immune Response to BPSL1204

To facilitate the generation of anti-BCAL2351/BPSL1204 primary antibodies, BPSL1204 was purified as in Section 3.6 into 10 mM Tris pH 8.0 to a concentration of 1 mg/ml. 2 x 1 ml of this fresh, pure BPSL1204 solution was used to inoculate two rats, after the removal of a sample of pre-immune blood.

The Sheffield University-based company BioServUK carried out all antibody production work under a Home Office licence. Both rats were allowed to mount an immune response to the BPSL1204 antigen for 7 weeks, before test sera were removed and subjected to an in-plate Enzyme-Linked Immunosorbant Assay (ELISA) using an anti-rat IgG-Fc HRP-linked antibody as a secondary reporter (King & Kochoumian, 1979) against a plate-immobilised BPSL1204 with anti-BPSL1204-containing sera. Against a pre-immune control, antibody titre (defined as the approximate dilution ratio that gave half of the maximum measurable signal, i.e. $<1.0 \text{ OD}_{450\text{nm}}$) was determined by the fluorescence change generated upon conversion of 3,3',5,5'-Tetramethylbenzidine Enhanced Chemiluminescence (TMB-ECL) reagent by HRP (Frey et al, 2000). Rat 1 gave a test-titre of 1:102400 and Rat 2 gave a test-titre of 1:51200 (Figure 8.1).

The immune response was boosted by the injection of a small volume of 1 mg/ml BPSL1204, and the rats were anaesthetically exsanguinated 1 week later. Final titres of 1:409600 for Rat 1 and 1:102400 for Rat 2 were achieved from ELISA's of the final anti-BPSL1204 sera (Figure 8.2). A titre greater than 1:4000 is considered a strong immune response in the strain of rats used (Dr Andrew Wright, BioServeUK, personal communication), and therefore Rat 1 gave a response 100 times greater than expected, whilst Rat 2 gave a response 26 times greater than expected. It was therefore proposed that the BPSL1204 sequence and structure might present epitopes that elicit a particularly antigenic response in rats, perhaps reflecting a biological role in regards to cellular location and function.

8.1.3 Isolation of *B. cenocepacia* Cellular Fractions

B. cenocepacia was grown overnight and centrifuged into a pellet. A sample of the supernatant media was taken as representative of the secreted fraction. Pelleted cells were then resuspended in 20% sucrose, 30 mM Tris pH 8.0, 1 mM EDTA (STE) buffer for the osmotic permeabilisation of the outer membrane,

	Test-Sera		
	Rat 1	Rat 2	Pre-Immune
1:100	1.137	1.470	0.090
1:200	1.258	1.440	0.062
1:400	1.249	1.467	0.059
1:800	1.214	1.686	0.054
1:1600	1.216	1.717	0.048
1:3200	1.147	1.880	0.052
1:6400	1.196	2.018	0.053
1:12800	1.307	1.847	0.052
1:25600	1.788	1.188	0.047
1:51200	1.506	0.690	0.048
1:102400	0.941	0.378	0.059
1:204800	0.565	0.237	0.054
1:409600	0.319	0.172	0.043
1:819200	0.191	0.128	0.049
1:1638400	0.122	0.092	0.049
1:3276800	0.098	0.079	0.050

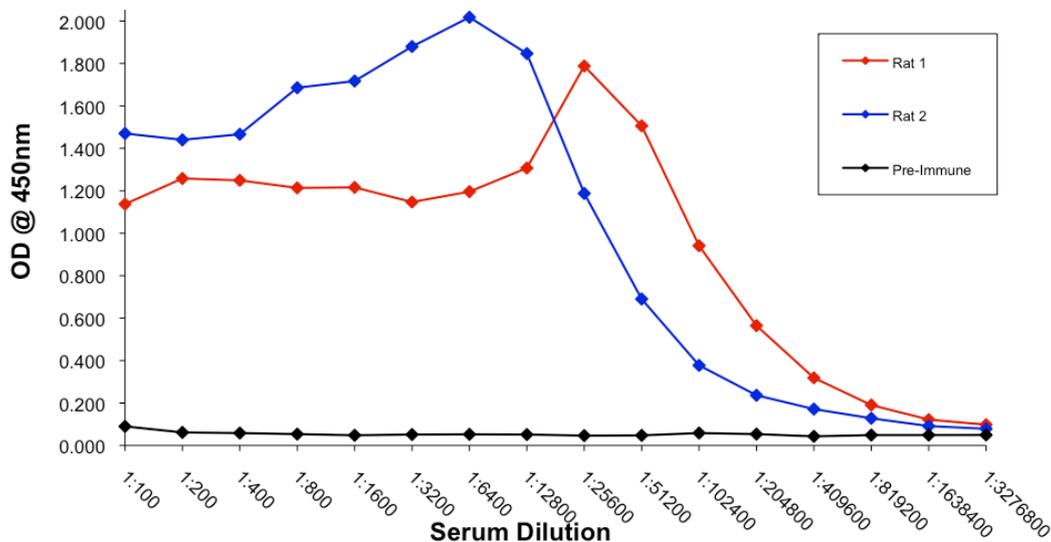


Figure 8.1: Results of test ELISA titre analysis of anti-BPSL1204 immune response in rats. *Top:* Table of ELISA dilution ratios and TMB-cleavage OD at 450nm. *Bottom:* Line graph of OD against serum dilution, plotted in Excel.

In the table, the titre dilution ratios are highlighted in red. A titre of 1:4000 was considered to show a significant immune response. In these test sera of Rat 1 and Rat 2, titres of 1:102400 and 1:51200 respectively were achieved, indicative of a highly immunogenic response to BPSL1204. Both rats were allowed to proceed in generating anti-BPSL1204 polyclonal antibodies. No response was seen in the pre-immune sera, indicating that the titres seen were due to BPSL1204 inoculation.

	Final-Sera		
	Rat 1	Rat 2	Pre-Immune
1:100	9.999	9.999	0.312
1:200	9.999	9.999	0.179
1:400	9.999	9.999	0.128
1:800	9.999	6.317	0.108
1:1600	9.999	9.999	0.081
1:3200	9.999	9.999	0.077
1:6400	9.999	9.999	0.076
1:12800	9.999	2.841	0.086
1:25600	9.999	2.171	0.163
1:51200	2.761	1.418	0.111
1:102400	2.009	0.827	0.095
1:204800	1.281	0.508	0.096
1:409600	0.793	0.289	0.086
1:819200	0.494	0.191	0.085
1:1638400	0.284	0.144	0.078

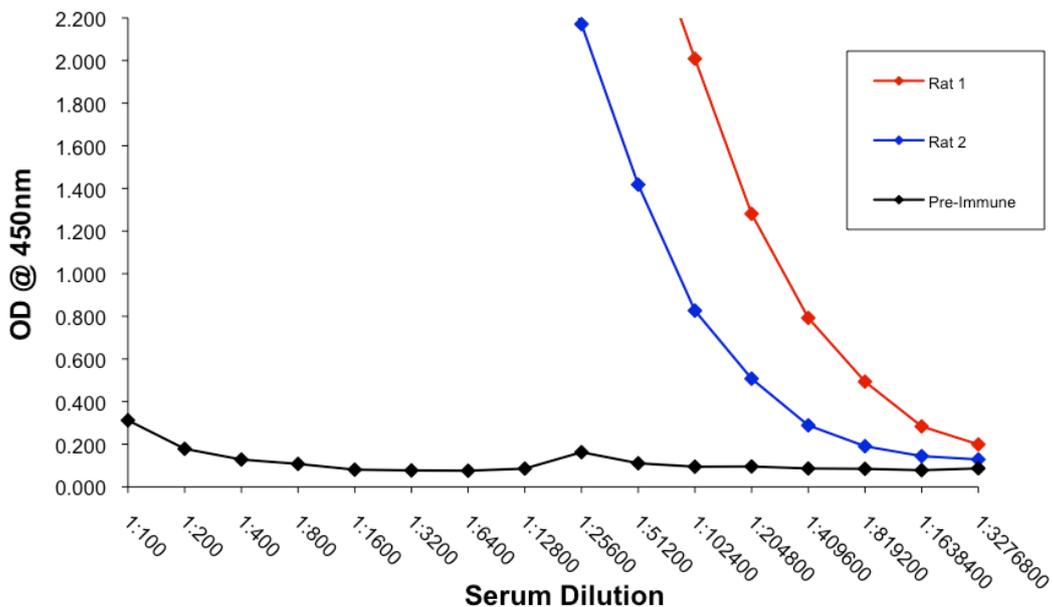


Figure 8.2: Results of final ELISA titre analysis of anti-BPSL1204 immune response in rats. *Top:* Table of ELISA dilution ratios and TMB-cleavage OD at 450nm. *Bottom:* Line graph of OD against serum dilution, plotted in Excel.

As can be seen in the graph, the immune response mounted by both rats is very large, with most serum dilutions off the scale. The final titres measured are highlighted **red**. No response seen in the pre-immune sera suggests that the large immune response is due solely to inoculation with pure BPSL1204. These sera were harvested by anaesthetic exsanguination to be used as the probes in BPSL1204 localisation studies.

from which the supernatant after centrifugation was taken as the soluble periplasmic fraction. A sonication step to shatter membranes was employed, from which an ultracentrifuged supernatant was then taken as the cytoplasmic fraction. The inner membrane was selectively solubilised using the detergent sarkosyl (sodium lauroyl sarcosinate), from which the outer membrane was pelleted via centrifugation. The pellet was resuspended in 10 mM HEPES pH 7.4 to create the outer membrane fraction. A full account of this method is given in Section 3.9.1.

Analysis of fractions was carried out by SDS-PAGE to discover the presence of differing protein profiles to assess success of fractionation (Figure 8.3). It was seen that each fraction did appear to differ in protein profile, except that many cytoplasmic and periplasmic proteins were present in both samples. This was taken as an indication that the fractionation had been reasonably precise in terms of membrane preparation, but that the periplasmic fraction was contaminated with partial cytoplasmic leakage. This was provisionally considered as acceptable, unless a situation in which antibodies bound to non-membrane fractions arose.

8.2 Immunochemiluminescence Assays for Cellular Location of BPSL1204

8.2.1 Experimental Procedure

13 µl of each fraction was run on a 13.5% SDS-PAGE gel alongside a pure 0.5 µg BPSL1204 control and a pre-stained marker lane, which were then Western-blotted onto a nitrocellulose membrane following the protocol described in Section 3.9.3. The post-Western blot SDS-PAGE gel was stained and then destained for comparison to a co-run pre-Western blot gel, to assess transfer efficiency. After blocking of non-sample-bound nitrocellulose regions overnight in 5% milk, 0.1% Tween-20 and phosphate-buffered saline (M-T-PBS) at 4°C, the membrane was incubated for 60 minutes at room temperature with a 1:2500 dilution of primary rat anti-BPSL1204 sera from Rat 2, washed and then incubated for 45 minutes at room temperature with a 1:5000 dilution of

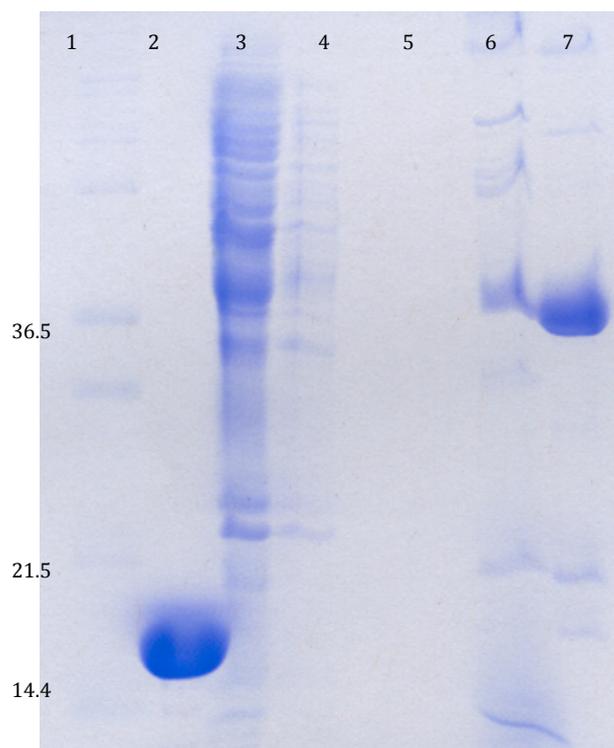


Figure 8.3: SDS-PAGE analysis of *B. cenocepacia* fractionation.

Lane 1 = Mark12 (kDa)

Lane 2 = 15 μ g BPSL1204

Lane 3 = 13 μ l cytoplasm

Lane 4 = 13 μ l periplasm

Lane 5 = 13 μ l secreted

Lane 6 = 13 μ l inner membrane

Lane 7 = 13 μ l outer membrane

15 μ g of pure BPSL1204 was run against the cytoplasmic fractions to allow for a rough estimation of concentration of fraction contents. As can be seen, the cytoplasmic fraction contains the most proteins, with some very faint bands seen in the periplasmic and secreted fractions. However cross contamination between the cytoplasm and periplasm is likely.

