A study of Outer Membrane Biogenesis in *Escherichia coli*

by

Riyaz Maderbocus

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

School of Cancer Sciences
University of Birmingham
September 2012

ABSTRACT

The outer membrane (OM) of *Escherichia coli* is an essential organelle. The OM allows *E. coli* to interact with its environment and has a critical function as a barrier to prevent the entry of toxic molecules into the cell. The OM is composed of phospholipids, lipoproteins, outer membrane β-barrel proteins (OMPs) and lipopolysaccharide (LPS). The correct ratio of these components is needed to ensure proper OM barrier function is maintained. Assembly of OMPs is performed by the Bam (β-barrel assembly machinery) complex, lipoproteins by the Lol (Localisation of lipoproteins) pathway and LPS by the Lpt (LPS transport) pathway. The factors responsible for the assembly of phospholipids at the OM are unknown.

This study presents two key areas in understanding OM biogenesis. Firstly, a comprehensive mutagenesis screen was performed on the Bam complex member BamE. This analysis along with the structure of BamE has indicated crucial regions for BamE function. Secondly, we have performed a structure and function analysis on the previously uncharacterised protein, YraP. The structure of YraP has been solved and represents a novel fold. Additionally, we have obtained some functional evidence that suggest that YraP is involved in phospholipid biogenesis in the OM.

Acknowledgements

First, I would like to thank my supervisors. Prof. Ian Henderson, who has given me continual support and great advice during my PhD. Likewise, Prof. Michael Overduin, who has also given a huge amount of support. I am indebted to their patience and guidance during this time and I am sure at times they must have felt like they were backing the wrong horse.

Really big thanks to Tim Knowles. His guidance has steered me onto the right track during this time and his invaluable help that led to solving the structure of YraP. I am very grateful for him going over countless drafts of the thesis and taking time to explain things. Working with him on a daily basis has been a joy over the past years.

Also, big thanks to Doug Browning, who helped me clone a large number of things during the BamE part of the project. I am very grateful for his help and advice and it was really fun to work with him.

I would also like to thank my colleagues, the Hendersons and the Overduins. In the Henderson camp, Faye (for allowing me to borrow her Eppys), Tim (for sorting out the phylogenetic tree), Yana (for putting up with me), Amanda (for being optimistic), Denisse (for also putting up with me), Riddhi (for being super nice), Cathy (for keeping things together and filling me up with hope during thesis writing), Karina (for also putting up with me), Matt (for being funny) and Andy (for knowing everything). All of these people have contributed to this work one way or another and made the time working in the Henderson lab a great experience.

In the Overduin camp, firstly I'd like to thank Piv. His presence in the Overduin lab has made it an unforgettable experience and I am going to miss nights either in the lab or playing our instruments terribly and watching bad films. I also thank him for letting me stay at his place during the period of printing and binding the thesis. I thank my colleagues Zee and Mike 'Giggsy' Tong, working with them has also been unforgettable and sadly an acknowledgment in a Thesis does not do the times we've spent together justice. I also thank Claudia (for talking about flippy-flippy nonsense), Nazia, Vicky, Penny, Jeeves, Monsieur Lenoir and Rajesh. I also thank the technicians, Pooja, Sandya and Jas, who do countless things to make sure the lab works, as well being really nice people to work with. Thank you all very much.

I also thank my family, who have been with me through good times and bad. Special thanks to my sister Hanna, who helped with a ton of tech problems during the last weeks of writing up and saved me loads of time.

I'd like to also thank my friends, Chris 'Stegabilly Slick', Amy, Phil, Ollie, Chris, Nick, Matt, Paul, Andrew, Ray, Ed, Daniel and Jerry.

I would also like to thank Phil, Andrew and Chris allowing me to perform at their various pubs. Sorry about the lacklustre 'material' that was under the guise of 'comedy', but you guys did grant me a great release and lots of laughs during the PhD.

I would also like to dedicate this work to my good friend Chris Gamble, who passed away during my PhD. RIP buddy.

Table of Contents

Chapter 1 Introduction 1

1.1 Introduction 2

1.2 Topological problem of OM biogenesis 6

1.3 Transport through the IM 7

1.4 Peptidoglycan 8

1.5 Regulation systems involved with OM homeostasis 9

1.6 Lipopolysaccharide assembly 11

1.6.1.1 Synthesis of lipid A 13

1.6.1.2 Synthesis of core oligosaccharide 14

1.6.1.3 MsbA 16

1.6.1.4 O-antigen 17

1.6.1.5 LPS transport 18

1.6.1.6 Extraction from the Inner Membrane - LptBCFG complex 18

1.6.1.7 Translocation across the periplasm - LptA 19

1.6.1.8 Insertion into the Outer Membrane - LptD and LptE 20

1.6.1.9 Concluding remarks 21

1.7 Phospholipids 23

1.7.1.1 Maintaining the asymmetry of the OM 25

1.7.1.2 PldA 25

1.7.1.3 PagP 29

1.7.1.4 The Mla pathway 32

1.7.1.5 Concluding remarks 34

1.8 Lipoproteins 36

1.8.1 Biogenesis of lipoproteins 38

1.8.2 Sorting of lipoproteins 39

1.8.3 The Lol system 41

1.8.3.1 The LolCDE-Lipoprotein complex 41

1.8.3.2 LolA 42

1.8.3.3 LolB 43

1.8.3.4 Lipoprotein Transfer 45

1.8.4 Concluding remarks 47

1.9 Outer membrane beta barrels....48
1.9.1 Transport through the IM....51
1.9.2 Late translocation steps....51
1.9.3 Interactions with periplasmic chaperones....52
1.9.3.1 SurA....53
1.9.3.2 Skp....57
1.9.3.3 DegP....59
1.9.3.4 Other periplasmic chaperones....61
1.9.4 OM substrate/target recognition....63
1.9.5 The Bam Complex....63
1.9.5.1 BamA....65
1.9.5.2 BamB....69
1.9.5.3 BamC....70
1.9.5.4 BamD....72
1.9.5.5 BamE....73
1.9.6 Interactions of Bam complex members....75
1.9.7 Concluding remarks....78
1.10 Aims of this Study....82
Chapter 2 Materials and Methods....83
2.1 Working concentrations of antibiotics used in this study....84
2.2 Regaents used in this study....84
2.2.1 Media Recipes....84
2.2.2 Reagents Used in this Study....85
2.3 Bacterial strains....87
2.4 Cell manipulations....88
2.4.1 Bacterial growths....88
2.4.2 Generation of competent cells....88
2.4.3 Transformation by heat shock....88
2.5 DNA manipulations....89
2.5.1 Polymerase chain reactions (PCR)....89
2.5.2 Gel electrophoresis of DNA....89
2.5.3 Cloning of genes....90
2.5.4 Primers used in this study....91

2.5.4.1 Primers used for cloning *bamE* 91
2.5.4.2 Primers used for cloning *yraP* 91
2.5.4.3 Primers used to confirm the Δ*yraP* strain 91
2.5.4.4 Primers used for YraP mutagenesis 92
2.6 Protein experimentation 95
2.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 95
2.6.2 Protein expression and purification 95
2.6.3 Western Blotting 97
2.7 NMR Experiments 98
2.7.1 General overview 98
2.7.2 1D NMR experiments 98
2.7.3 HSQC 98
2.7.4 3D experiments 99
2.7.5 Hydrogen/Deuterium (H/D) exchange 100
2.7.6 Structure Calculations 100
2.7.7 NMR for lipid binding titrations 101
2.7.8 SurA binding by NMR 101
2.8 Functional experiments studying YraP 102
2.8.1 Screening of OM defects 102
2.8.2 Analytical Ultracentrifugation Sedimentation Velocity 102
2.8.3 Analytical gel filtration 103
2.8.4 Immunfluorescence Microscopy 103
2.8.5 Sucrose density centrifugation 104
2.8.6 Co-Immunoprecipitations 105
2.8.7 Periplasmic Co-IP 107
2.8.8 Identification of proteins 107
2.8.9 Generation of suppressor mutants 108
2.8.10 Mutagenesis of genes 109
2.8.11 LPS visualisation 110
2.8.12 Population difference identification of OM phospholipids 111
2.8.13 Relative quantification of outer membrane proteins 112
2.8.14 Screening of phenotypes using BioLog™ 113
Chapter 3 Functional and Structural Study of BamE 114

3.1 Introduction 115
3.2 Results 118
3.2.1 OM defects associated with Δ*bamE* strains 118
3.2.2 Cloning of *bamE* into pET20b 119
3.2.3 Complementing the defective phenotype associated with Δ*bamE* strains 120
3.2.4 Mutagenesis of the *bamE* gene 122
3.2.5 Screening of mutants 124
3.3 Discussion 130
Chapter 4 Bioinformatic analysis of YraP 136
4.1 Introduction 137
4.2 Results 139
4.2.1 YraP is a predicted OM lipoprotein 139
4.2.2 YraP is widely distributed in Gram-negative bacteria 141
4.2.3 The YraM-P module is conserved in Gram-negative bacteria 144
4.2.4 YraP contains 2 BON domains 147
4.2.5 Possible interactions with *yraP* 150
4.2.6 Secondary structure prediction of YraP 152
4.2.7 Tertiary structure predictions 154
4.3 Discussion 157
Chapter 5 Structure determination of YraP by NMR 160
5.1 Introduction 161
5.1.1 NMR as a structural biology tool 161
5.2 NMR Theory 162
5.2.1 Physical properties of the nucleus 162
5.2.2 Vector model of NMR 166
5.2.3 Relaxation 169
5.2.4 Chemical Shifts 171
5.2.5 Chemical shift perturbations 172
5.2.7 Sequential backbone assignment 174
5.2.8 HN(CA)CO and HNCO 175
5.2.9 HNCA and HN(CO)CA 176
5.2.10 HNCACB and CBCA(CO)NH 178
5.2.11 H(CCO)NH 179