The periplasmic and secreted proteins were at very low concentrations, but due to the extremely high titre of anti-BPSL1204 antibodies (Section 8.1.2) it was hoped that the high sensitivity of the probe used would compensate for this.

The inner membrane and outer membrane fractions (Lane 6 and 7) exhibited significantly different protein profiles, suggestive of efficient separation.

secondary anti-rat rabbit IgG HRP-conjugate. Full details of this protocol can be found in Section 3.9.4. After being washed to remove all secondary reporter antibodies, the side of the membrane onto which proteins were blotted was then coated in a solution of HRP-substrate ECL reagent and the membrane was sealed in clean air-bubble-free plastic wrap. In a dark room, the fluorescent signal generated by the catalysis of ECL reagent by BPSL1204/BCAL2351-bound HRP was visualised by exposure to X-ray film. Exposure time was varied to achieve the best signal:noise image, and then the film was developed. This experimental procedure was carried out in triplicate, alongside controls of secondary antibody alone or use of rat pre-immune sera as the primary antibody source. For a detailed description of this protocol, see Section 3.9.5.

8.2.2 Results of the Experiment

Confirmation of successful Western blotting of cellular fractions was achieved by resolution of pre- and post-Western SDS-PAGEs (Figure 8.4). It was seen that nearly complete transfer of low molecular weight proteins was achieved (such as BPSL1204 in Lane 2), but in general all lanes post-transfer to nitrocellulose appeared much weaker. It had been well documented that the rate of lower weight protein transfer was far more efficient than that of higher weights. In addition to this, both low and high molecular weight pre-stained proteins in the marker lane were seen to have transferred to the nitrocellulose membrane (Figure 8.5, Lane 1).

As seen in Figure 8.5, the anti-BPSL1204 antibodies bound BPSL1204 with high affinity, with no signal generated from a control using secondary anti-rat HRP-conjugate alone (Figure 8.6). The lack of signal from replacing post-immune anti-BPSL1204 sera with pre-immune sera, seen in Figure 8.7, showed that binding events were solely a result of an anti-BPSL1204:BPSL1204 complex generated as a result of inducing a rat immune response. Analysis of Figure 8.5 showed that rat anti-BPSL1204 bound BCAL2351 exclusively in the outer membrane fraction of *B. cenocepacia*.

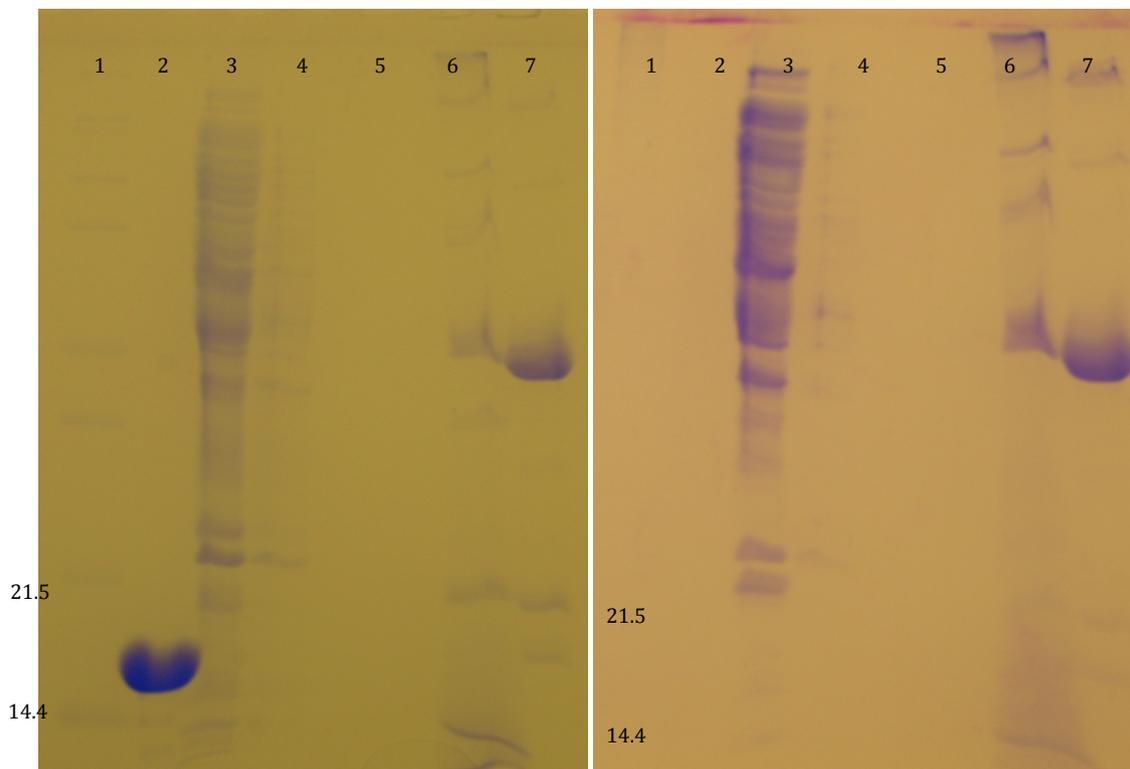


Figure 8.4: SDS-PAGE analysis of Western blot transfer of BPSL1204 and *B. cenocepacia* fractions. *Left:* Pre-Western. *Right:* Post-Western.

Lane 1 = Mark12 (kDa)	Lane 5 = 13 μ l secreted
Lane 2 = 15 μ g BPSL1204	Lane 6 = 13 μ l inner membrane
Lane 3 = 13 μ l cytoplasm	Lane 7 = 13 μ l outer membrane
Lane 4 = 13 μ l periplasm	

As can be seen, there was total transfer (as judged by eye) of BPSL1204 and most proteins < 20 kDa in all fractions. However, proteins with higher molecular weights seem less efficient at transferring during Western blotting.

Nevertheless, as the aim of the experiment was to detect the BCAL2351 homologue of BPSL1204, then as long as proteins of similar mass to BPSL1204 transferred then BCAL2351 localisation was still possible.

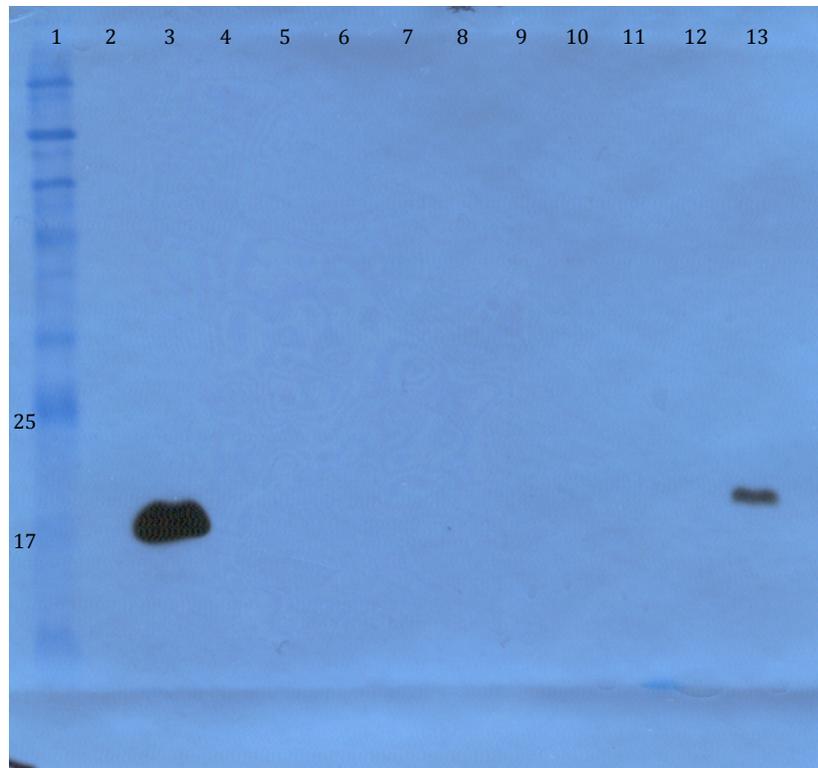


Figure 8.5: BPSL1204 and *B. cenocepacia* fractions immuno-blotted with anti-BPSL1204. X-ray film was exposed for 30 seconds.

Lane 1 = BroadRange pre-stained marker
 Lane 3 = Pure BPSL1204
 Lane 5 = Cytoplasm
 Lane 7 = Periplasm

Lane 9 = Secreted
 Lane 11 = Inner membrane
 Lane 13 = Outer membrane
 Even lanes = Empty

Even against a small amount of protein (0.5 μ g) a considerable binding event was seen against pure BPSL1204 (Lane 3). The only other binding seen was in the outer membrane fraction. This band (Lane 13) ran slightly slower than pure BPSL1204, indicative of the detection of a native BCAL2351 processed-lipoprotein covalently associated to the additional mass of a lipid-anchor. BCAL2351 was not detected in any other fraction



Figure 8.6: Secondary antibody control of BPSL1204 localisation immunoblots. Lane 1 = BroadRange pre-stained marker. Lane 2 = BPSL1204. Lane 3 = Outer membrane. X-ray film was exposed for 5 minutes.

This immunoblot was probed with in the anti-rat HRP-conjugated secondary antibody. No binding is witnessed, indicating that the bands seen in Figure 8.5 are due to the primary binding of an antibody in the rat anti-BPSL1204 sera

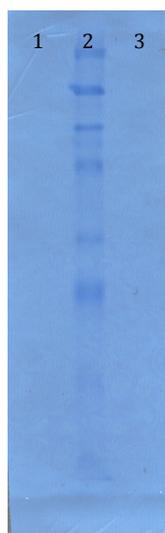


Figure 8.7: Pre-immune sera control of BPSL1204 localisation immunoblots. Lane 1 = BPSL1204. Lane 2 = BroadRange pre-stained marker. Lane 3 = Outer membrane. X-ray film was exposed for 5 minutes.

When probed with rat pre-immune sera as a source of primary antibody, no secondary antibody anti-rat HRP-conjugate binding was seen. This indicated that the primary binding seen in Figure 8.5 was due to antibodies raised against BPSL1204 post-inoculation, and not elements within native rat sera.

8.2.3 Conclusions Drawn From the Experiment

From the single band discovered in Figure 8.5 it was concluded that BPSL1204/BCAL2351 is an outer-membrane protein, anchored to the membrane via a probable glycerylcystiene moiety. Although the band observed appeared to be of slightly larger MW than that of pure BPSL1204-MA, this was probably because native BCAL2351 identified in the outer membrane of *B. cenocepacia* had the addition of covalently associated lipids. The outer membrane fluorescent signal was not seen in the control using only secondary HRP reporter antibody (Figure 8.6), therefore indicative of true localisation of BCAL2351/BPSL1204 and not an interaction between a non-specific protein and the secondary reporter. However, no information could be gleaned from the experiment in terms of orientation of BPSL1204 on the outer membrane, and therefore further investigation was required to determine if BPSL1204/BCAL2351 was located to the extracellular or periplasmic face of the outer membrane.

8.3 Immunofluorescence Microscopy of the *B. cenocepacia* Outer Membrane Protein BCAL2351 (BPSL1204)

8.3.1 Reasoning for the Fluorescence Microscopy Experiment

Although the cellular localisation of BCAL2351 to the outer membrane had been confirmed by immunoblot analysis of cellular fractions (Section 8.2.2), it was still unclear as to the orientation of the protein in the outer membrane. It was hypothesised that due to the lack of disulphide-bridge potential in the primary amino acid sequence that BPSL1204/BCAL2351 was unlikely to exist as an extracellular protein without the structural strength endowed by disulphide covalent bonds (Raina & Missiakas, 1997). Therefore it was proposed that immunofluorescence confocal microscopy of whole, fixed *B. cenocepacia* cells challenged with a rat anti-BPSL1204 primary antibody and a secondary anti-rat FITC fluorophore conjugate (Riggs et al, 1958) would allow the discovery of the orientation of BCAL2351 on the outer membrane of *B. cenocepacia* and therefore of BPSL1204 in *B. pseudomallei*. If BPSL1204 were an internal lipoprotein, then there would be no green fluorescence from whole cells, with a green fluorescent

signal only recorded if BPSL1204 was externally anchored to the outer membrane and thus exposed to the extracellular, antibody-containing environment.

8.3.2 Experimental Procedure

Following the protocol described in greater detail in Section 3.9.6, whole cells from a log-phase *B. cenocepacia* culture were fixed with a formaldehyde and glutaraldehyde solution and then dried to a poly-L-lysine-coated glass microscope slide and bound with rat anti-BPSL1204 primary and goat anti-rat IgG-FITC antibodies. Slides were then viewed and focussed for confocal Z-stack image collection using a DeltaVision deconvolution microscope and SoftWoRx 3.5.0 software. Fluorescence of FITC was achieved by setting a laser excitation wavelength of 490nm. Slide-preparation and image-collection was carried out with the aid and supervision of Victoria Kent from the lab of Professor Simon Foster of the University of Sheffield Department of MBB.