5.2.12 (H)C(CO)NH 180
5.2.13 HCCH TOCSY 180
5.2.14 NOESY experiments 183
5.2.15 Hydrogen/Deuterium (H/D) exchange 186
5.2.16 Structure calculations with CYANA 186
5.3 Results 189
5.3.1 Expression and purification of YraP for NMR studies 189
5.3.2 HSQC analysis of ^{13}C, ^{15}N double labelled YraP 193
5.3.3 Spectra and assignment of the backbone experiments 194
5.3.4 Side chain assignments 200
5.3.5 Hydrogen - Deuterium exchanged HSQC 203
5.3.6 Nuclear Overhauser Effect Experiments 205
5.3.7 Secondary structure predictions using backbone assignment data 210
5.3.8 Determining Hydrogen Bonds 213
5.3.8.1 Hydrogen - Deuterium exchanged HSQC 213
5.3.8.2 Identification of β-strands using deuterated ^{15}N-edited NOESY data 215
5.3.9 Solving the structure of YraP using NMR data 217
5.3.9.1 Initial structure 218
5.3.9.2 Intermediate structures - Resolving anti-parallel β-strands 219
5.3.9.3 Intermediate structure - Resolving parallel β-strands 220
5.3.10 Final calculations 222
5.3.11 The final structure of YraP 223
5.4 DISCUSSION 232
Chapter 6 Biochemical and functional analysis of YraP 241
6.1 Introduction 242
6.2 Results 244
6.2.1 Confirmation of Δ*yraP* strain 244
6.2.2 Cloning of *yraP* 246
6.2.3 Δ*yraP* cells have an OM defect 247
6.2.4 YraP is OM localised 249
6.2.5 YraP is not surface exposed 251
6.2.6 YraP is monomeric in solution 252
6.2.7 YraP may not be a member of a protein complex 255

6.2.8 YraP does not interact with SurA 260
6.2.9 YraP binds specifically to the OM lipid PG 264
6.2.10 Mutagenesis of YraP 269
6.2.11 Generation of suppressor mutants 280
6.2.12 *S. enterica* Δ*yraP* cells have a similar phenotype to BW25113 Δ*yraP* cells 284
6.2.13 Δ*yraP* cells have an altered LPS profile 288
6.2.14 OM proteomic comparison of WT and Δ*yraP* cells 290
6.2.15 The OM of *Δyrap* cells have a different phospholipid composition to WT cells 292
6.3 Discussion 294
Chapter 7 Final Discussion 306
7.1 Final Discussion 307
7.2 BamE 307
7.3 YraP 309
REFERENCES 313

Appendix I BamE Publication by Knowles *et al.* 2011
Appendix II YraP assignment data
Appendix III YraP NOESY assignment data
Appendix IV H/D exchanged peaks
Appendix V Hydrogen bond restraints used

List of Abbreviations Used in This Study

E. coli – *Escherichia coli*
OM – Outer Membrane
IM – Inner Membrane
OMP – Outer Membrane β-barrel Protein
LPS – Lipopolysaccharide
Bam – β-barrel Assembly Machinery
Lpt – Lipopolysaccharide Transport
Lol – Localisation of Lipoproteins
ATP – Adenosine triphosphate
PMF – Proton Motive Force
SRP – Signal Recognition Particle
GTP – Guansoine Triphosphate
ABC – ATP Binding Cassette
PE – Phosphatidylethanolamine
PG – Phosphatidylglycerol
PC – Phosphaytidylcholine
EDTA – Ethylenediaminetetraacetic acid
MLA – Maintenance of Lipid Asymmetry
DNA – Deoxyribonucleic Acid
NMR – Nuclear Magnetic Resonance
SDS – Sodium dodecyl sulphate
PBP – Penicillin-Binding Protein
PPIase – Peptidyl-Prolyl-Isomerase
UPEC – Uropathogenic *E. coli*
POTRA – Polypeptide Translocation Associated
RMSD – Root Mean Square Deviation
SAXS – Small Angle X-ray Scattering
PELDOR – Pulsed Electron Double Resonance
SAM – Sorting and Assembly Machinery
TPR – Tetratricopeptide Repeat
Amp – Ampicillin
Kan – Kanamycin
Vanc – Vancomycin
Tet – Tetracycline
SDW – Sterile Distilled Water
LB – Luria Bertani
PCR – Polymerase Chain Reaction
SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
MW – Molecular Weight
IPTG – Isopropyl β-D-1-thiogalactopyranoside

PBS – Phosphate Buffered Saline
D_2O – Deuterium Oxide
H/D – Hydrogen/Deuterium
AUC – Analytical Ultracentrifugation
DTT – Dithiothreitol
DSP – Dithiobis(succinimidyl) propionate
RT – Room Temperature
ZW 3-14 – Zwittergent 3-14
DDM – n-Dodecyl β-D-Maltopyranoside
MALDI-TOF – Matrix Assisted Laser Desorption Ionisation – Time of Flight
ESI-qTOF – Electrospray Ionisation - Quadrupole Time of Flight
FT-ICR – Fourier Transform – Ion Cyclotron Resonance
PCA – Principal Component Analysis
TFA – Trifluoroacetic acid
BON Domain – Bacterial and OsmY Nodulation Domain
HSQC – Heteronuclear Spin Quantum Coherence
NOESY – Nuclear Overhauser Effect Spectroscopy
TOCSY – Total Correlation Spectroscopy
CSI – Chemical Shift Index
RDC – Residual Dipolar Coupling
WT – Wildtype
KO – Knockout
Co-IP – Co-Immunoprecipitation
XL – Cross-Linked

Chapter 1

Introduction

1.1 Introduction

Bacteria can be classified into two major groups, Gram-positive or Gram-negative, based on the physical and chemical properties of their membranes. Gram-positive bacteria are surrounded by a single lipid bilayer whereas Gram-negative contain two distinct membranes, the inner membrane (IM), which surrounds the cytoplasm and the outer membrane (OM), which contains the aqueous periplasm and a layer of peptidoglycan (Sperandeo *et al*., 2011) and contacts the extracellular environment (**Figure1.1**). The presence of the OM is a defining feature of Gram-negative bacteria.

The structures of the IM and OM are remarkably different. The IM is composed of a phospholipid bilayer whereas the OM is asymmetric, containing phospholipids in its inner leaflet and lipopolysaccharide (LPS) in its outer leaflet. The protein compositions between the two membranes are also quite different. Both membranes contain lipoproteins (proteins that are attached to the membrane by N-terminal N-acyl-diacylglycerylcysteine), but the IM mainly harbours α-helical proteins whilst the OM contain β-barrels (OMPs).

The OM is an essential organelle. The OM mediates contact with the extracellular environment as well as crucially acting as a barrier and preventing the entry of bactericidal compounds, which is critically important during events such as invasion and persistence. The barrier function of the OM is very robust, with only small hydrophobic molecules of < 600 Da being permeable. This can be attributed to the tight interactions between LPS molecules.