Several control slides were also prepared, the first of which was a secondary antibody only slide to account for background fluorescence generated by non-specific binding of the secondary FITC reporter. In addition, a slide was prepared using a sample of the pure, shattered outer membrane sample used in Section 8.2.1. This slide, containing outer membrane with no regulated orientation, was designed to play a dual purpose. Firstly it was used as a control to show whether primary antibody from the sera would still bind BCAL2351, as seen in Section 8.2.2, under microscopy conditions. Upon confirmation that anti-BPSL1204 bound to allow secondary FITC fluorophore signalling, this could be used to infer the results of the whole-cell slide. If whole cells did not fluoresce but the shattered outer membrane sample did, then it may have indicated that BPSL1204 was internalised, and not available for external binding in whole cells. This slide also had a counterpart secondary FITC antibody-only control slide prepared, to account for any non-specific signal generated by exposure of both sides of the outer membrane.

8.3.3 Preliminary Results of Fluorescence Microscopy

Due to the low level of detail required from the microscopy experiments, with the external fluorescence under investigation either present or absent, Z-stack images were processed using a Quick Projection function in which all images were layered on top of one another to generate one image. Investigation of the control slides treated with secondary-reporter-antibody only for both the whole cells and shattered OM-fraction showed no generation of fluorescent signal. Images are not shown, as they were completely blank. This indicated that any green fluorescence seen was due to specific binding of the reporter to primary antibodies associated with BCAL2351/BPSL1204, and not random association of the reporter with incompletely blocked slides.

The slides presenting random orientations of shattered outer membrane exhibited speckles of green (Figure 8.8), witnessed in several repeats of the slides. Due to the fragmented nature of the outer membrane sample, this was the result predicted if BPSL1204 and BCAL2351 were presented on one of the two sides of the outer membrane, with both sides exposed to binding by the primary anti-BPSL1204 antibody.

However, analysis of fixed, whole cells showed mixed results, with localised regions of green-fluorescence on the surface of bacterial cells in some samples (5 out of 12 slides) and no fluorescence in other slides of whole cells (7 out of 12 slides). Examples of these results are shown in Figure 8.9. This implies a lack of binding of the primary antibody to BCAL2351 in the slides with no fluorescence, but potentially external binding to BCAL2351 in the other slides, with some indication of specific localisation and clustering of the protein where fluorescence was seen. No fluorescence is seen in the areas of these slides where there are no cells. The normal, light microscope slides in Figure 8.9 show that whole cells were indeed fixed and present on the slides.

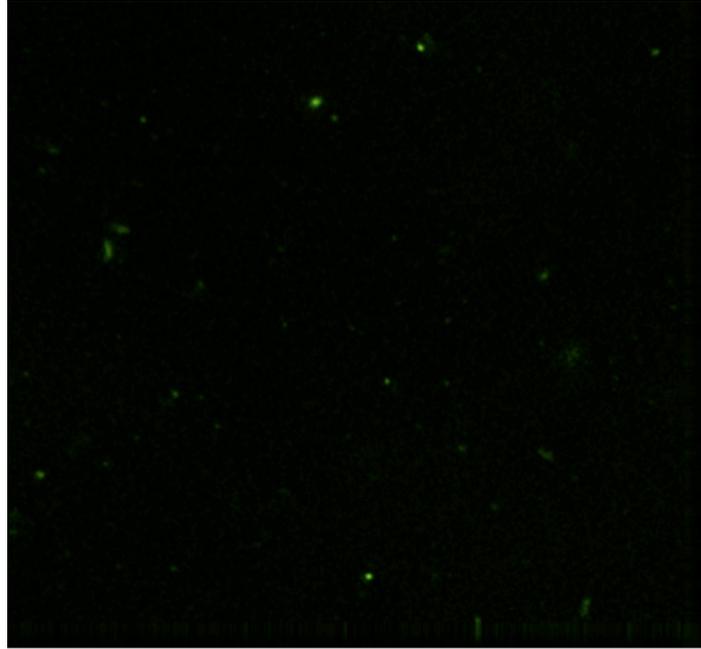


Figure 8.8: DeltaVision fluorescence microscopy of FITC-tagged BCAL2351 in the shattered outer membrane.

The slide above is an example of multiple slides coated in sonicated and purified *B. cenocepacia* outer membrane. The membrane sample applied was assumed to be immobilised on the slide in small fragments of random orientation. The slide was bound first by rat anti-BPSL1204 and then goat anti-rat FITC conjugate.

When excited by a laser, the secondary FITC-antibody fluoresced when bound to BCAL2351 via the primary binding. BCAL2351 is present, represented by the green 'speckling', reaffirming that BCAL2351 and BPSL1204 are outer-membrane-associated. This does not allow the conclusion of the orientation of BCAL2351/BPSL1204 in the outer membrane.

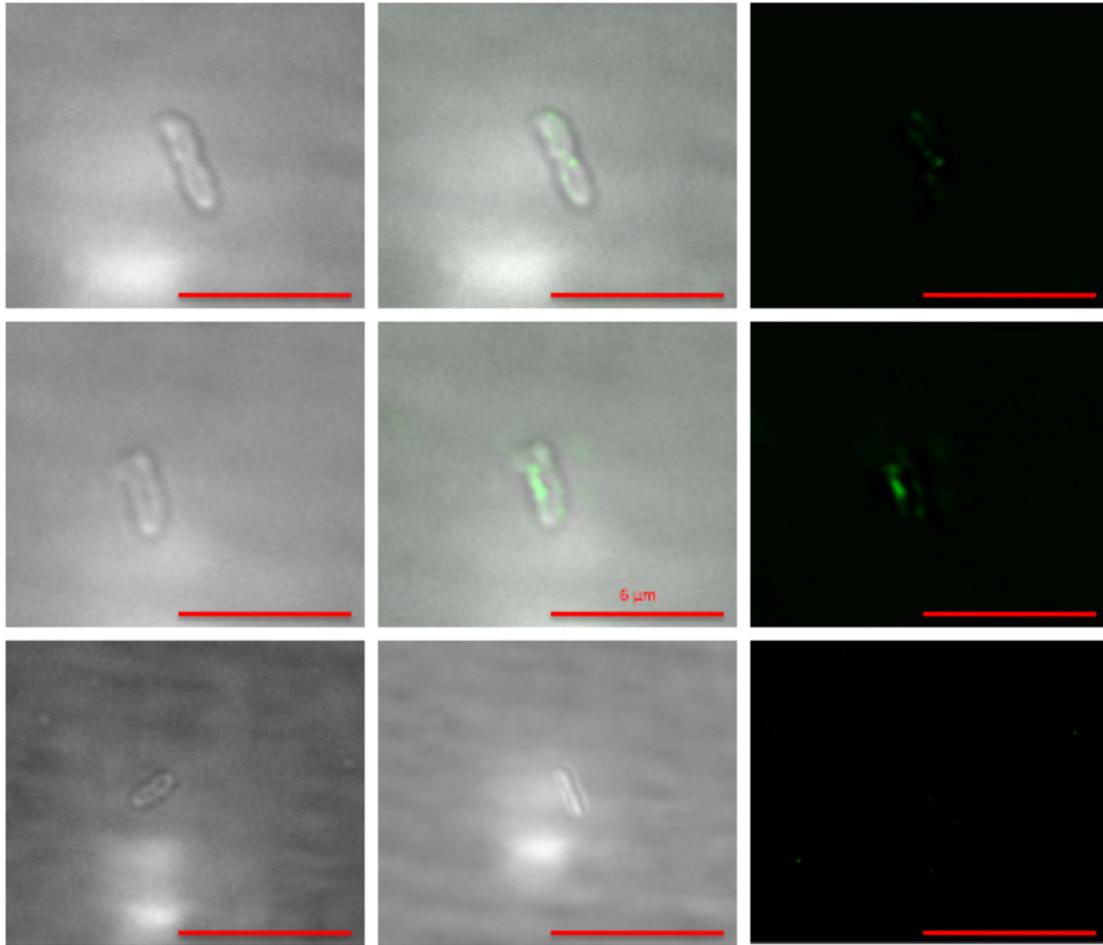


Figure 8.9: Results of whole-cell DeltaVision fluorescence microscopy. *B. cenocepacia* cells were challenged with anti-BPSL1204 and a secondary FITC-conjugate. *Left, center and right columns:* Bright-field, overlay (of left and right) and fluorescence images respectively. *Top, middle, bottom rows:* Positive, positive and negative examples for external binding respectively.

The images in the *Top* and *Middle* row show (in the centre column) that a binding event occurs only within the boundaries of the bacterium, with no fluorescent noise seen from the surrounding slide. *Top* shows a cell potentially mid-division, with *Middle* a ‘normal’ cell. Both exhibit patches of green-FITC binding, suggestive of specifically localised binding of anti-BPSL1204 to BCAL2351, however no pattern was seen. These images represent 42% of the samples investigated.

The *Bottom* row images show whole-cell slides that do not show fluorescence within the boundaries of the bacterium or in the surrounding slide. The overlay and fluorescence images are independent from the bright-field image, due to availability of images, however all images in the *Bottom* row are representative of 58% of the samples investigated. This suggests that BCAL2351 is not accessible to be bound by anti-BPSL1204, in contradiction with the *Top* and *Middle* images.

8.3.4 Conclusions and Limitations of Microscopy Results

When compared to the total lack of signal in the secondary-only control-slides, the patches of green speckles in the fragmented outer membrane slides confirm that BPSL1204 and BCAL2351 are outer membrane proteins, and that under these experimental conditions the rat anti-BPSL1204 primary antibody can bind to BCAL2351 when the protein is exposed and accessible on the outer membrane. However, the lack of any significant fluorescent signal in 7/12 slides when whole cells were treated with antibodies suggests that BCAL2351 is not usually accessible to the primary antibody, and thus no secondary binding of the FITC-conjugated goat anti-rat antibody occurs. In contrast however, 5/12 whole-cell slides suggest that BCAL2351 is present on the outside of some of the cells where it can be bound by anti-BPSL1204, and further suggests specifically localised binding on the bacterial cell surface.

The cells used in the preparation of slides came from the same formaldehyde/glutaraldehyde-fixed sample. The sample was taken from an overnight culture of *B. cenocepacia* and therefore had the potential to contain cells in various stages of the cell cycle. This led to the conclusion that from the preliminary experimental results in Section 8.3.3 BPSL1204/BCAL2351 are external outer-membrane proteins only expressed at certain stages of the *B. pseudomallei*/*B. cenocepacia* cell-cycle. However, an alternative explanation would be that the 42% green signal witnessed in whole cells could be due to damage to the outer membrane in the fixing process, thus exposing the periplasmic face of the outer membrane to antibody-binding, and that ordinarily BCAL2351/BPSL1204 are internalised to the periplasmic face of the outer membrane and therefore give no green fluorescent signal, as seen 58% of the time in these experiments.

These experiments were carried out towards the end of the funding/writing-up period for this project, and due to time constraints concerning the preparation of fresh whole cells and outer membrane fractions of *B. cenocepacia*, further repeats of this experiment were not possible. Although whole cells were

immobilised on slides (Figure 8.9) they were low in cell count per slide and did not appear to fully reflect various stages of the cell-cycle as aimed for - for example only a few cells were seen in mid-division. This did leave the potential for BCAL2351 to be expressed on the outside of the cell at only specific time-points in the *B. cenocepacia* cell cycle not fully represented in the sample used.

In addition to this, the conclusion that absence of exterior binding in the majority of whole cells reflects an internalised outer membrane protein was made on the visualisation of the outer membrane as a simple barrier. It may be more realistic to treat the external environment of the outer membrane as a dynamic system (Rajendran & Simons, 2005), with constant processes such as protein secretion and exopolysaccharide production underway (Osborn & Wu, 1980) which may block external antibody binding in whole cells at certain stages of the cell-cycle, as illustrated in Figure 8.10. Although treated with formaldehyde and glutaraldehyde to 'kill-and-freeze' life-processes, this may have also shielded external BPSL1204/BCAL2351 from antibody binding behind other outer membrane processes immobilised by fixing. The shattered outer-membrane fraction, purified away from an *in vivo* state, may have always allowed external antibody binding to the slides containing only outer membrane or in the immunoblots of Section 8.2.2 due to the stripping-away of other biological factors during the membrane purification process.

Therefore the conclusions made are that BPSL1204 is clearly an outer-membrane protein. At this point it is unclear whether it is located on the external face of the outer membrane at specific stages of the cell-cycle or is located on the inner face of the outer membrane, with any antibody binding seen as an experimental by-product due to ruptured outer membranes. Further experimentation is required to resolve these two possibilities. Optimisation of fluorescence microscopy slides to incorporate a greater cell-count field per slide would allow for analysis of many cells across the exact same conditions, reducing variation due to disparity in slide-preparation. Microbiological growth techniques could be improved so that cells could also be viewed at defined time-

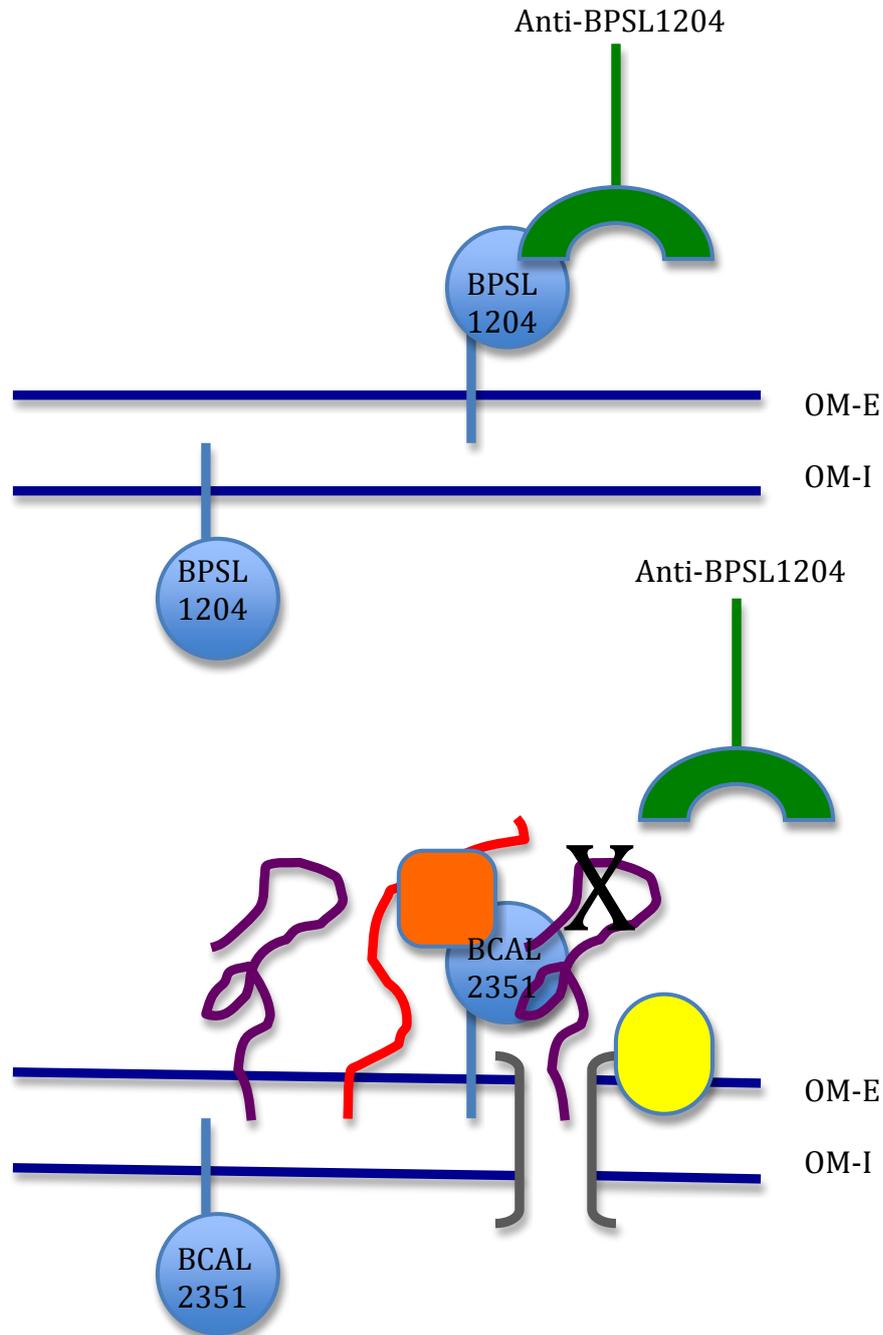


Figure 8.10: Illustration of the difference between pure and *in vivo* outer membrane. OM-E = Outer Membrane Exterior. OM-I = Outer Membrane Internal.

The top image represents the scenario in which the outer membrane is considered as a simple barrier between the periplasm and the environment. The bottom image represents a more likely scenario where the outer membrane is a dynamic system within which BCAL2351 may function, and due to which *in vivo* binding of BCAL2351 by anti-BPSL1204 primary antibody may be prevented.

The purple shapes denote other membrane-anchored proteins in the outer membrane, whilst the red shape is representative of exopolysaccharides secreted from and associated with membranes. The yellow and grey shapes represent peripheral and integral membrane proteins that can be found associated with membranes and the orange shape represents secreted proteins that can also be found in an external membrane environment. All these molecules form a biologically functional environment that may prevent antibodies from binding a target.

points of the *B. cenocepacia* cell cycle, which may present different external exposure or expression of BCAL2351/BPSL1204.

In addition, the bioengineered conjugation of a fluorescent moiety such as green-fluorescent protein (GFP) to BCAL2351 may allow for direct visualisation via fixed and live-cell microscopy of the orientation of BCAL2351 and BPSL1204 when associated with the outer membrane of *B. cenocepacia*. Application of a secondary fluorophore-conjugate known to be periplasm-specific may allow for the analysis of cell rupturing as the cause for variable signal in whole cells.

8.4 Investigation into Immunogenic Targets

8.4.1 Experimental Theory

With the discovery of BPSL1204 and BCAL2351 as outer-membrane proteins, and the large immune response observed against BPSL1204 in rats (Section 8.1.2), it was decided that an investigation into the potential immunogenic status of BPSL1204 in melioidosis was needed. It was therefore proposed to employ a similar methodology to that used in Section 8.2.1. Fractionated *B. cenocepacia* cells and BPSL1204 would be run on SDS-PAGE, Western-blotted to nitrocellulose and challenged with *Burkholderia*-infected human sera as a source of primary antibodies. Proteins present that elicited an immune-response during infection would be bound by the infected sera, to which an anti-human HRP-conjugate secondary reporter antibody would bind for detection by X-ray film chemiluminescence. This would result in the identification of any major antigens shared by *B. cenocepacia* and the *Burkholderia* species against which antibodies were raised, and perhaps also would reveal BPSL1204 as an immunogenic outer membrane protein.

8.4.2 Components of the Assay

Access to *B. pseudomallei*-infected human sera was impossible, due to low survival rates of patients and the dangers of handling blood contaminated with the bacterium. However, following the hypothesis of the original localisation

immunoblot assays, the antigenic effect of *B. pseudomallei* infection was modelled in the infection of a human by *B. mallei*. Sera recovered from a human Glanders patient (sera of Dr David DeShazer, U.S. Army Medical Research Institute of Infectious Diseases, provided by Professor David Rice, University of Sheffield) was used as a source of anti-*Burkholderia mallei* primary antibodies. However, growth of *B. mallei* is also prohibited in Sheffield, and therefore *B. pseudomallei* infection was once again modelled in *B. cenocepacia* cell fractions (Section 8.1.3) using a response against *B. mallei*. As very high sequence identity homologs of BPSL1204 exist in all species used in the model, with the *B. mallei* homolog 100% identical to BPSL1204, this was considered a viable test for BPSL1204 immunogenicity. However any other proteins detected would require further bioinformatical and experimental investigation before being conclusively identified as true *B. pseudomallei* immunogenic proteins.

8.4.3 Experimental Procedure

As with the localisation studies (Section 8.1), 13 µl of each *B. cenocepacia* fraction was run on a 13.5% SDS-PAGE gel alongside 10 µg of pure BPSL1204. BPSL1204 was relatively overloaded to supply plenty of antigen should binding have proven present but weak. Western blots were carried out as in Section 3.9.3 and the resulting nitrocellulose membrane was subjected to overnight blocking by M-T-PBS buffer, before being challenged with a 1:5000 dilution of primary antibody-supplying human *B. mallei*-infected sera in M-T-PBS. After washing, a 1:2500 dilution of goat anti-human HRP-conjugate secondary reporter antibody in M-T-PBS was incubated with the membrane. Alongside this experiment, a control that challenged the *B. cenocepacia* fractions and BPSL1204 with secondary goat anti-human HRP alone was carried out to ensure that any bands seen were not a product of non-specific binding by the reporter. The membrane was then soaked in ECL reagent and fluorescence of bands from bound antibodies was recorded by the development of X-ray film exposed to the resulting chemiluminescent reaction. Greater details of this process can be found in Section 3.9.5.

8.4.4 Results of the Experiment

Comparison of pre- and post-Western blot SDS-PAGEs gels (Figure 8.11) suggested that lower MW proteins, including BPSL1204, had indeed transferred efficiently with their absences being obvious on the post-Western gel. Higher MW bands, though still present on the post-Western gel appeared weaker, suggesting at least partial transfer of every protein loaded onto the gel, supported by the total transfer of pre-stained protein MW markers, seen as blue bands on the nitrocellulose post-Western blot (Figure 8.12).

It was also noticed that in the outer membrane fraction of the pre-Western SDS-PAGE (Figure 8.11) that a single band at around 40 kDa contributed to the majority of the protein present in the fraction. Consequently it was proposed that this protein might act as an antigen during infection or might represent a potential diagnostic marker, so therefore was identified by mass spectrometry (Section 8.4.5).

Analysis of the immunoblots generated revealed that any bands seen in the main assay (Figure 8.12a) were a product of primary sera binding, as shown by a blank film in the secondary-antibody only control (Figure 8.12b). The anti-*B. mallei* human sera did not contain antibodies specific to BPSL1204, as seen by the lack of bands at ~17.2-20 kDa (variation due to potentially covalently associated lipids). However, many other bands were detected from several of the *B. cenocepacia* fractions, amongst which a major band was seen in the secreted fraction (~60kDa) and three from the cytoplasmic fraction (~58, ~40 and ~17 kDa). Bands were seen at similar weights in multiple lanes, but this was considered a product of lane cross-contamination. Regrettably assays could not be repeated in efforts to 'clean-up' due to the limited anti-*B. mallei* sera available. The large ~40 kDa protein, subjected to mass spectrometry (see Section 8.4.5), seen in the SDS-PAGE outer membrane fraction did not appear to elicit an immune response in this model.

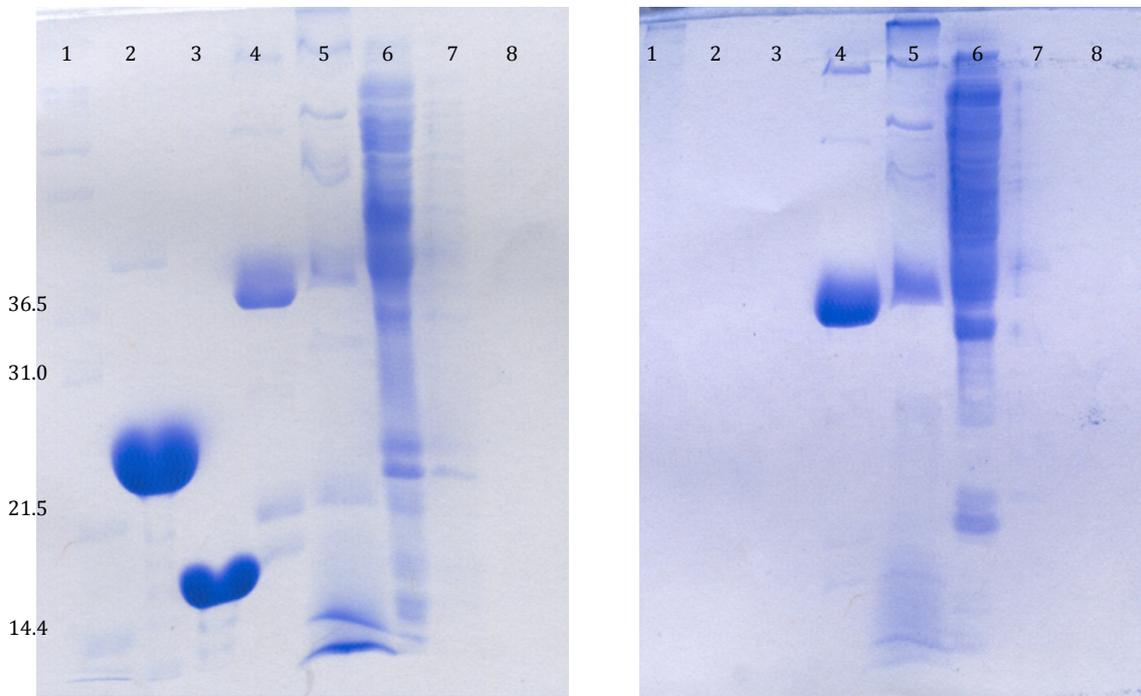


Figure 8.11: SDS-PAGE analysis of Western blot transfer of BPSL1204 and *B. cenocepacia* fractions for immunogenicity studies.

Lane 1 = Mark12 (kDa)	Lane 5 = Inner membrane
Lane 2 = Pure BPSL1549	Lane 6 = Cytoplasm
Lane 3 = Pure BPSL1204	Lane 7 = Periplasm
Lane 4 = Outer membrane	Lane 8 = Secreted

The 23.3 kDa BPSL1549 sample used in Lane 2 was used to assess the limit between transfer efficiency and molecular weight. It was seen that BPSL1549 was totally electroblotted from the SDS-PAGE gel, though some proteins of similar mass in the cytoplasmic fraction (Lane 6) seemed to only partially transfer. The 17.2 BPSL1204 appeared to completely transfer.