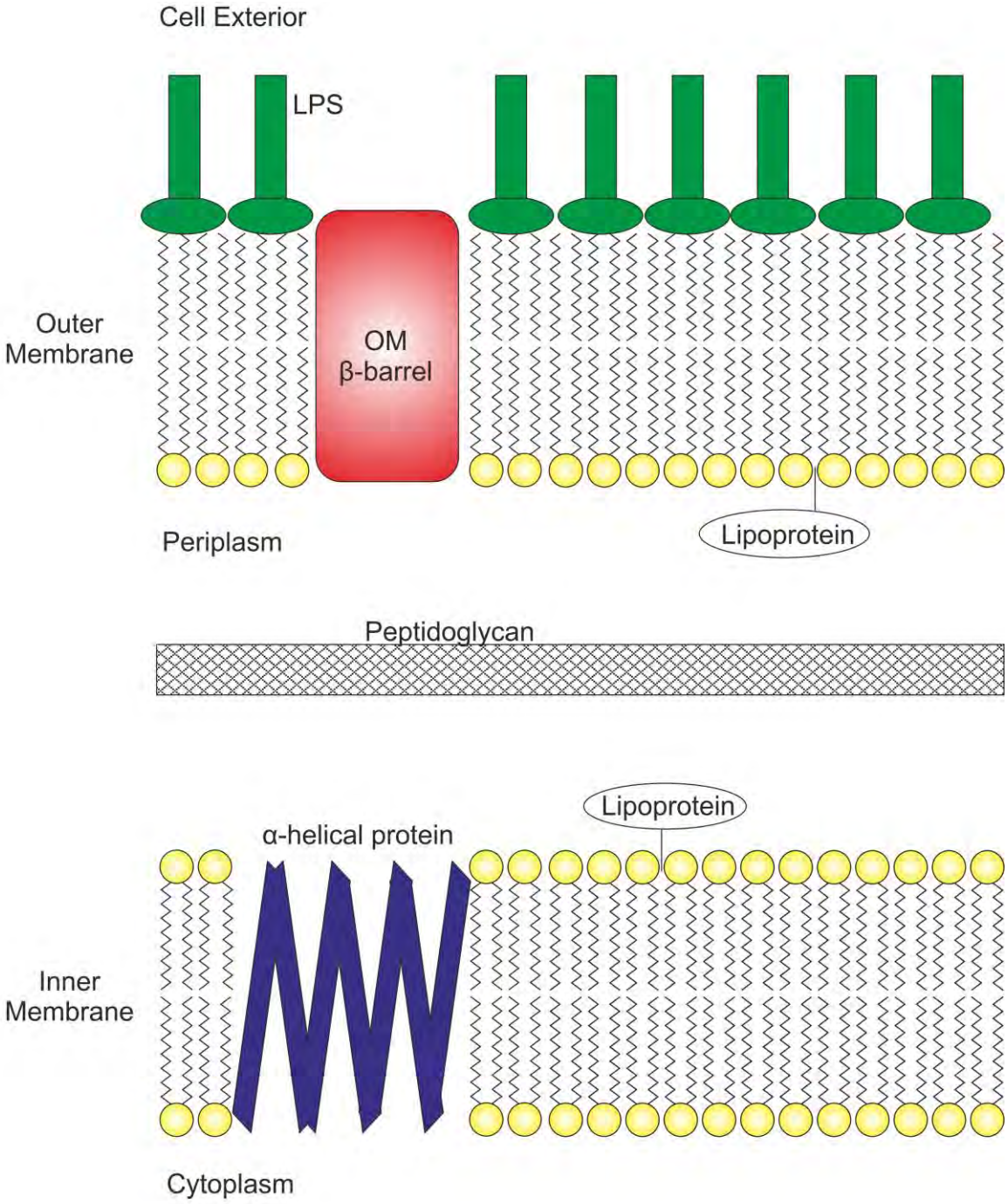


Figure 1.1 Schematic of the membrane topology in Gram negative bacteria. α-helical proteins (blue) reside in the IM, whereas β-barrel proteins (red) are present exclusively in the OM. LPS (green) is present on the outer leaflet of the OM.

The barrier function is dependent on the correct composition of phospholipids, proteins and LPS. Deviations to this ratio can lead to a compromised OM barrier. In order to understand how this barrier is maintained the pathways that are responsible for assembly of the individual components must be studied. It is now known that LPS assembly is dependent on the Lpt (lipopolysaccharide transport) pathway, lipoproteins with the Lol (localisation of lipoproteins) pathway and OMPs are assembled by the Bam (β-barrel assembly machinery) complex. The pathway that is involved in the assembly of phospholipids is still unknown (**Figure 1.2**).

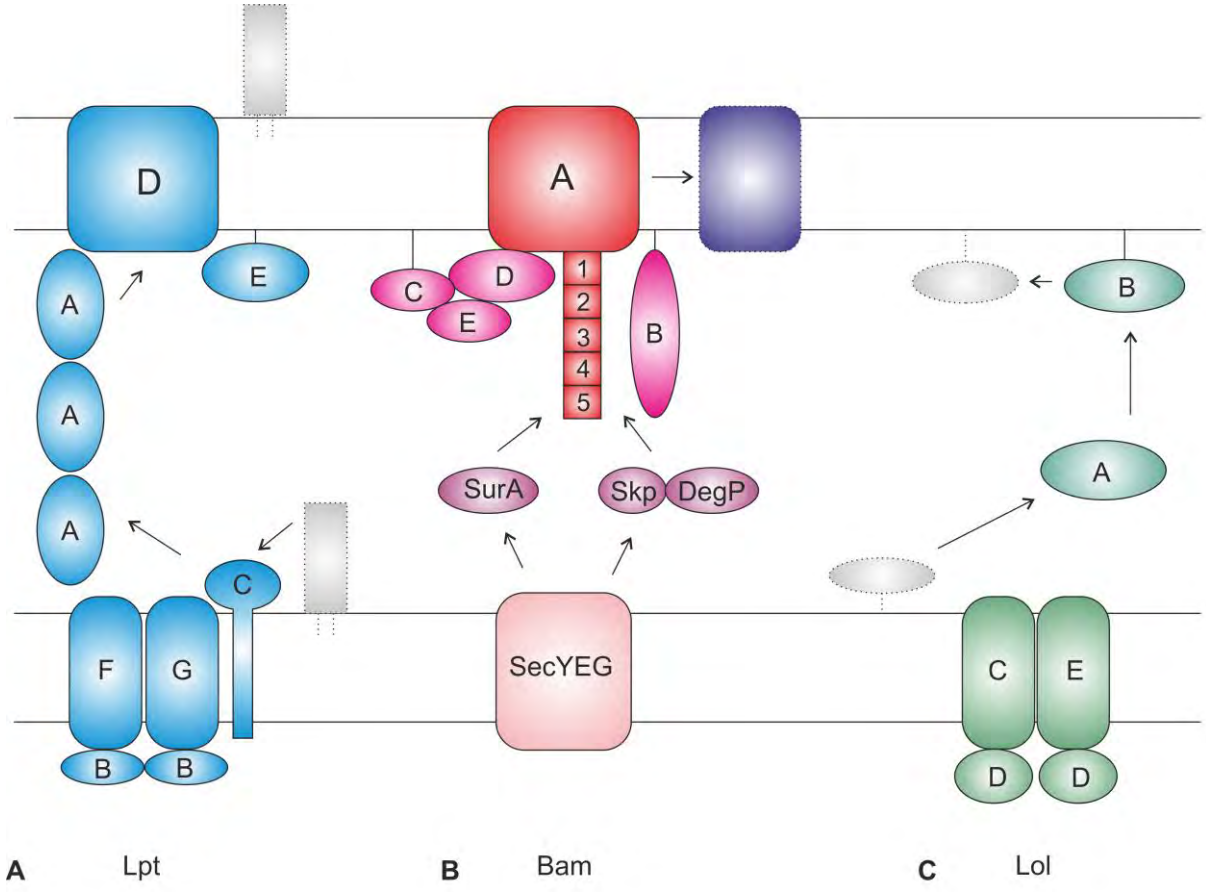


Figure 1.2 Summary of the pathways involved in OM biogenesis in *E. coli*. **A**The Lpt pathway (blue) is responsible for assembly of LPS molecules in the OM. The LptBCFG complex releases LPS from the IM in an ATP-dependent manner, which is passed onto the chaperone LptA and to the OM β-barrel and lipoprotein LptD and LptE, which insert LPS into the OM by an unknown mechanism.

B The Bam complex pathway (pink) is responsible for the assembly of β-barrels. Proteins are translocated through the IM by the Sec machinery and interact with one of two chaperone pathways, SurA or Skp and DegP. Precurssor OMPs then interact with either the accessory lipoproteins of BamB, C, D or E or with the POTRA domains of BamA (1-5) and are then inserted into the OM by BamA.

C The Lol pathway (green) assembles lipoproteins. Lipoproteins are released from the IM using ATP by the LolCDE complex and the interact with the periplasmic chaperone LolA. LolA passes the lipoprotein to LolB, which inserts the lipoprotein into the OM.

1.2 Topological problem of OM biogenesis

None of the constituent parts of the OM are synthesised *in situ*. Instead, they are all synthesised in the cytoplasm or the inner leaflet of the IM and have to overcome a series of obstacles to get to their final location of the OM. OM components must first be translocated across the IM, which is an energetically unfavourable reaction that must be coupled to an exergonic reaction (du Plessis *et al*., 2010). Additionally, OM components may contain hydrophobic motifs that need to be shielded from the aqueous periplasm by dedicated factors, which would otherwise cause aggregation or hydrophobic collapse in the periplasm. Furthermore, transport to the OM has to be unidirectional as there is the possibility of these components being assembled in the IM, which can have disastrous effects to the cell, such as the dissipation of proton gradients present in the IM (Carlson and Silhavy, 1993, Guilvout *et al*., 2006). Finally, assembly seems to be a passive process as the cell envelope lacks obvious energy sources, such as ATP or ion gradients (Hagan *et al*., 2011). All of these criteria have to be fulfilled to maintain a robust permeability barrier.

Gram-negative bacteria have specific pathways that are involved in the assembly of the individual components to the OM. These pathways have some common features. Firstly, OM components are synthesised in the cytosol or cytoplasmic face of the IM. They are then translocated through the IM by a dedicated system in the IM, using an energy source (such as ATP). Once in the periplasm, chaperones bind to the substrate and shield the hydrophobic patches and aid assembly. Finally, these components are assembled in the OM by a dedicated OM protein (or protein complex).

Each transport system accomplishes this with the concerted efforts of protein factors involved in each pathway. This introductory section will describe the known pathways involved in OM biogenesis. It should be noted that these pathways should not be thought of as entirely separate aspects of OM biogenesis, as there is genetic interplay between them, which ensures that the correct ratio of OM components is maintained (Ruiz *et al*., 2005, Tam and Missiakas, 2005).

1.3 Transport through the IM

All of the protein species that are transported to the OM, are synthesised in the cytoplasm and have to cross the IM. There are several translocation pathways available, transport through the Sec translocon will be discussed briefly.

The evolutionary conserved Sec translocase is responsible for the trafficking of the majority of protein species through or into the IM. The SEC translocase consists of a heterotrimeric complex of SecYEG to form a protein-conducting membrane channel. Additional accessory factors, such as YidC and SecDF have been shown to improve translocation of certain membrane proteins and mediate proton motive force (PMF) dependent translocation (du Plessis *et al*., 2006, Kol *et al*., 2009, van Bloois *et al*., 2006, Nouwen *et al*., 2001). The Sec translocase is able to accept a range of substrates, from very hydrophilic to very hydrophobic proteins.

There are two modes of translocation through the Sec machinery, co-translational and post-translational. The co-translational pathway is mainly used for hydrophobic membrane proteins and is mediated by the Signal Recognition Particle (SRP), Ffh in *E. coli*, which is a conserved structure ribonucleoprotein associated with a 4.5S RNA (Poritz *et al*., 1990, Ribes *et al*., 1990, Pool, 2009, Valent *et al*., 1995). The SRP is able to bind to ribosome-nascent chains and guide them to the SRP receptor, FtsY. GTP hydrolysis mediates transfer to the translocon (Valent *et al*., 1998, de Leeuw *et al*., 2000).

Post-translational translocation across the Sec machinery is generally used for proteins that are secreted across the membrane (Ulbrandt *et al*., 1997). The newly synthesised preprotein is maintained in an unfolded state by the cytosolic chaperone SecB (Bechtluft *et al*., 2007, Fekkes and Driessen, 1999). SecB-preprotein complex is targeted to the translocon, where SecB binds to SecA, an ATPase motor protein. Translocon is mediated by the expense of ATP (Schiebel *et al*., 1991).

1.4 Peptidoglycan

The peptidoglycan cell wall is an elastic mesh-like network of rigid cross-linked glycan strands. The petptidoglycan is an essential organelle used to maintain cell shape, resist turgor pressure and promote cell growth and separation. This critical structural element is located in the cell envelope, in the periplasmic space between the IM and OM of Gram-negative bacteria. Due it its location, this would suggest

that it presents a physical barrier for OM components en route to the OM. However, no available evidence supports this. Additionally, the peptidoglycan layer has been shown to be permeable to proteins up to 100 kDa (Vazquez-Laslop *et al*., 2001). It seems that the peptidoglycan contributes very little as physical barrier. Because of this, the peptidoglycan will be largely ignored for the discussion of OM biogenesis.

1.5 Regulation systems involved with OM homeostasis

The OM can be exposed to many external stimuli which can alter its properties. In order to maintain the OM, there are three signal transduction pathways that are induced by envelope perturbations. These are the extracytoplasmic alternative sigma factor, σ^E (Mecsas *et al*., 1993) and the two component systems, CpxAR (Raivio *et al*., 1999) and BaeSR (Raffa and Raivio, 2002). These pathways are involved in upregulating genes that are involved in OM biogenesis and some have overlapping genes (Danese *et al*., 1995, Dartigalongue *et al*., 2001). These pathways will be discussed briefly.

The σ^E response is activated by events that lead to alterations in the OM from external conditions, such as the presence of heat or ethanol (Mecsas *et al*., 1993). σ^E is an alternative sigma factor that is sequestered by the membrane-bound anti-sigma factor RseA, under unstressed conditions. However, under stressed conditions, cleavage of RseA occurs by DegS, which releases σ^E (Alba *et al*., 2002) (Kanehara *et al*., 2002) (Walsh *et al*., 2003) (Flynn *et al*., 2004). This cleavage can also be fine tuned by the periplasmic regulatory protein, RseB (Grigorova *et al*., 2004). Degradation of RseA, leads to the release of σ^E, which binds to core RNA

polymerase leading to the transcription of σ^E-regulated genes (Erickson and Gross, 1989, Danese and Silhavy, 1997, Dartigalongue *et al*., 2001).

The Cpx envelope stress response is controlled by the two-component system of CpxA (a membrane-localised histidine kinase) and CpxR (the response regulator). The Cpx pathway is activated by several conditions that lead to protein misfolding, for example alkaline pH and alterations in membrane conditions amongst others (Snyder *et al*., 1995, Mileykovskaya and Dowhan, 1997, Danese and Silhavy, 1998). The current model of activation suggests that the periplasmic inhibitor CpxP is degraded by proteolysis, which allows CpxA to phosphorylate CpxR. CpxR-P can then act as a transcription factor and activate OM homeostasis genes (Pogliano *et al*., 1997, Raivio and Silhavy, 1997, Dartigalongue and Raina, 1998).

The BaeSR is also a two-component system consisting of the membrane-localised histidine kinase BaeS and the response regulator BaeR (Raffa and Raivio, 2002). This pathway is sensitive to antimicrobial substances such as indole and is involved in the resistance to several classes of antimicrobial compounds, such as bile salts and novobiocin (Baranova and Nikaido, 2002, Nagakubo *et al*., 2002). Phosphorylation of BaeR by BaeS can lead to the transcription of several genes, such as drug efflux pumps (Baranova and Nikaido, 2002, Bury-Mone *et al*., 2009).

1.6 Lipopolysaccharide assembly

Lipopolysaccharide (LPS) is a glycolipid that resides exclusively in the outer leaflet of the OM (Kamio and Nikaido, 1976). LPS is one of the major components that contributes greatly to the barrier function of the OM. LPS monomers exhibit strong lateral interactions with each other which reduces the fluidity of the OM. This leads to the production of a rigid and gel-like structure in the OM that can act as a barrier to a variety of hydrophobic and hydrophilic molecules (e.g. antibiotics) (Nikaido, 2003).

Apart from its barrier function, LPS has been shown to be highly antigenic and has been found to be the main causative agent of septic shock (Drake *et al*., 1993, Parillo, 1993, Esmon, 2000, van Deuren *et al*., 2000, Bernard *et al*., 2001). It is detected by TLR4 (Toll-like receptor 4) present on macrophages and endothelial animal cells (Aderem and Ulevitch, 2000, Medzhitov and Janeway, 2000, Poltorak *et al*., 1998, Hoshino *et al*., 1999) and activation leads to the biosynthesis of mediators of inflammation, such as IL1-β and TNF-α as well as the production of molecules required for the adaptive immune response (Medzhitov and Janeway, 2000, Beutler and Cerami, 1988, Dinarello, 1991).

LPS can be subdivided into three distinct components, lipid A (which secures the LPS to the OM by a hexa-acylated sugar moiety), core oligosaccharide and O-antigen (which is absent in derivatives of *E. coli* K-12) (**Figure 1.3**).

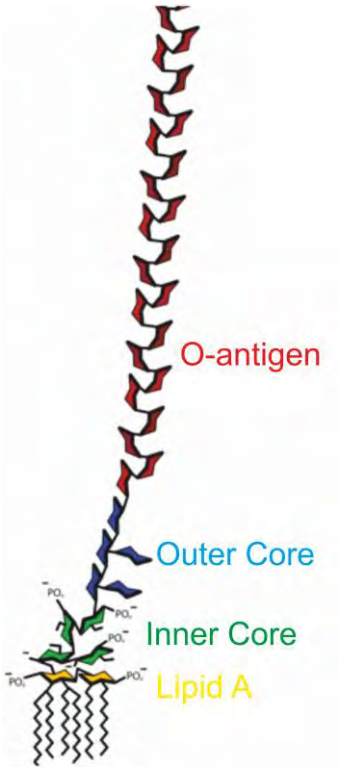


Figure 1.3 Schematic of the main components of LPS. Lipid A (yellow), inner core (green) and outer core (blue) comprise rough LPS. The addition of O-antigen (red) produces smooth LPS, which is absent in derivatives of *E. coli* K-12).
Synthesis of lipid A and core oligosaccharide takes place on the cytoplasmic leaflet of the IM, whereas O-antigen is ligated on the periplasmic face of the IM, after the nascent glycolipid has been flipped to the periplasmic leaflet by an ABC transporter (Raetz and Whitfield, 2002). LPS is transferred to the OM by the Lpt pathway.

1.6.1.1 Synthesis of lipid A

Synthesis of LPS originates on the cytoplasmic leaflet of the IM. Each of the component parts are synthesised separately and then combined. The biogenesis of lipid A is well characterised in *E. coli* and requires many enzymatic reactions in a highly conserved pathway (Raetz, 1990, Raetz, 1993, Wyckoff *et al*., 1998) (**Figure 1.4**). In brief, LpxA (UDP-GlcNAc acyltransferase) acylates the sugar nucleotide UDP-GlcNAc (Anderson and Raetz, 1987). This product is then deacetylated by the zinc metalloprotease LpxC (Anderson *et al*., 1988, Young *et al*., 1995, Sorensen *et al*., 1996, Jackman *et al*., 1999, Jackman *et al*., 2001). A β-hydroxymyristate moiety is incorporated by LpxD to generate UDP-2,3-diacylglucosamine (Kelly *et al*., 1993). This product is then cleaved at its pyrophosphate bond by LpxH to form 2,3-diacylglucosamine-1-phosphate (lipid X). The kinase LpxK phosphorylates the 4' position of the disaccharide to form lipid IV_A (Ray and Raetz, 1987). WaaA (also known as KdtA) transfers two 3-Deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) residues to lipid IV_A and finally lauroyl and myristoyl residues are added by LpxL and LpxM respectively, to produce lipid A (Clementz *et al*., 1996, Clementz *et al*., 1997, Mohan and Raetz, 1994, Karow and Georgopoulos, 1991, Karow *et al*., 1992).