A large proportion of the outer membrane fraction (Lane 4) appeared to be a single protein at ~40 kDa.

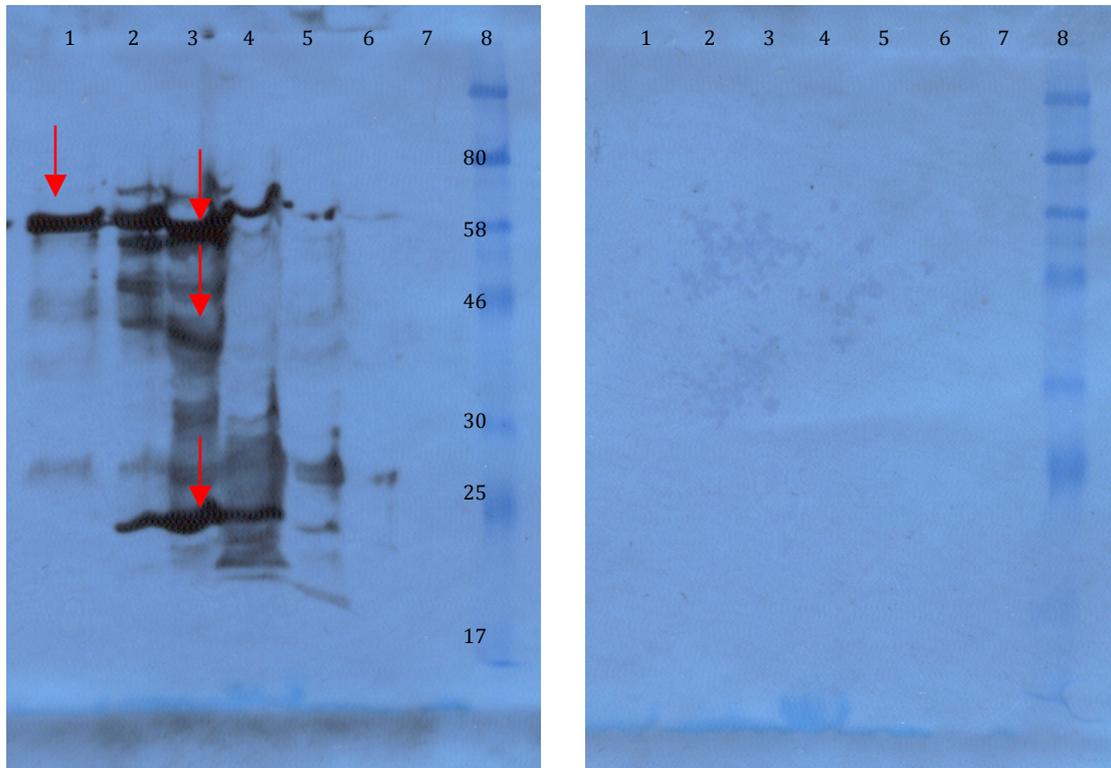


Figure 8.12: Immunoblot analysis of BPSL1204 and *B. cenocepacia* fraction immunogenicity. *Left:* a) immuno-blotted with primary and secondary-reporter antibodies, X-ray film exposed for 30 seconds. *Right:* b) immuno-blotted with secondary-reporter antibody only, X-ray film exposed for 5 minutes.

Lane 1 = Secreted
 Lane 2 = Periplasm
 Lane 3 = Cytoplasm
 Lane 4 = Inner Membrane

Lane 5 = Outer membrane
 Lane 6 = Pure BPSL1204
 Lane 7 = Empty
 Lane 8 = BroadRange prestained marker

The secondary-only control was exposed to X-ray film for 10 times longer than the fully blotted nitrocellulose membrane, and detected no bands. Therefore any bands seen in the left membrane represent a binding event of *B. mallei*-infected sera against *B. cenocepacia*, and not non-specific secondary binding.

BPSL1204 (Lane 5 and 6), and by inference its homologues, did not elicit a significant immune response in *B. mallei* infection of a human. Several *B. cenocepacia* bands did give strong immune responses in *B. mallei* infection, indicated by arrows. These represent potential future targets for the investigation of immunogenic targets in *Burkholderia*.

8.4.5 Mass Spectrometry of a Potential Antigen or Diagnostic Marker

A 13.5% SDS-PAGE gel of the outer membrane fraction of *B. cenocepacia* was run, and stained with InstantBlue colloidal Coomassie G-250 to be compatible with identification for gel-excision, tryptic digestion and gel-extraction for identification by trypsin-fragment OrbiTrap mass spectrometry (Hu et al, 2005). After the staining of the SDS-PAGE, all procedures were carried out by Dr Richard Beniston of the University of Sheffield Department of Biomedical Science. Analysis of peptide fragments by the MASCOT server identified several potential proteins within the band, summarised as an Excel file in Figure 8.13. Based upon the returned results, such as the high score, high % coverage and greatest number of peptides identified it was decided that the main constituent of the band was the putative porin with accession:206563776 (BCAM1931 in *B. cenocepacia* and similar to BPSS0879 in *B. pseudomallei*).

8.4.6 Conclusions of Immunogenic Studies

Though modelled in a *B. mallei* infection, the 100% identity shared with the *B. mallei* protein and BPSL1204 allowed for the inference that *in vivo* BPSL1204 does not elicit an immune response during *B. pseudomallei* infection, regardless of the hyperimmune response witnessed in rats on exposure to pure protein (Section 8.1.2). This supports the preliminary conclusions of Section 8.2.3 and supports one of the conclusions in 8.3.4, that BPSL1204 is an inward-facing outer-membrane protein that would not be presented to a host's immune system during infection. However, the study did reveal several *B. cenocepacia* proteins that did elicit an immune response during a *B. mallei* infection that may represent potential targets for the identification of *B. cenocepacia* infection in CF-patients, or Glanders in both humans and cattle.

Although the putative porin protein discovered in *B. cenocepacia* appeared to be a major component of the outer membrane, it too did not elicit an immune response in *B. mallei*-infection, though a 78% homologue of the porin was discovered to exist in *B. mallei*. As porins are transmembrane proteins, and

Accession	Description	Score	Coverage	# Proteins
325521856	extracellular solute-binding protein [Burkholderia sp. TJI49]	138.90	15.98	15
358076252	outer membrane protein (porin) [Burkholderia cenocepacia H111]	567.25	39.43	1
206562870	putative porin [Burkholderia cenocepacia J2315]	712.90	40.41	1
206561353	putative outer membrane porin protein [Burkholderia cenocepacia J2315]	765.81	39.95	1
206563776	putative porin [Burkholderia cenocepacia J2315]	11468.82	81.34	1
197295256	zinc metalloprotease ZmpA [Burkholderia cenocepacia J2315]	502.77	21.24	3
206563581	putative porin [Burkholderia cenocepacia J2315]	229.67	17.17	2

Accession	# Unique Peptides	# Peptides	# PSMs	Molecular Function	# AAs	MW [kDa]	calc. pI
325521856	3	3	3	transporter activity	219	23.8	8.76
358076252	8	8	10	transporter activity	383	40.2	9.58
206562870	9	10	14	transporter activity	386	39.6	8.18
206561353	8	8	15	transporter activity	383	40.3	9.63
206563776	22	23	252	transporter activity	359	37.5	9.45
197295256	8	8	10	catalytic; metal binding	565	59.9	7.99
206563581	5	5	6	transporter activity	367	39.7	7.27

Figure 8.13: Mass spectroscopic investigation of the ~40 kDa major OM protein.

The MASCOT output summarised in Microsoft Excel suggests that the most likely identity of the ~40 kDa band (based upon factors such as the number of peptides analysed, highest MASCOT score and sequence coverage) is the putative porin with accession: **206563776**.

Although the 78% *B. mallei* homologue of this protein did not generate an immune response in *B. mallei*-infected human sera, it was still considered a potential diagnostic marker, due to its high occurrence in the outer membrane of *B. cenocepacia*.

therefore likely to be partially exposed on the cell surface, the lack of response was considered to be due to differing epitopes exposed on the cell-surface. The porin was however still considered to hold the potential of a diagnostic marker due to its high occurrence in the outer membrane fraction.

Chapter 9: Conclusions and Suggestions for Future Work

As described in Chapter 4, from the examination of Kyte-Doolittle hydrophathy plots (Kyte & Doolittle, 1982) of the entire proteome of *B. pseudomallei*, a short-list of potentially membrane-anchored proteins was created. Bioinformatical investigation shortened this list further to include only targets that exhibited no structural homologues in the PDB, with no folds predicted based upon sequence-threading and that for safety reasons had homologues in the non-virulent *B. thailandensis*. In addition, targets were selected that were not already under investigation by other structural genomics consortia, exhibited low predicted tertiary disorder favourable for crystallisation (Yang et al, 2005) and contained methionines for seleno-met phasing. The final target selected was BPSL1204, a putative lipoprotein with an N-terminal LLAGCA glyceryl-cysteine covalent lipid-anchoring motif.

The *bpsl1204* gene was obtained by PCR from the *B. pseudomallei* genome and truncated to remove the first 20 largely hydrophobic residues up to and including the LLAGCA motif. This was to solubilise the functional domain(s) by removing its insoluble anchoring region. Upon cloning pure PCR product into the pETBlue system and over-expressing in *E. coli* Tuner cells, BPSL1204-MA (minus membrane anchor) was purified from cell-free extract by DEAE ion-exchange, ammonium sulphate precipitation and gel filtration. 30 mg/ml of monomeric BPSL1204-MA was purified to >95% pure and used to carry out robotic, 96 well vapour diffusion crystallisation trials using the Hydra II microdispenser and screens from Nextal Qiagen.

Initial crystal hits were optimised over a period of months until large rod-clusters were obtained by 2 weeks growth in 9% MPD and 0.1 M sodium acetate pH 5.5. A 1.3 Å native dataset with cell dimensions $a=99.96$ Å, $b=32.91$ Å, $c=48.35$ Å, $\alpha =90.0^\circ$ $\beta=116.14^\circ$ and $\gamma = 90.0^\circ$ in space group C2 was collected at the Diamond Synchrotron Light Source in Oxford. Attempts to solve the phase problem by MAD/SAD data collected from heavy atom-soaked derivatives of native BPSL1204-MA generated no anomalous signal. Therefore over-expression

of BPSL1204-MA in *E. coli* grown in minimal media, denied methionine and supplemented with seleno-methionine allowed the incorporation of selenium into the primary sequence of BPSL1204-MA.

Once purified by similar methods to those used for native BPSL1204-MA, seleno-met BPSL1204-MA was crystallised into the same morphology and under similar conditions to the native protein by a process of dilution-seeding. MAD data isomorphous to native data was collected, processed in iMOSFLM (Battye et al, 2011) and CCP4i (Potterton et al, 2003), phases calculated and density modified in SHELX (Sheldrick, 2008) and then phases were extended to 1.3 Å native resolution. Later a further native dataset was collected to 1.05 Å, into which phases were extended. A model was built into the calculated electron density, using Coot (Emsley & Cowtan, 2004), and the structure was refined using the CCP4i program Refmac5 (Murshudov et al, 2011). The model was validated using MolProbity (Davis et al, 2007) and PROCHECK (Laskowski et al, 1998).

The final 1.05 Å resolution BPSL1204 structure was displayed in PyMol (Shrödinger, LLC), and revealed a protein consisting of an $\alpha 1\beta 4$ N-terminal domain and an $\alpha 2\beta 4$ C-terminal domain connected by a short loop. Structural comparison using the DALI server found the nearest structure with similarity to BPSL1204 to be the β -lactamase inhibitory protein I, BLIP-I (Joint Center for Structural Genomics), which is comprised of two $\alpha 2\beta 4$ domains. However, the domains in BLIP-I are orientated side-by-side, whilst the domains in BPSL1204 are stacked on top of each other. The two proteins were seen to have no sequence similarity to each other.

When the inhibitory N-terminal domain of BLIP-I in complex with the β -lactamase TEM1 (Strynadka et al, 1996) was superimposed with the N-terminal domain of BPSL1204, it was seen that the inhibitory loop of BLIP-I was matched by a loop with similar chemical properties in BPSL1204. However, assays designed to assess the inhibition of TEM1 using the chromogenic penicillin analogue Nitrocefin (CalbioChem) showed that BPSL1204 did not inhibit the *E. coli* TEM1. It was proposed that BPSL1204-MA possessed a biological function

that was either a BLIP-like role specific to *B. pseudomallei* or functioned in a process analogous but evolutionarily divergent to BLIP.

In order to obtain more biochemical data on the function of BPSL1204-MA, a pull-down assay was carried out with the aim of discovering potential binding partners of the protein. The process was modelled in the less dangerous *B. cenocepacia* using the 92% identical BPSL1204 homologue BCAL2351. Although a potential binding partner was isolated and purified, the methionyl-tRNA synthetase (metRS) BCAL2646, a complex was not seen to form between the two pure proteins upon gel filtration. Thus it was concluded that the pull-down interaction seen was an artefact of the bead-immobilised pull-down method and not a true biological interaction.