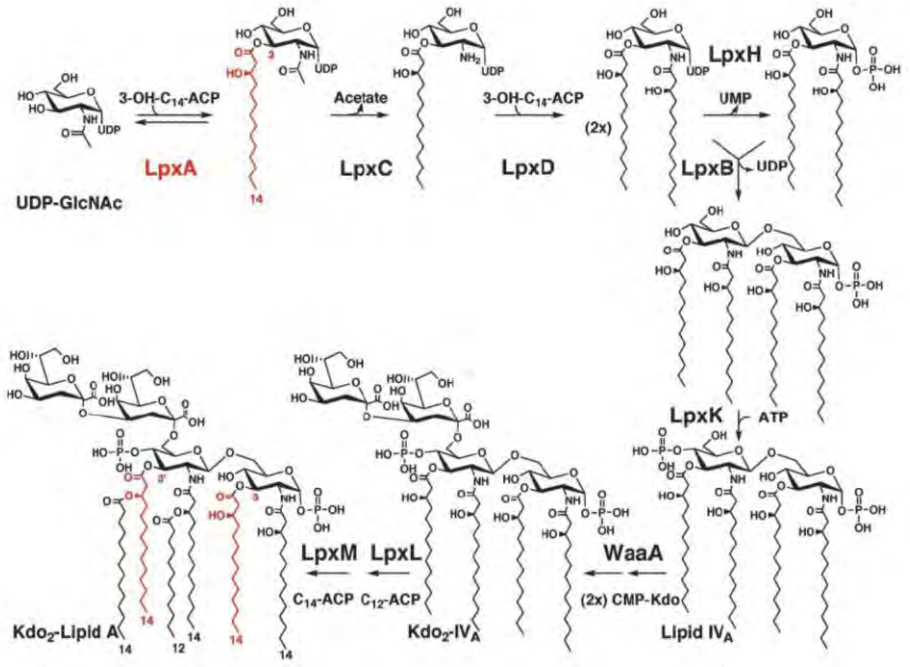


Figure 1.4 The synthesis of Lipid A. Schematic of the enzymatic reactions that are involved in the synthesis of Lipid A (taken from Raetz and Whitfield 2002).

1.6.1.2 Synthesis of core oligosaccharide

As with lipid A, biosynthesis of core oligosaccharide is well characterised in *E. coli*. It should be noted that a variety of organisms with differing lifestyles can produce remarkably different core oligosaccharides, demonstrating the complexity of this molecule (Muller-Loennies *et al*., 1999).

Core oligosaccharide can be divided into two distinctive regions; inner core (lipid A proximal) and outer core. The inner core tends to be well conserved, containing residues of Kdo and L-*glycero*-D-*manno*heptose (Kadrmas *et al*., 1996). This conservation possibly indicates the importance of the core in OM integrity. The outer core provides a site for O-antigen attachment and shows more structural diversity.

This is expected for a region with more exposure to selective pressures of host responses and environmental stresses, although the structural variation between core oligosaccharides within a species or genus is still limited (Raetz and Whitfield, 2002).

The synthesis of core oligosaccharide follows a multiple enzyme-catalysed pathway, similar to lipid A biogenesis (**Figure 1.5**). However, this pathway is much more complex and can vary between strains and thus will not be discussed in detail (Muller-Loennies *et al*., 1999). The biogenesis of inner core oligosaccharides requires enzymes WaaP and WaaY (both LPS kinases), WaaQ (a heptose transferase) and WaaZ (a Kdo transferase) (Heinrichs *et al*., 1998b, Yethon *et al*., 1998, Yethon *et al*., 2000, Walsh *et al*., 2000, White *et al*., 1999, Brabetz *et al*., 2000). With outer core oligosaccharides, the Waa family of glycosyltransferases are used, with WaaI being involved in the ligation of lipid A to the core oligosaccharide to form rough LPS (Creeger *et al*., 1979, Campbell *et al*., 1997, Heinrichs *et al*., 1998a, Heinrichs *et al*., 1998b, Saxena *et al*., 1995).

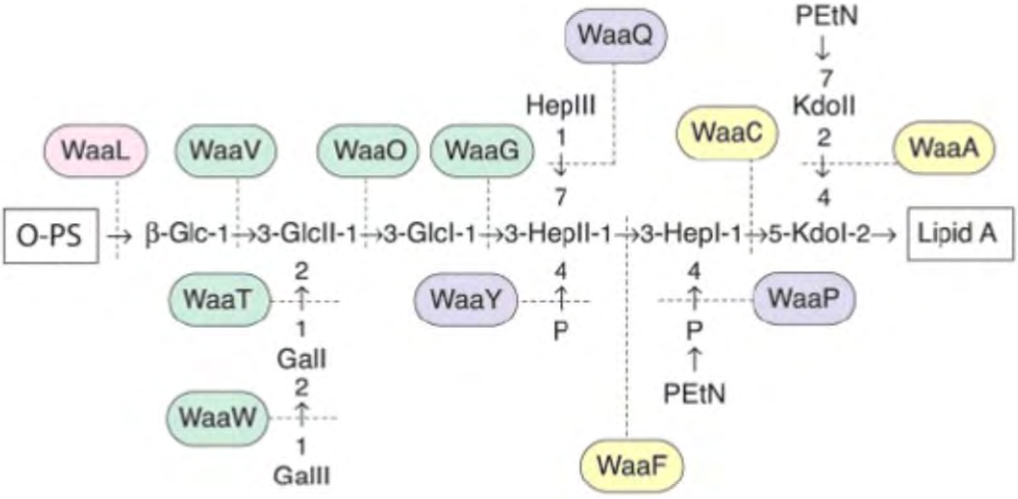


Figure 1.5 The biosynthesis of core oligosaccharide. Schematic of the enzymatic process of synthesis of core oligosaccharide (taken from Raetz and Whitfield 2002). Yellow boxes indicate glycosyltransferases that form the inner core backbone and enzymes that modify the structure are indicated in blue. Outer core glycosyltransferases are indicated in green and the ligase enzyme WaaL is highlighted in pink.

1.6.1.3 MsbA

Further maturation of LPS takes place on the periplasmic leaflet of the IM. For this to occur, rough LPS must be flipped across the IM. This transportation step is performed by the essential ABC transporter MsbA in an ATP dependent manner (Zhou *et al*., 1998).

The crystal structure of MsbA from *E. coli*. has been solved to 5.3 Å resolution(Ward *et al*., 2007). From the structure, MsbA is a homodimer that contains 6 trans-membrane helices and a large cavity between the two subunits that could accommodate lipid A and/or phospholipid molecules. How MsbA mediates lipid flip-flop is unclear however it is thought that the nucleotide binding domains may come

together upon binding to ATP, following lipid entry into the cavity, which could force the lipid to move to the periplasmic side of the IM.

1.6.1.4 O-antigen

O-antigen is a repetitive glycan polymer that is attached to the outer core oligosaccharide. O-antigen is structurally diverse, with more than 60 monosaccharides and 30 different non-carbohydrate components having been identified (Knirel' Iu and Kochetkov, 1994). O-antigen displays remarkable diversity, as different structures comprise varying numbers of monosaccharides, can either be linear or branched and they can be made up of homo or hetero-polymers. This diversity defines the O-antigen serological specificity of an organism.

O-antigen is synthesised in the cytoplasm by several genes that are present in the *rfb* locus, which encodes enzymes such as glycosyltranserases and synthesisers of sugar nucleotide precursors, that are either soluble or peripheral membrane proteins (Popoff and Le Minor, 1985, Keenleyside *et al*., 1994). Three pathways have been identified that can transfer O-antigen to the periplasmic leaflet, Wzy-dependent, ABC-transport dependent and synthase dependent (Liu *et al*., 1993, Liu *et al*., 1996, Batchelor *et al*., 1991, Morona *et al*., 2000, Rick *et al*., 1994, Kido *et al*., 1995). O-antigen is then ligated to the lipid A-Core oligosaccharide complex at the periplasmic face of the IM by the ligase WaaL, to form smooth LPS (McGrath and Osborn, 1991, Daniels and Morona, 1999).

1.6.1.5 LPS transport

Mature LPS molecules are attached to the periplasmic face of the IM and have to be transported to the OM. Recent advances have led to the discovery of proteins involved in this process and are known as the Lpt (LPS transport) pathway. This pathway contains proteins LptA through G and all are essential in *E. coli*, gene depletions lead to LPS accumulation (Sperandeo *et al*., 2006, Ruiz *et al*., 2008, Sperandeo *et al*., 2007, Sperandeo *et al*., 2008, Wu *et al*., 2006). LptBCFG form a subcomplex that is present in the IM that acts as an ABC transporter (Narita and Tokuda, 2009). LptA is a periplasmic chaperone (Ma *et al*., 2008) and LptD and LptE are a β-barrel and lipoprotein respectively in the OM (Braun and Silhavy, 2002).

1.6.1.6 Extraction from the Inner Membrane - LptBCFG complex

It is thought that the LptBCFG complex releases LPS from the IM using ATP. LptBCFG has been found to be in a complex with a 2:1:1:1 ratio (Narita and Tokuda, 2009). LptB is 26.7 kDa protein, with an ABC signature (Sperandeo *et al*., 2007). LptF and G possess typical transmembrane topologies of ABC transporters, with 6 transmembrane helices and cytoplasmic N and C-termini (Linton and Higgins, 2007, Daley *et al*., 2005). It is thought that LptBFG, used ATP hydrolysis to transfer LPS to LptC (Sperandeo *et al*., 2009). LptC is a bitopic 21.6 kDa protein with a single transmembrane domain (Sperandeo *et al*., 2008). LptC was shown not to be essential for the ATPase activity of the complex and that the ATPase activity of the complex is unaffected by LPS, phospholipids and lipidA, unlike MsbA (Narita and

Tokuda, 2009). The structure of the periplasmic domain of LptC was solved and it shows a similar domain to LptA (discussed below).

1.6.1.7 Translocation across the periplasm - LptA

LptA is a18.6 kDa periplasmic protein that is thought to act as a chaperone for the transit of LPS to the OM (Sperandeo *et al*., 2007, Tefsen *et al*., 2005). The crystal structure of LptA reveals a novel fold, consisting of 16 consecutive anti-parallel β-strands to form a slightly twisted "beta-jellyroll" (**Figure 1.6**) and has the ability to bind LPS via LipidA (Suits *et al*., 2008).

Interaction studies have been performed between LptA and LptC. It has been shown that LptC can bind to LPS *in vitro* (Tran *et al*., 2010) and that LptA can displace LPS from LptC. Additionally, LptC cannot displace LPS from LptA, indicating that transfer is unidirectional and probably driven by the competing affinities for LPS between LptA and LptC (Tran *et al*., 2010). This suggests a role of LptC being used as a mediator to transfer LPS to LptA.

Two hypotheses have been suggested for how LptA transfers LPS to the OM. The "chaperone hypothesis" suggests that LptA acts as a chaperone, in a similar manner to SurA in assembly of OMPs and LolA in lipoprotein biogenesis (Sperandeo *et al*., 2007). The second hypothesis is the "periplasmic bridge" model, in which multiple LptA subunits form a protein bridge that connects both membranes and LPS is passed between the subunits until it reaches the OM (Sperandeo *et al*., 2007, Tefsen *et al*., 2005). Evidence to support this model comes from a crystal structure of LptA showing long fibre-like structures are formed from multiple LptA

molecules arranged in a head-to-tail manner in the presence of LPS (Suits *et al*., 2008, Merten *et al*., 2012).

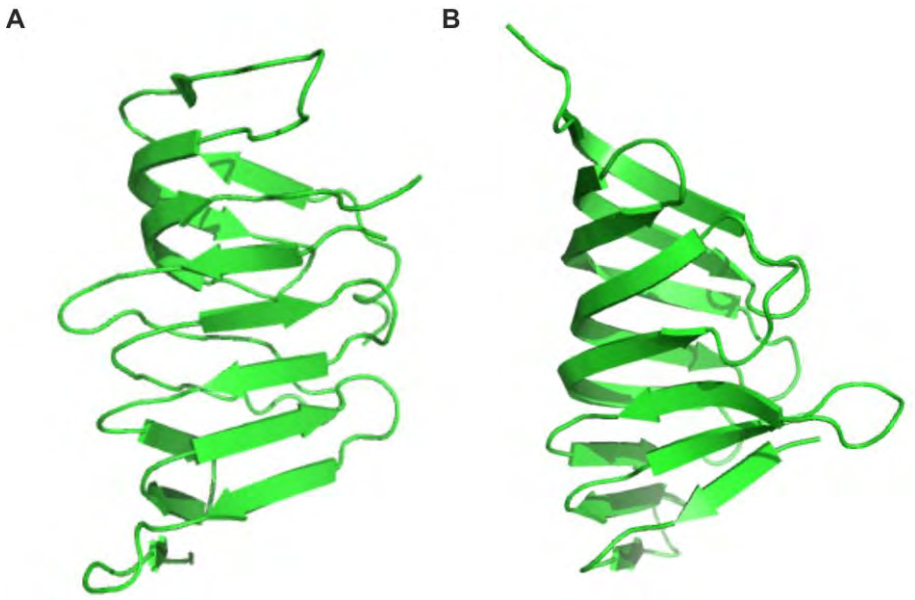


Figure 1.6 Structures of LptA and LptC. Crystal structures of LptA (PDB:2RIA) (**A**) and LptC (PDB:3MY2) (**B**) exhibit a very similar fold of a slightly twisted "β-jellyroll", comprised of 16 anti-parallel β-strands.

1.6.1.8 Insertion into the Outer Membrane - LptD and LptE

LptD is an essential 87 kDa OMP that is expressed at low levels in *E. coli* (Abe *et al*., 2003, Braun and Silhavy, 2002). LptD is associated with another low expression protein, LptE (Takase *et al*., 1987, Wu *et al*., 2006). Over-expression of LptD was only possible when LptE is expressed simultaneously, reflecting the importance of this complex for function (Chng *et al*., 2010b).

LptE has been shown to interact directly with LPS, possibly suggesting a role in substrate recognition (Chng *et al*., 2010a). However, evidence also suggests that LptE is involved in trafficking LPS to LptD (Bos and Tommassen, 2011). Interaction

studies have demonstrated that several residues of LptE interact directly with LptD *in vivo* and that a "plug-and-barrel" architecture is observed, with LptE residing in the LptD barrel (Freinkman *et al*., 2011). The interaction between LptD and LptE has also been shown to occur while LptD is assembled into the OM by the Bam complex, suggesting LptE may have a role in the correct folding of LptD (Chimalakonda *et al*., 2011).

Two models have been proposed of how LPS is inserted into the OM by LptD and LptE. In one model, "the two-step model", LPS is inserted into the inner leaflet of the OM first. The LPS is then flipped to the outer leaflet, performed by either LptD or LptE (Sperandeo *et al*., 2009). Another model, the "plug and barrel model", suggests that LPS is released laterally into the OM by LptD, possibly gated by LptE (Chng *et al*., 2010b, Freinkman *et al*., 2011). In this model, the fatty acyl chains of the LPS substrate are recognised by LptE. LptE guides the LPS into the lumen of LptD. This interaction triggers conformational changes in LptD that leads to insertion of LPS laterally into the outer leaflet of the OM.

1.6.1.9 Concluding remarks

LPS is synthesised in the cytoplasm and IM and transported to the outer leaflet of the OM. Characterisation of the synthesis of LPS has been well documented, although there is some complexity with different enzymes used for the production of outer-core and O-antigen. This ultimately leads to a high degree of variability between organisms that produce LPS.

The molecular mechanisms of transport to the OM are not yet fully understood, but great progress has been made within the last decade. It seems very likely that the LptBCFG complex can act as an ABC transporter to release LPS from the IM, with LptC acting as a holder of LPS which is then transferred to the periplasmic chaperone LptA. Whether or not LptA acts as a chaperone or forms a protein bridge is still not known. Also, the molecular details of how LptD and LptE insert LPS into the OM still needs to be investigated.

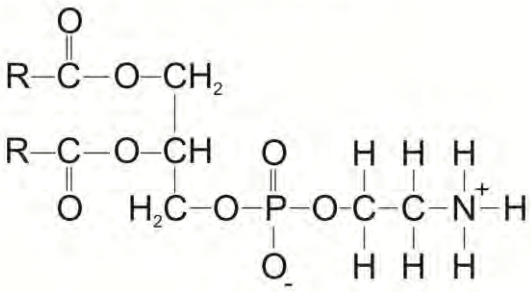
1.7 Phospholipids

The majority species of phospholipid in the OM of *E. coli* is the zwitterionic phosphatidylethanolamine (PE), which comprises roughly 75-80% of the total phospholipid content in *E. coli*. The second major species is phosphatidylglycerol (PG), which is negatively charged and comprises 15 -20%. Finally, cardiolipin (a PG condensate) is present in small amounts. The presence of these phospholipids with saturated acyl chains, slightly increases membrane rigidity and distinguishes the OM from the IM (Lugtenberg and Peters, 1976, Huijbregts *et al*., 2000).

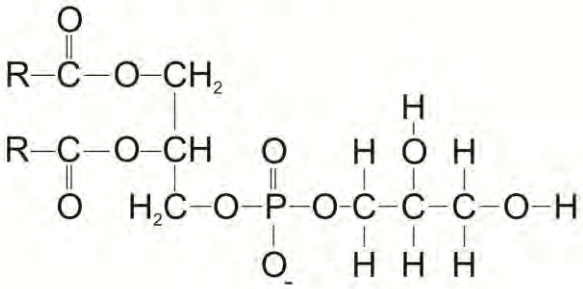
Nothing is known about the mechanisms of how phospholipids are transported from their site of synthesis at the IM (Osborn *et al*., 1974), to their site of function in the OM. Flip-flop mechanisms have been observed, whereby lipids are transported from one leaflet to another in a bi-layer. For example in the IM, flip-flop from the cytoplasmic leaflet to the periplasmic leaflet can be mediated passively by α-helical transmembrane proteins (e.g. the potassium channel) KcsA or actively by MsbA (Doerrler *et al*., 2001, Kol *et al*., 2003, Kol *et al*., 2004). However transport from one membrane to another remains elusive. Periplasmic lipid vesicles have not been observed visually and it seems unlikely that they can be used as the peptidoglycan layer may restrict their passage through the periplasm (Huijbregts *et al*., 2000). Zones of adhesion between the IM and OM have been postulated, "Bayer bridges", although the existence of these remains controversial (Bayer, 1968, Bayer, 1991, Kellenberger, 1990). It has been demonstrated that phospholipid shuttling does not

require protein synthesis or ATP, but is heavily dependent on the proton motif force (PMF) (Donohue-Rolfe and Schaechter, 1980). It should be noted however, that unlike protein or LPS transport, the transportation of phospholipids is bi-directional (Jones and Osborn, 1977), further adding to the complexity of an unknown system.