In order to determine the cellular location of BPSL1204/BCAL2351, and thus gain more information about BPSL1204, rat anti-BPSL1204-MA antibodies were produced. Antibody production in rats appeared to reflect a hyper-immune response, with 100 times more antibody produced than in a standard inoculation. Whole *B. cenocepacia* cells were subjected to fractionation via sucrose-gradient osmotic shock, sonication, ultra-centrifugation and selective Sarkosyl solubilisation to generate pure secreted, periplasmic, cytoplasmic, inner membrane and outer membrane samples. Western blot transfer of these fractions to a nitrocellulose membrane allowed for the probing of each fraction for the location of native BCAL2351 using rat anti-BPSL1204 as a primary antibody. A secondary rabbit anti-rat HRP-conjugate was used to identify sites of BCAL2351 binding by chemiluminescence of the HRP ECL-substrate visualised as bands on developed X-ray film. BCAL2351, and thus BPSL1204, was bound only in the pure outer membrane fraction with the only other band seen that of a pure BPSL1204 control lane.

In order to determine the orientation of BPSL1204 and BCAL2351 in the outer membrane, fluorescence microscopy of whole cells and pure, fractured outer membrane samples was carried out. Slides of immobilised whole *B. cenocepacia* cells and slides covered in immobilised, randomly orientated outer membrane

fragments were washed with rat anti-BPSL1204 primary antibody, to which a secondary goat anti-rat FITC-conjugate was bound. Upon laser excitation of the FITC, green fluorescent 'speckling' was seen in the fragmented outer membrane, representative of BCAL2351 binding. When applied to whole-cells, a portion of samples used exhibited fluorescence within the boundaries of the bacterium alone, suggestive of antibodies binding to surface-exposed BCAL2351/BPSL1204. Furthermore, binding appeared to specifically localise to patches of the cell surface rather than all over the bacterium, perhaps reflecting a specific biological function. No conclusive pattern was seen in localisation however. In contrast, no green fluorescence was seen in the remaining whole-cell slides. This may indicate that BCAL2351 and BPSL1204 are inward-facing, outer membrane proteins and thus not exposed to be bound by antibodies applied to whole cells.

It was proposed that BPSL1204 and BCAL2351 could be external outer membrane proteins selectively expressed or conditionally exposed on the cell surface to fulfil a specific biological function, perhaps associated with the cell-cycle. Conversely, it was hypothesised that BPSL1204 and BCAL2351 are both outer membrane proteins facing the periplasm, and that the inconsistent fluorescent signal seen in some slides was due to outer-membrane rupture-damage incurred during slide preparation, thus exposing the periplasm in patches.

A sample of serum from a human infected with *B. mallei* was obtained, a species that contains a 100% homologue of BPSL1204 and thus a 92% homolog of BCAL2351. After Western blot transfer to a nitrocellulose membrane, *B. cenocepacia* fractions were challenged with the anti-*B. mallei* human antibodies in the sera, which were then bound by a secondary goat anti-human HRP-conjugate. Proteins from all fractions that elicited an immune response during infection were located by the development of X-ray film exposed to the secondary HRP chemiluminescent conversion of ECL reagent upon the nitrocellulose membrane. Several unknown proteins were found to give a large

signal (compared to no signal in secondary antibody only controls), considered as potential targets for future research into diagnostic and therapeutic targets.

However, pure BPSL1204 and native BCAL2351 did not exhibit an immune response in *B. mallei* infection of humans. This lent support to the conclusion that BPSL1204/BCAL2351 is an internal outer membrane protein, as infection by whole *B. cenocepacia*, *B. pseudomallei* or *B. mallei* cells would not present BPSL1204 and its homologues to a host's immune system, and instead conceal it within the periplasmic space.

A series of cross-discipline experiments comprising structural X-ray crystallography, biochemical pull-down and enzyme assays combined with immunohistochemical and microbiological fluorescence experiments have been used to investigate BPSL1204. It was concluded that BPSL1204 possesses a two-domain BLIP-I-like fold, though it differs in domain orientation from BLIP-I and does not possess the same general β -lactamase inhibitory properties of BLIP-I. BPSL1204 also appeared not to pull down any strong binding partners from cell-extract. Localisation studies modelled using the *B. cenocepacia* BCAL2351 revealed BPSL1204 to be an outer membrane-anchored protein of unknown function covalently associated via a glyceryl-cysteine to the outer lipid membrane of *B. pseudomallei*. The orientation of BPSL1204 about the outer membrane is suggested to be periplasmic, but possibly may be conditionally found exposed on the cell-surface. When modelled in *B. mallei* infection of humans, BPSL1204 was found to be non-immunogenic.

In the future, further fluorescence microscopy experiments should be undertaken in order to determine whether the conclusion that BPSL1204 is facing the inside of the outer membrane is valid or not. All future experiments require that multiple repeats with a greater field of bacteria per slide be used in order to confirm that different results seen between different slides are due to experimentally altered factors and not variation in the quality of slide preparation. Further to this all microscopy experiments seen in this thesis, performed as preliminary investigations, used antibody dilutions of 1:600.

Optimisation of antibody dilutions used may allow for better analysis of binding events and maybe improve the resolution of specific sites of anti-BPSL1204/BCAL2351 binding.

Obtaining samples of cells in various, defined cell-cycle stages may allow the investigation of any stages at which BPSL1204 may be exposed on the cell-surface, perhaps due to a role in cell-wall transpeptidase regulation, a process evolutionarily analogous to β -lactamases. The application of a secondary fluorophore-reporter known to be periplasm-specific could act as a control for fluorescence due to outer membrane rupturing. In addition, the exact location of BPSL1204 and BCAL2351 could be further defined by the creation of a GFP-fusion BCAL2351 mutant in *B. cenocepacia*, in order to directly monitor the location of BCAL2351 in both fixed and live-cell fluorescence microscopy.

Furthermore, microscopic localisation studies could be supported by the development of a whole-cell protease assay in order to determine if BPSL1204 and BCAL2351 are exposed on the bacterial cell-surface and thus accessible to proteolysis by an enzyme specific to part of the target protein sequence. The protease proline-endopeptidase would cleave BCAL2351 after Pro¹¹³ and in BPSL1204 after Pro¹⁰⁹, therefore releasing an extracellular 53- or 57-residue polypeptide respectively. These could then be isolated by SDS-PAGE and subjected to MALDI-TOF MS to identify, or perhaps even identified by use of rat anti-BPSL1204 in an immunoblot assay if the released cleavage-product presented epitopes recognised by the antibody used.

Continuing with a microbiological theme, a potential route for the discovery of a biological function for BPSL1204 and BCAL2351 may entail the creation of a knock-out mutant for either *B. pseudomallei* or *B. cenocepacia*. Upon the removal of a functional *in vivo* protein, the phenotype of a knock-out strain may reveal a process in which BPSL1204 is involved, for example if involved in the regulation of cell-wall transpeptidases (analogous to β -lactamase activity) then morphological differences in cell-shape and cell-wall construction might be seen.

In addition, further crystallographic investigations of BPSL1204 homologues from multiple *Burkholderia* species should be undertaken, in order to conclusively determine the biological domain conformation of BPSL1204 and its homologues with regards to BLIP-I. Furthermore, the identification and isolation of β -lactamases and cell-wall transpeptidases from *Burkholderia* species (including *pseudomallei*) may allow for the investigation of complex formation, either by enzymatic assays or biophysical techniques such as isothermal titration calorimetry (ITC). Any complex formation seen could then be investigated crystallographically in order to discover structural aspects of the complexes.

Although to this stage BPSL1204 appeared to not elicit an immune response during human infection (when modelled in *B. mallei* infection), an extraordinarily high immune response against BPSL1204 was seen to occur in rats, with detection of BPSL1204 to the ng (and potentially pg) level comprehensively seen in immuno-blots. Therefore BPSL1204 was considered to hold the potential to act as a rapid diagnostic marker for *B. pseudomallei* infection, perhaps upon antibody probing of *in vitro* blood-samples from patients suspected to be infected with the bacterium. If a reliable and rapid ELISA-based diagnostic assay for BPSL1204 could be developed, it may help speed the identification of *B. pseudomallei* infection, leading to quicker application of therapeutic treatments against the lethal and potentially fast-acting disease melioidosis.

References

- Aldhous P (2005) Tropical medicine: melioidosis? Never heard of it. *Nature* **434**(7034): 692-693
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**(3): 403-410
- Ashdown LR (1979) Nosocomial infection due to *Pseudomonas pseudomallei*: two cases and an epidemiologic study. *Rev Infect Dis* **1**(5): 891-894
- Axford D, Owen RL, Aishima J, Foadi J, Morgan AW, Robinson JI, Nettleship JE, Owens RJ, Moraes I, Fry EE, Grimes JM, Harlos K, Kotecha A, Ren J, Sutton G, Walter TS, Stuart DI, Evans G (2012) In situ macromolecular crystallography using microbeams. *Acta Crystallogr D Biol Crystallogr* **68**(Pt 5): 592-600
- Banse AV, Hobbs EC, Losick R (2011) Phosphorylation of Spo0A by the histidine kinase KinD requires the lipoprotein med in *Bacillus subtilis*. *J Bacteriol* **193**(15): 3949-3955
- Barlow DJ, Thornton JM (1983) Ion-pairs in proteins. *J Mol Biol* **168**(4): 867-885
- Bartlett JM, Stirling D (2003) A short history of the polymerase chain reaction. *Methods Mol Biol* **226**: 3-6
- Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 271-281
- Ben-Bassat A, Bauer K, Chang SY, Myambo K, Boosman A, Chang S (1987) Processing of the initiation methionine from proteins: properties of the *Escherichia coli* methionine aminopeptidase and its gene structure. *J Bacteriol* **169**(2): 751-757
- Benvenuti M, Mangani S (2007) Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat Protoc* **2**(7): 1633-1651
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**(1): 235-242
- Bertani G (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J Bacteriol* **186**(3): 595-600
- Bianco N, Neshat S, Poole K (1997) Conservation of the multidrug resistance efflux gene *oprM* in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **41**(4): 853-856
- Bishop RE (2000) The bacterial lipocalins. *Biochim Biophys Acta* **1482**(1-2): 73-83

- Blow DM (2003) How Bijvoet made the difference: the growing power of anomalous scattering. *Methods Enzymol* **374**: 3-22
- Bonk T, Humeny A (2001) MALDI-TOF-MS analysis of protein and DNA. *Neuroscientist* **7**(1): 6-12
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254
- Bragg W.L. (1913). "The Diffraction of Short Electromagnetic Waves by a Crystal". *Proceedings of the Cambridge Philosophical Society* **17**: 43-57.
- Bramhill D, Thompson CM (1994) GTP-dependent polymerization of Escherichia coli FtsZ protein to form tubules. *Proc Natl Acad Sci U S A* **91**(13): 5813-5817
- Brunger AT (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* **49**(Pt 1): 24-36
- Buchanan SK (1999) Beta-barrel proteins from bacterial outer membranes: structure, function and refolding. *Curr Opin Struct Biol* **9**(4): 455-461
- Burnette WN (1981) "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal Biochem* **112**(2): 195-203
- Carballido-Lopez R, Errington J (2003) A dynamic bacterial cytoskeleton. *Trends Cell Biol* **13**(11): 577-583
- Carugo O, Bordo D (1999) How many water molecules can be detected by protein crystallography? *Acta Crystallogr D Biol Crystallogr* **55**(Pt 2): 479-483
- Chaowagul W, White NJ, Dance DA, Wattanagoon Y, Naigowit P, Davis TM, Looareesuwan S, Pitakwatchara N (1989) Melioidosis: a major cause of community-acquired septicemia in northeastern Thailand. *J Infect Dis* **159**(5): 890-899
- Chayen NE (1998) Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr D Biol Crystallogr* **54**(Pt 1): 8-15
- Chayen NE, Saridakis E (2008) Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods* **5**(2): 147-153