Phosphatidylethanolamine (PE)



Phosphatidylglycerol (PG)



Cardiolipin

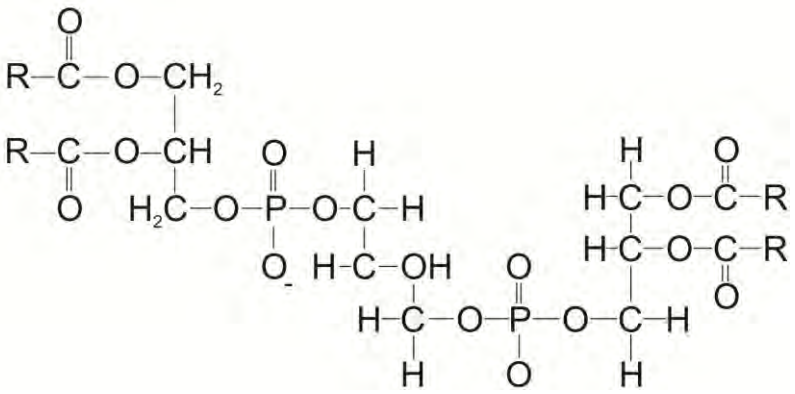


Figure 1.7 Structures of the three major phospholipid species found in the OM of *E. coli*. The head groups of phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin are shown, with R representing the acyl chains.

1.7.1.1 Maintaining the asymmetry of the OM

The OM is organised asymmetrically, with LPS in the outer leaflet and phospholipids in the inner leaflet (Kamio and Nikaido, 1976). The LPS molecules exhibit strong lateral interactions between themselves which contribute to the barrier function of the OM (Nikaido, 2003). The effect of OM disruptions can perturb these LPS interactions, for example the addition of EDTA, displaces the divalent cations needed to reduce the repulsive charges between the LPS molecules (Nikaido, 2003). As a result of this, LPS is shed (Leive, 1965) and phospholipids are forced to migrate from the inner leaflet of the OM into the disrupted areas of the outer leaflet. The effect of these phospholipid patches at the OM compromises the OM barrier function (Nikaido, 2005). It is essential for Gram-negative bacteria to remove these patches and maintain the asymmetry of the OM.

There are three systems that have been identified for the removal of phospholipids from the OM. Two of these are OMPs that have enzymatic activity, PldA and PagP (Bishop, 2008), and the third is the retrograde (OM to IM) phospholipid transport system, Mla (Maintenance of Lipid Asymmetry) (Malinverni and Silhavy, 2009). As these systems contribute to OM maintenance, they will be discussed briefly.

1.7.1.2 PldA

PldA (or OMPLA) is an OMP with Ca^{2+}-dependent phospholipase activity (Scandella and Kornberg, 1971, Bell *et al*., 1971). Usually it is constitutively expressed indicating a possible role as a house-keeping gene (Dekker, 2000). Mutants of PldA have been identified that show increased activity, resulting in the observation of OM

defects (Audet *et al*., 1974, Michel and Starka, 1979). The WT protein is triggered by a variety of conditions or compounds that compromise the OM. Such examples include colicins, phage transfection of DNA, phage-induced lysis, heat shock, spheroplast formation, EDTA-treatment and the action of antimicrobial peptides (Pugsley and Schwartz, 1984, Cascales *et al*., 2007, Taketo, 1974, Cronan and Wulff, 1969, Patriarca *et al*., 1972, de Geus *et al*., 1983, Hardaway and Buller, 1979, Weiss *et al*., 1979, Wright *et al*., 1990). The effect of these conditions leads to increased activity of PldA and an increase in levels of degraded phospholipids (Audet *et al*., 1974). These findings suggested a role in OM lipid homeostasis for PdlA.

The elucidation of the PldA crystal structure led to the identification of the active site being localised to the external leaflet of the OM and that activity could be regulated by a reversible dimerisation mechanism (Snijder *et al*., 1999) (**Figure 1.8A**). When the OM lipid asymmetry is intact, PldA adopts a non-active monomeric state. However, under OM stressed conditions, which leads to the presence of phospholipids at the outer leaflet of the OM, induction of PldA dimerisation occurs. The effect of dimerisation leads to the formation of two active sites at the subunit interface, where Ca^{2+} ions are able to bind and phospholipase activity is activated (Dekker *et al*., 1997).

The crystal structure of *E. coli* PldA has shown how it hydrolyses glycerophospholipids in the OM (Snijder *et al*., 1999). PldA consists of β-barrel made up of 12 anti-parallel β-strands. An intricate hydrogen bonding network blocks

the central cavity of PldA and three active site residues, Asn156, His142 and Ser144, are organised on the exterior of the β-barrel (**Figure 1.8B**) (Kingma *et al*., 2000). This catalytic triad is normally associated with the inert glucosamine backbone of lipid A in the outer leaflet (Snijder *et al*., 2001). During activity, Ser144 is irreversibly sulfonylated by the palmitate analogue, hexadecanesulfonyl fluoride, whilst the Ca^{2+} are involved in stabilising this complex during hydrolysis, by re-ordering the local water molecules (Snijder *et al*., 1999, Kingma *et al*., 2002, Baaden *et al*., 2003).

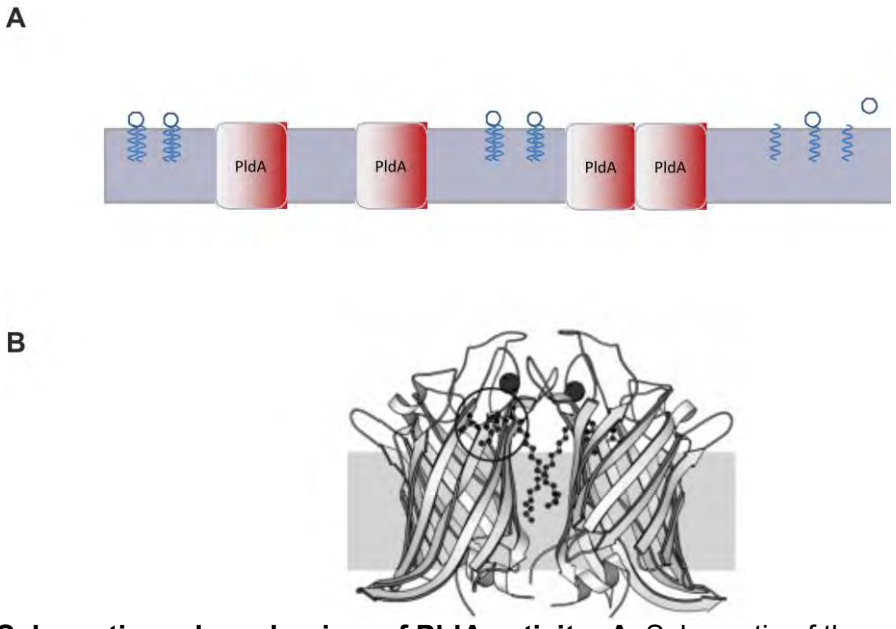


Figure 1.8 Schematic and mechanism of PldA activity. **A** Schematic of the reversible dimerisation that occurs between PldA monomers (red) in response to the accumulation of phospholipids in the outer leaflet of the OM (blue). Phospholipids are chemically cleaved by PldA. **B** Crystal structure representation of the dimer produced (taken from Kingma *et al.* 2002).

PldA is able to hydrolyse a wide range of phospholipid substrates and can remove the fatty acyl-chain at the sn-1 and sn-2 from the glycerophosphodiester backbone of both phospholipids and lysophospholipids (Horrevoets *et al*., 1989)]. PldA is largely unspecific for polar head groups but has the ability to select fatty-acyl chains of 14 carbons or more in length by the presence of acyl-chain binding pockets (Stanley *et al*., 2006, Stanley *et al*., 2007).

1.7.1.3 PagP

Like PldA, PagP is also an OMP that utilises phospholipids in the outer leaflet of the OM as a substrate (Bishop, 2008). Similarly to PldA, PagP is expressed at low levels and remains dormant in unstressed cells, but undergoes a conformational change in response to stress (Jia *et al*., 2004).

PagP is induced by the PhoP/PhoQ signal transduction pathway in response to decreased divalent cations (Bishop, 2005). PhoQ is a transmembrane sensory transducer, which directs a patch of acidic amino acids into the anionic glycerophospholipids located at the outer surface of the IM (Bader *et al*., 2005). Mg^{2+} ions are thought to bind the acidic patch to the lipid polar head groups, which keeps PhoQ in a repressed state (Cho *et al*., 2006). Limitation of Mg^{2+} ions activates PhoQ triggering a phosphorylation cascade that leads to the activation of the transcription factor PhoP (Bader *et al*., 2003). This in turn increases the expression of lipid A modification enzymes, of which PagP is a member.