- Chen L, Oughtred R, Berman HM, Westbrook J (2004) TargetDB: a target registration database for structural genomics projects. *Bioinformatics* **20**(16): 2860-2862
- Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 12-21
- Cheng AC, Currie BJ (2005) Melioidosis: epidemiology, pathophysiology, and management. *Clin Microbiol Rev* **18**(2): 383-416
- Chesnel L, Pernot L, Lemaire D, Champelovier D, Croize J, Dideberg O, Vernet T, Zapun A (2003) The structural modifications induced by the M339F substitution in PBP2x from *Streptococcus pneumoniae* further decreases the susceptibility to beta-lactams of resistant strains. *J Biol Chem* **278**(45): 44448-44456
- Cladiere L, Hamze K, Madec E, Levnikov VM, Wilkinson AJ, Holland IB, Seror SJ (2006) The GTPase, CpgA(YloQ), a putative translation factor, is implicated in morphogenesis in *Bacillus subtilis*. *Mol Genet Genomics* **275**(4): 409-420
- Cottrell JS (2012) Protein identification using MS/MS data. *J Proteomics* **74**(10): 1842-1851
- Cowtan KD, Main P (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1): 43-48
- Cruz-Migoni A, Ruzhenikov SN, Sedelnikova SE, Obeng B, Chieng S, Mohamed R, Nathan S, Baker PJ, Rice DW (2011) Cloning, purification and crystallographic analysis of a hypothetical protein, BPSL1549, from *Burkholderia pseudomallei*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**(Pt 12): 1623-1626
- Currie BJ, Fisher DA, Howard DM, Burrow JN, Lo D, Selva-Nayagam S, Anstey NM, Huffam SE, Snelling PL, Marks PJ, Stephens DP, Lum GD, Jacups SP, Krause VL (2000) Endemic melioidosis in tropical northern Australia: a 10-year prospective study and review of the literature. *Clin Infect Dis* **31**(4): 981-986
- Dale GE, Oefner C, D'Arcy A (2003) The protein as a variable in protein crystallization. *J Struct Biol* **142**(1): 88-97
- Dance DA (1991) Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* **4**(1): 52-60
- Dance DA (2000) Melioidosis as an emerging global problem. 74 Thesis, Acta Trop,
- Dance DA, Smith MD, Aucken HM, Pitt TL (1999) Imported melioidosis in England and Wales. *Lancet* **353**(9148): 208

Dance DA, Wuthiekanun V, Naigowit P, White NJ (1989) Identification of *Pseudomonas pseudomallei* in clinical practice: use of simple screening tests and API 20NE. *J Clin Pathol* **42**(6): 645-648

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**(Web Server issue): W375-383

Debye, P (1913). "Interferenz von Röntgenstrahlen und Wärmebewegung". *Ann. d. Phys.* **348** (1): 49-92

DeLano WL, Brunger AT (1995) The direct rotation function: Patterson correlation search applied to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **51**(Pt 5): 740-748

DeShazer D (2004) Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in *B. pseudomallei* 1026b. *J Bacteriol* **186**(12): 3938-3950

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* **312**(5771): 212-217

Ealick SE (2000) Advances in multiple wavelength anomalous diffraction crystallography. *Curr Opin Chem Biol* **4**(5): 495-499

Eisenberg D, Schwarz E, Komaromy M, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol* **179**(1): 125-142

Einstein, A (1926) *Investigations on the Theory of Brownian Movement*, ed. R. Fürth, translated by A.D. Cowper; *Collected Papers*, vol. 2, 170-82, 206-22.

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132

Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 72-82

Ewald, P. P. (1969). "Introduction to the dynamical theory of X-ray diffraction". *Acta Crystallographica Section A* **25**: 103.

Eyal E, Chennubhotla C, Yang LW, Bahar I (2007) Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. *Bioinformatics* **23**(13): i175-184

- Falagas ME, Kasiakou SK (2005) Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. *Clin Infect Dis* **40**(9): 1333-1341
- Frey A, Meckelein B, Externest D, Schmidt MA (2000) A stable and highly sensitive 3,3',5,5'-tetramethylbenzidine-based substrate reagent for enzyme-linked immunosorbent assays. *J Immunol Methods* **233**(1-2): 47-56
- Fujita M, Kinoshita T (2012) GPI-anchor remodeling: Potential functions of GPI-anchors in intracellular trafficking and membrane dynamics. *Biochim Biophys Acta* **1821**(8): 1050-1058
- Galyov EE, Brett PJ, DeShazer D (2010) Molecular insights into Burkholderia pseudomallei and Burkholderia mallei pathogenesis. *Annu Rev Microbiol* **64**: 495-517
- Gandhi G, Londono D, Whetstine CR, Sethi N, Kim KS, Zuckert WR, Cadavid D (2010) Interaction of variable bacterial outer membrane lipoproteins with brain endothelium. *PLoS One* **5**(10): e13257
- Garman E (2003) 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr Opin Struct Biol* **13**(5): 545-551
- Garman E, Nave C (2002) Radiation damage to crystalline biological molecules: current view. *J Synchrotron Radiat* **9**(Pt 6): 327-328
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**(13): 3784-3788
- Govindarajan S, Recabarren R, Goldstein RA (1999) Estimating the total number of protein folds. *Proteins* **35**(4): 408-414
- Gretes M, Lim DC, de Castro L, Jensen SE, Kang SG, Lee KJ, Strynadka NC (2009) Insights into positive and negative requirements for protein-protein interactions by crystallographic analysis of the beta-lactamase inhibitory proteins BLIP, BLIP-I, and BLP. *J Mol Biol* **389**(2): 289-305
- Grosse-Kunstleve RW, Adams PD (2003) Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 11): 1966-1973
- Guerin ME, Buschiazzi A, Ugalde JE, Ugalde RA, Alzari PM (2003) Preliminary crystallographic studies of glycogen synthase from Agrobacterium tumefaciens. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 3): 526-528
- Hammersley AP, Brown K, Burmeister W, Claustre L, Gonzalez A, McSweeney S, Mitchell E, Moy JP, Svensson SO, Thompson AW (1997) Calibration and application of an X-ray image intensifier/ charge-coupled device detector for

monochromatic macromolecular crystallography. *J Synchrotron Radiat* **4**(Pt 2): 67-77

Harland DN, Chu K, Haque A, Nelson M, Walker NJ, Sarkar-Tyson M, Atkins TP, Moore B, Brown KA, Bancroft G, Titball RW, Atkins HS (2007) Identification of a LolC homologue in *Burkholderia pseudomallei*, a novel protective antigen for melioidosis. *Infect Immun* **75**(8): 4173-4180

Hayashi S, Wu HC (1990) Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**(3): 451-471

Hayes MV, Orr DC (1983) Mode of action of ceftazidime: affinity for the penicillin-binding proteins of *Escherichia coli* K12, *Pseudomonas aeruginosa* and *Staphylococcus aureus*. *J Antimicrob Chemother* **12**(2): 119-126

Heger A, Holm L (2000) Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* **41**(2): 224-237

Hendrickson WA (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**(5028): 51-58

Hendrickson WA (2000) Synchrotron crystallography. *Trends Biochem Sci* **25**(12): 637-643

Hendrickson WA, Smith JL, Sheriff S (1985) Direct phase determination based on anomalous scattering. *Methods Enzymol* **115**: 41-55

Heras B, Martin JL (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* **61**(Pt 9): 1173-1180

Hodeau JL, Favre-Nicolin V, Bos S, Renevier H, Lorenzo E, Berar JF (2001) Resonant diffraction. *Chem Rev* **101**(6): 1843-1867

Holden MT, Titball RW, Peacock SJ, Cerdeno-Tarraga AM, Atkins T, Crossman LC, Pitt T, Churcher C, Mungall K, Bentley SD, Sebaihia M, Thomson NR, Bason N, Beacham IR, Brooks K, Brown KA, Brown NF, Challis GL, Cherevach I, Chillingworth T, Cronin A, Crossett B, Davis P, DeShazer D, Feltwell T, Fraser A, Hance Z, Hauser H, Holroyd S, Jagels K, Keith KE, Maddison M, Moule S, Price C, Quail MA, Rabinowitsch E, Rutherford K, Sanders M, Simmonds M, Songsivilai S, Stevens K, Tumapa S, Vesaratchavest M, Whitehead S, Yeats C, Barrell BG, Oyston PC, Parkhill J (2004) Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A* **101**(39): 14240-14245

Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**(23): 2780-2781

Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**(Web Server issue): W545-549

Hopwood D (1972) Theoretical and practical aspects of glutaraldehyde fixation. *Histochem J* **4**(4): 267-303

Horii T, Kobayashi M, Sato K, Ichiyama S, Ohta M (1998) An in-vitro study of carbapenem-induced morphological changes and endotoxin release in clinical isolates of gram-negative bacilli. *J Antimicrob Chemother* **41**(4): 435-442

Howard J, Hyman AA (2003) Dynamics and mechanics of the microtubule plus end. *Nature* **422**(6933): 753-758

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R (2005) The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **40**(4): 430-443

Huang WM, Gibson SJ, Facer P, Gu J, Polak JM (1983) Improved section adhesion for immunocytochemistry using high molecular weight polymers of L-lysine as a slide coating. *Histochemistry* **77**(2): 275-279

Ilari A, Savino C (2008) Protein structure determination by x-ray crystallography. *Methods Mol Biol* **452**: 63-87

Inglis TJ, Garrow SC, Henderson M, Clair A, Sampson J, O'Reilly L, Cameron B (2000) Burkholderia pseudomallei traced to water treatment plant in Australia. *Emerg Infect Dis* **6**(1): 56-59

Jelsch C, Mourey L, Masson JM, Samama JP (1993) Crystal structure of Escherichia coli TEM1 beta-lactamase at 1.8 Å resolution. *Proteins* **16**(4): 364-383

Jones AM, Dodd ME, Govan JR, Barcus V, Doherty CJ, Morris J, Webb AK (2004) Burkholderia cenocepacia and Burkholderia multivorans: influence on survival in cystic fibrosis. *Thorax* **59**(11): 948-951

Jones DT (1999a) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287**(4): 797-815

Jones DT (1999b) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**(2): 195-202

Judge RA, Jacobs RS, Frazier T, Snell EH, Pusey ML (1999) The effect of temperature and solution pH on the nucleation of tetragonal lysozyme crystals. *Biophys J* **77**(3): 1585-1593

Kang G, Rajan DP, Ramakrishna BS, Aucken HM, Dance DA (1996) Melioidosis in India. *Lancet* **347**(9014): 1565-1566

Kantardjieff KA, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**(9): 1865-1871

- Kendrew JC, Perutz MF (1957) X-ray studies of compounds of biological interest. *Annu Rev Biochem* **26**: 327-372
- Kespichayawattana W, Rattanachetkul S, Wanun T, Utaisincharoen P, Sirisinha S (2000) Burkholderia pseudomallei induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. *Infect Immun* **68**(9): 5377-5384
- King TP, Kochoumian L (1979) A comparison of different enzyme-antibody conjugates for enzyme-linked immunosorbent assay. *J Immunol Methods* **28**(3-4): 201-210
- Knox JR, Moews PC, Frere JM (1996) Molecular evolution of bacterial beta-lactam resistance. *Chem Biol* **3**(11): 937-947
- Koshland D, Botstein D (1980) Secretion of beta-lactamase requires the carboxy end of the protein. *Cell* **20**(3): 749-760
- Kovacs-Simon A, Titball RW, Michell SL Lipoproteins of bacterial pathogens. *Infect Immun* **79**(2): 548-561
- Kraft P, Bergamaschi A, Broennimann C, Dinapoli R, Eikenberry EF, Henrich B, Johnson I, Mozzanica A, Schleputz CM, Willmott PR, Schmitt B (2009) Performance of single-photon-counting PILATUS detector modules. *J Synchrotron Radiat* **16**(Pt 3): 368-375
- Kricka LJ (1991) Chemiluminescent and bioluminescent techniques. *Clin Chem* **37**(9): 1472-1481
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**(1): 105-132
- Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**(5259): 680-685
- Langer G, Cohen SX, Lamzin VS, Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**(7): 1171-1179
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**(21): 2947-2948
- Laskowski RA (2009) PDBsum new things. *Nucleic Acids Res* **37**(Database issue): D355-359
- Laskowski RA, MacArthur MW, Thornton JM (1998) Validation of protein models derived from experiment. *Curr Opin Struct Biol* **8**(5): 631-639

Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* **231**(4): 1049-1067

Lee SH, Hannink M (2001) The N-terminal nuclear export sequence of IkkappaBalpha is required for RanGTP-dependent binding to CRM1. *J Biol Chem* **276**(26): 23599-23606

Leelarasamee A (1998) Burkholderia pseudomallei: the unbeatable foe? *Southeast Asian J Trop Med Public Health* **29**(2): 410-415

Leslie AG (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 48-57

Lever MS, Nelson M, Ireland PI, Stagg AJ, Beedham RJ, Hall GA, Knight G, Titball RW (2003) Experimental aerogenic Burkholderia mallei (glanders) infection in the BALB/c mouse. *J Med Microbiol* **52**(Pt 12): 1109-1115

Leverrier P, Declercq JP, Denoncin K, Vertommen D, Hiniker A, Cho SH, Collet JF (2011) Crystal structure of the outer membrane protein RcsF, a new substrate for the periplasmic protein-disulfide isomerase DsbC. *J Biol Chem* **286**(19): 16734-16742

Li L, Lu Z, Han O (1994) [Epidemiology of melioidosis in China]. *Zhonghua Liu Xing Bing Xue Za Zhi* **15**(5): 292-295

Li Z, Gillis KA, Hegg LA, Zhang J, Thurlow DL (1996) Effects of nucleotide substitutions within the T-loop of precursor tRNAs on interaction with ATP/CTP:tRNA nucleotidyltransferases from Escherichia coli and yeast. *Biochem J* **314** (Pt 1): 49-53

Lichtman JW, Conchello JA (2005) Fluorescence microscopy. *Nat Methods* **2**(12): 910-919

Liscum L, Faust JR (1987) Low density lipoprotein (LDL)-mediated suppression of cholesterol synthesis and LDL uptake is defective in Niemann-Pick type C fibroblasts. *J Biol Chem* **262**(35): 17002-17008

Mahenthalingam E, Baldwin A, Dowson CG (2008) Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. *J Appl Microbiol* **104**(6): 1539-1551

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**(Pt 4): 658-674

Mikol V, Rodeau JL, Giege R (1990) Experimental determination of water equilibration rates in the hanging drop method of protein crystallization. *Anal Biochem* **186**(2): 332-339

- Moore RA, DeShazer D, Reckseidler S, Weissman A, Woods DE (1999) Efflux-mediated aminoglycoside and macrolide resistance in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother* **43**(3): 465-470
- Morris RJ, Bricogne G (2003) Sheldrick's 1.2 Å rule and beyond. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 3): 615-617
- Morris RJ, Perrakis A, Lamzin VS (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**: 229-244
- Mukhopadhyay S, Cho W (1996) Interactions of annexin V with phospholipid monolayers. *Biochim Biophys Acta* **1279**(1): 58-62
- Munoz V, Blanco FJ, Serrano L (1995) The hydrophobic-staple motif and a role for loop-residues in alpha-helix stability and protein folding. *Nat Struct Biol* **2**(5): 380-385
- Murray JW, Rudino-Pinera E, Owen RL, Grininger M, Ravelli RB, Garman EF (2005) Parameters affecting the X-ray dose absorbed by macromolecular crystals. *J Synchrotron Radiat* **12**(Pt 3): 268-275
- Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 355-367
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-255
- Nestle M, Roberts WK (1969) An extracellular nuclease from *Serratia marcescens*. I. Purification and some properties of the enzyme. *J Biol Chem* **244**(19): 5213-5218
- Neubauer H, Sprague LD, Zacharia R, Tomaso H, Al Dahouk S, Wernery R, Wernery U, Scholz HC (2005) Serodiagnosis of *Burkholderia mallei* infections in horses: state-of-the-art and perspectives. *J Vet Med B Infect Dis Vet Public Health* **52**(5): 201-205
- Ngauy V, Lemeshev Y, Sadkowski L, Crawford G (2005) Cutaneous melioidosis in a man who was taken as a prisoner of war by the Japanese during World War II. *J Clin Microbiol* **43**(2): 970-972
- Niall HD (1973) Automated Edman degradation: the protein sequenator. *Methods Enzymol* **27**: 942-1010
- Nielsen JB, Lampen JO (1982) Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria. *J Bacteriol* **152**(1): 315-322

Nierman WC, DeShazer D, Kim HS, Tettelin H, Nelson KE, Feldblyum T, Ulrich RL, Ronning CM, Brinkac LM, Daugherty SC, Davidsen TD, Deboy RT, Dimitrov G, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Khouri H, Kolonay JF, Madupu R, Mohammoud Y, Nelson WC, Radune D, Romero CM, Sarria S, Selengut J, Shamblin C, Sullivan SA, White O, Yu Y, Zafar N, Zhou L, Fraser CM (2004) Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci U S A* **101**(39): 14246-14251

Osborn MJ, Wu HC (1980) Proteins of the outer membrane of gram-negative bacteria. *Annu Rev Microbiol* **34**: 369-422

Otto K, Silhavy TJ (2002) Surface sensing and adhesion of *Escherichia coli* controlled by the Cpx-signaling pathway. *Proc Natl Acad Sci U S A* **99**(4): 2287-2292

Padilla JE, Yeates TO (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 7): 1124-1130

Pal A, Kraetzner R, Gruene T, Grapp M, Schreiber K, Gronborg M, Urlaub H, Becker S, Asif AR, Gartner J, Sheldrick GM, Steinfeld R (2009) Structure of tripeptidyl-peptidase I provides insight into the molecular basis of late infantile neuronal ceroid lipofuscinosis. *J Biol Chem* **284**(6): 3976-3984

Peacock SJ, Chieng G, Cheng AC, Dance DA, Amornchai P, Wongsuvan G, Teerawattanasook N, Chierakul W, Day NP, Wuthiekanun V (2005) Comparison of Ashdown's medium, *Burkholderia cepacia* medium, and *Burkholderia pseudomallei* selective agar for clinical isolation of *Burkholderia pseudomallei*. *J Clin Microbiol* **43**(10): 5359-5361

Perkins SJ (1988) Structural studies of proteins by high-flux X-ray and neutron solution scattering. *Biochem J* **254**(2): 313-327

Pflugrath JW (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 10): 1718-1725

Phillips JC, Wlodawer A, Yevitz MM, Hodgson KO (1976) Applications of synchrotron radiation to protein crystallography: preliminary results. *Proc Natl Acad Sci U S A* **73**(1): 128-132

Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 7): 1131-1137

Puthucheary SD, Parasakthi N, Lee MK (1992) Septicaemic melioidosis: a review of 50 cases from Malaysia. *Trans R Soc Trop Med Hyg* **86**(6): 683-685

Raina S, Missiakas D (1997) Making and breaking disulfide bonds. *Annu Rev Microbiol* **51**: 179-202

- Rajendran L, Simons K (2005) Lipid rafts and membrane dynamics. *J Cell Sci* **118**(Pt 6): 1099-1102
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**: 95-99
- Read RJ (1997) [Model phases: probabilities and bias. *Methods Enzymol* **277**: 110-128
- Reechaipichitkul W (2004) Pulmonary melioidosis presenting with right paratracheal mass. *Southeast Asian J Trop Med Public Health* **35**(2): 384-387
- Reynolds KA, Thomson JM, Corbett KD, Bethel CR, Berger JM, Kirsch JF, Bonomo RA, Handel TM (2006) Structural and computational characterization of the SHV-1 beta-lactamase-beta-lactamase inhibitor protein interface. *J Biol Chem* **281**(36): 26745-26753
- Riggs JL, Seiwald RJ, Burckhalter JH, Downs CM, Metcalf TG (1958) Isothiocyanate compounds as fluorescent labeling agents for immune serum. *Am J Pathol* **34**(6): 1081-1097
- Ruzheinikov SN, Das SK, Sedelnikova SE, Baker PJ, Artymiuk PJ, Garcia-Lara J, Foster SJ, Rice DW (2004) Analysis of the open and closed conformations of the GTP-binding protein YsxC from *Bacillus subtilis*. *J Mol Biol* **339**(2): 265-278
- Samraoui B, Sutton BJ, Todd RJ, Artymiuk PJ, Waley SG, Phillips DC (1986) Tertiary structural similarity between a class A beta-lactamase and a penicillin-sensitive D-alanyl carboxypeptidase-transpeptidase. *Nature* **320**(6060): 378-380
- Sapay N, Guermeur Y, Deleage G (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics* **7**: 255
- Saravu K, Mukhopadhyay C, Eshwara VK, Shastry BA, Ramamoorthy K, Krishna S, Sathyanarayanan V (2012) Melioidosis presenting with mediastinal lymphadenopathy masquerading as malignancy: a case report. *J Med Case Rep* **6**(1): 28
- Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**(Pt 10 Pt 2): 1772-1779
- Scholz HC, Joseph M, Tomaso H, Al Dahouk S, Witte A, Kinne J, Hagen RM, Wernery R, Wernery U, Neubauer H (2006) Detection of the reemerging agent *Burkholderia mallei* in a recent outbreak of glanders in the United Arab Emirates by a newly developed fliP-based polymerase chain reaction assay. *Diagn Microbiol Infect Dis* **54**(4): 241-247

- Seubert A, Hiestand R, de la Cruz F, Dehio C (2003) A bacterial conjugation machinery recruited for pathogenesis. *Mol Microbiol* **49**(5): 1253-1266
- Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* **64**(Pt 1): 112-122
- Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 479-485
- Sheldrick GM, Dauter Z, Wilson KS, Hope H, Sieker LC (1993) The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Crystallogr D Biol Crystallogr* **49**(Pt 1): 18-23
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**(6): 2856-2860
- Staden R (1994) Staden: statistical and structural analysis of protein sequences. *Methods Mol Biol* **25**: 125-130
- Stehle T, Gamblin SJ, Yan Y, Harrison SC (1996) The structure of simian virus 40 refined at 3.1 Å resolution. *Structure* **4**(2): 165-182
- Stone R (2007) Infectious disease. Racing to defuse a bacterial time bomb. *Science* **317**(5841): 1022-1024
- Strub MP, Hoh F, Sanchez JF, Strub JM, Bock A, Aumelas A, Dumas C (2003) Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **11**(11): 1359-1367
- Strynadka NC, Jensen SE, Alzari PM, James MN (1996) A potent new mode of beta-lactamase inhibition revealed by the 1.7 Å X-ray crystallographic structure of the TEM-1-BLIP complex. *Nat Struct Biol* **3**(3): 290-297
- Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**(1): 113-130
- Suparak S, Kespichayawattana W, Haque A, Easton A, Damnin S, Lertmemongkolchai G, Bancroft GJ, Korbsrisate S (2005) Multinucleated giant cell formation and apoptosis in infected host cells is mediated by Burkholderia pseudomallei type III secretion protein BipB. *J Bacteriol* **187**(18): 6556-6560
- Swairjo MA, Morales AJ, Wang CC, Ortiz AR, Schimmel P (2000) Crystal structure of trbp111: a structure-specific tRNA-binding protein. *EMBO J* **19**(23): 6287-6298
- Taylor G (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* **59**(Pt 11): 1881-1890

Taylor GL (2010) Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 325-338

Ten Eyck LF (1985) Fast Fourier transform calculation of electron density maps. *Methods Enzymol* **115**: 324-337

Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* **56**(Pt 8): 965-972

Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 4): 849-861

Thomas DH, Rob A, Rice DW (1989) A novel dialysis procedure for the crystallization of proteins. *Protein Eng* **2**(6): 489-491

Tomaso H, Pitt TL, Landt O, Al Dahouk S, Scholz HC, Reisinger EC, Sprague LD, Rathmann I, Neubauer H (2005) Rapid presumptive identification of *Burkholderia pseudomallei* with real-time PCR assays using fluorescent hybridization probes. *Mol Cell Probes* **19**(1): 9-20

Torrens JK, McWhinney PH, Tompkins DS (1999) A deadly thorn: a case of imported melioidosis. *Lancet* **353**(9157): 1016

Tronrud DE (2004) Introduction to macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2156-2168

Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2184-2195

Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr* **55**(Pt 1): 191-205

Vianney A, Muller MM, Clavel T, Lazzaroni JC, Portalier R, Webster RE (1996) Characterization of the tol-pal region of *Escherichia coli* K-12: translational control of tolR expression by TolQ and identification of a new open reading frame downstream of pal encoding a periplasmic protein. *J Bacteriol* **178**(14): 4031-4038

Vidyalakshmi K, Chakrapani M, Shrikala B, Damodar S, Lipika S, Vishal S (2008) Tuberculosis mimicked by melioidosis. *Int J Tuberc Lung Dis* **12**(10): 1209-1215

Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res* **31**(13): 3688-3691

- von Heijne G, Abrahmsen L (1989) Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts. *FEBS Lett* **244**(2): 439-446
- Vongphayloth K, Rattanavong S, Moore CE, Phetsouvanh R, Wuthiekanun V, Sengdouangphachanh A, Phouminh P, Newton PN, Buisson Y (2012) Burkholderia pseudomallei detection in surface water in southern Laos using Moore's swabs. *Am J Trop Med Hyg* **86**(5): 872-877
- Walburger A, Lazdunski C, Corda Y (2002) The Tol/Pal system function requires an interaction between the C-terminal domain of TolA and the N-terminal domain of TolB. *Mol Microbiol* **44**(3): 695-708
- Waller, I (1923). "Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen". *Zeitschrift für Physik A* (Berlin / Heidelberg: Springer) **17**: 398-408.
- Wang BC (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol* **115**: 90-112
- White NJ (2003) Melioidosis. *Lancet* **361**(9370): 1715-1722
- Wiersinga WJ, van der Poll T, White NJ, Day NP, Peacock SJ (2006) Melioidosis: insights into the pathogenicity of Burkholderia pseudomallei. *Nat Rev Microbiol* **4**(4): 272-282
- Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 235-242
- Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* **374**: 300-321
- Wukovitz SW, Yeates TO (1995) Why protein crystals favour some space-groups over others. *Nat Struct Biol* **2**(12): 1062-1067
- Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M (1992) Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. *Microbiol Immunol* **36**(12): 1251-1275
- Yamaguchi K, Yu F, Inouye M (1988) A single amino acid determinant of the membrane localization of lipoproteins in E. coli. *Cell* **53**(3): 423-432

Yang ZR, Thomson R, McNeil P, Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**(16): 3369-3376

Yoshizawa S, Fourmy D, Puglisi JD (1998) Structural origins of gentamicin antibiotic action. *EMBO J* **17**(22): 6437-6448

Zimmermann L, Peterhans E, Frey J (2010) RGD motif of lipoprotein T, involved in adhesion of *Mycoplasma conjunctivae* to lamb synovial tissue cells. *J Bacteriol* **192**(14): 3773-3779

Zwart PH (2005) Anomalous signal indicators in protein crystallography. *Acta Crystallogr D Biol Crystallogr* **61**(Pt 11): 1437-1448