PagP catalyses the transfer of a palmitate chain from the sn-1 position of a phospholipid to the hydroxyl group on the R-3-hydroxymyristate chain at position 2 of lipid A. This leads to the formation of a hepta-acylated LPS molecule and a lysophospholipid (**Figure 1.9A**) (Bishop, 2000). It is presumed that these lipid A modifications improve the quality of the OM under conditions where divalent cations are limited, as these modified lipid A molecules are more hydrophobic (Miyadai *et al*., 2004).

The structure and dynamics of *E. coli* PagP have been determined by both X-ray crystallography and NMR (Ahn *et al*., 2004, Hwang *et al*., 2002). PagP contains 8 anti-parallel β-strands with a short α-helix at its N-terminus. It is orientated in the membrane with the barrel axis titled by 25 degrees (**Figure 1.9B**) (Evanics *et al*., 2006). PagP contains a palmitate recognition pocket, known as the "hydrocarbon ruler", which resides in the barrel and is localised in the outer LPS exposed region of the protein.

The exact molecular mechanism of enzymatic activity of PagP is not fully understood, but it has been shown that PagP exists in two dynamic states, a repressed R-state and a catalytically active T-state (Hwang *et al*., 2004). Catalysis most likely proceeds when both the phospholipid and lipid A form a ternary complex with PagP (Bishop, 2005). Conserved proline residues are present in a loop region between strands A and B (L1), and it is the reordering of L1 that gives rise to the T-state (Bishop, 2000). Lipid A palmitoyltransferase activity seems to be dependent on L1, although other residues (His33, Asp76 and Ser77) are also thought to play a role (Bishop, 2005, Hwang *et al*., 2004). In addition to the conserved prolines in L1, there are conserved prolines present between strands F and G, which are thought to have a role in the lateral migration of lipids in the outer leaflet (Ahn *et al*., 2004).

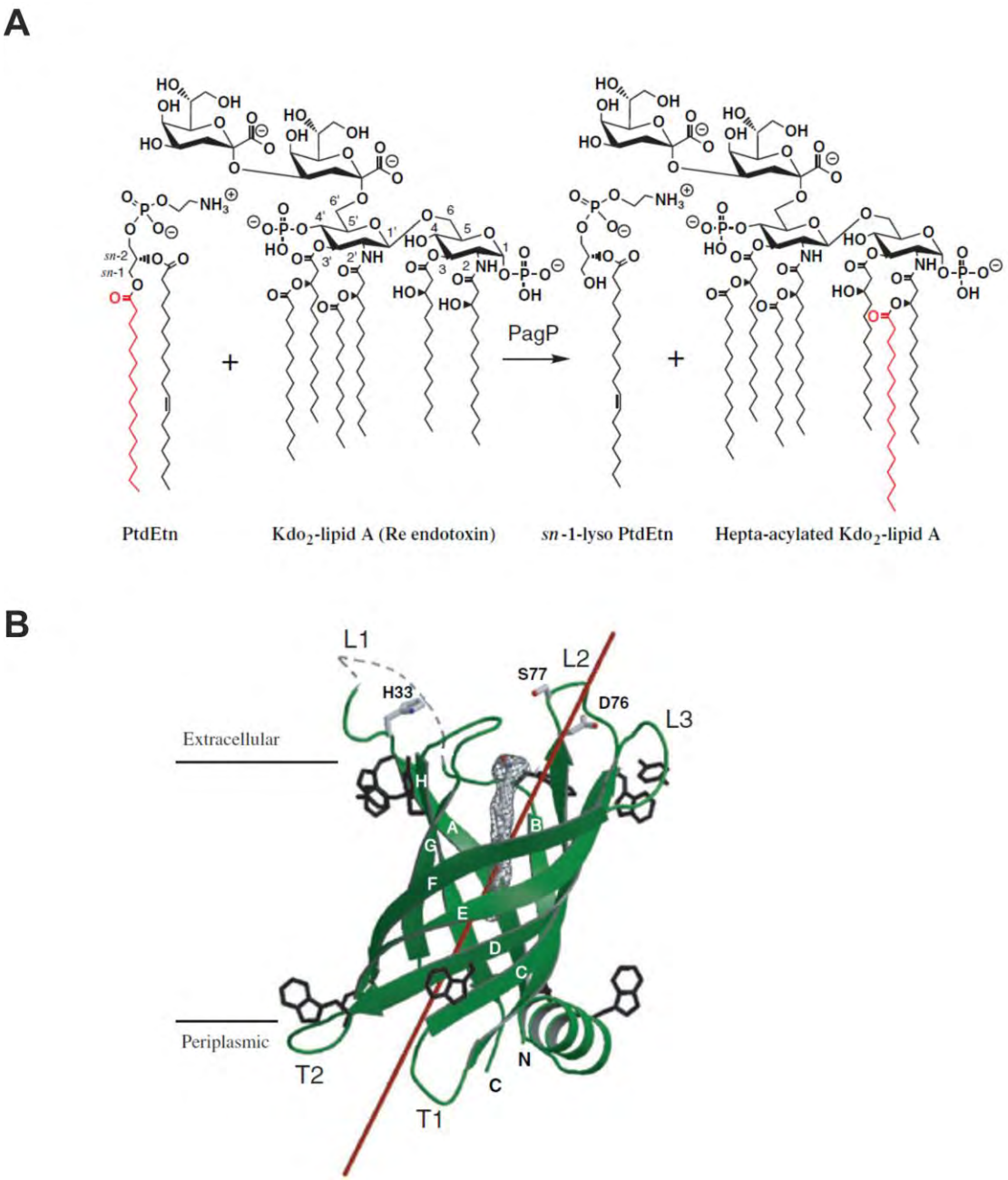


Figure 1.9 Models of PagP activity (taken from Bishop 2005). **A** Enzymatic reaction that PagP catalyses. **B** The 25 degree tilted β-barrel of PagP. L1, the site that converts PagP from an inactive R-state and a catalytically active T-state is indicated by a dashed line. Residues H33, D76 and S77 that are involved in this process are also indicated.

The exact contribution of PagP activity to the OM barrier function seems to be context dependent and can negatively influence the properties of the OM under unstressed conditions (Miyadai *et al.*, 2004). It should be noted that PagP activity is

insensitive to divalent cations and thus provides a specific response to the presence of phospholipids in outer leaflet (Bishop, 2000, Brozek *et al*., 1987). Under these conditions, transacylation to lipid A might provide a superior adaptive response, compared to the hydrolysis of phospholipids which is otherwise catalysed by PldA, albeit an inefficient one (Weiss *et al*., 1979). It should be noted that the presence of palmitoylated lipid A in the OM can lead to induction of the σ^E stress response (Tam and Missiakas, 2005).

1.7.1.4 The Mla pathway

The Mla pathway is a conserved pathway in Gram-negative bacteria with homologous pathways identified in chloroplasts and Actinobacteria (Benning, 2008, Mohn *et al*., 2008, Pandey and Sassetti, 2008). The function identified with this pathway is the retrograde trafficking of phospholipids from the OM (Malinverni and Silhavy, 2009).

The Mla pathway consists of 6 proteins, MlaA through to MlaF. These genes are present on the same operon and are co-expressed, although in *E. coli*, MlaA is present elsewhere from the MlaB-F subset (Gama-Castro *et al*., 2008, Casali and Riley, 2007). These genes have differing or unknown functions that are present in many of the subcellular locations. MlaA is an OM lipoprotein (Suzuki *et al*., 1994, Juncker *et al*., 2003) and MlaC is a periplasmic, substrate binding protein (Lopez-Campistrous *et al*., 2005, Linton and Higgins, 1998). At the IM, MlaB, -D, -E and -F comprise an IM ABC transport machine, with MlaF containing a classic ABC

transport nucleotide binding component (Linton and Higgins, 1998). MlaE is an integral IM protein with a sequence similar to those found in prokaryotic ABC import proteins (Casali and Riley, 2007). MlaD is a substrate-binding protein which is localised to the periplasmic face of the IM (Linton and Higgins, 1998). Finally, MlaB is a cytoplasmic protein with a STAS domain that is thought to bind to NTPs (Casali and Riley, 2007, Nikaido, 2003).

Individual deletions of these genes demonstrated an OM defect with increased sensitivity to SDS and EDTA. Further to this, increased PldA activity rescued this sensitivity and the use of PagP activity demonstrated the presence of phospholipids in the outer leaflet of the OM, in these mutant strains (Malinverni and Silhavy, 2009). It should be noted that loss-of-function *mla* mutations lead to OM defects, whereas with the loss of either PagP or PldA, no such phenotypes are observed (Bishop, 2008). This highlights the importance of this pathway in regards to maintaining lipid asymmetry.

The Malinveri 2009 study proposed a possible pathway involving these Mla proteins (**Figure 1.10**). MlaA may take phospholipids out of the OM and pass them onto MlaC. From here, MlaC passes the phospholipids to the MlaBDEF complex. The fate of the phospholipids are unknown in this pathway, but it is significantly different to the activity of PldA or PagP, in that only removal of phospholipid is involved rather than destruction of it.

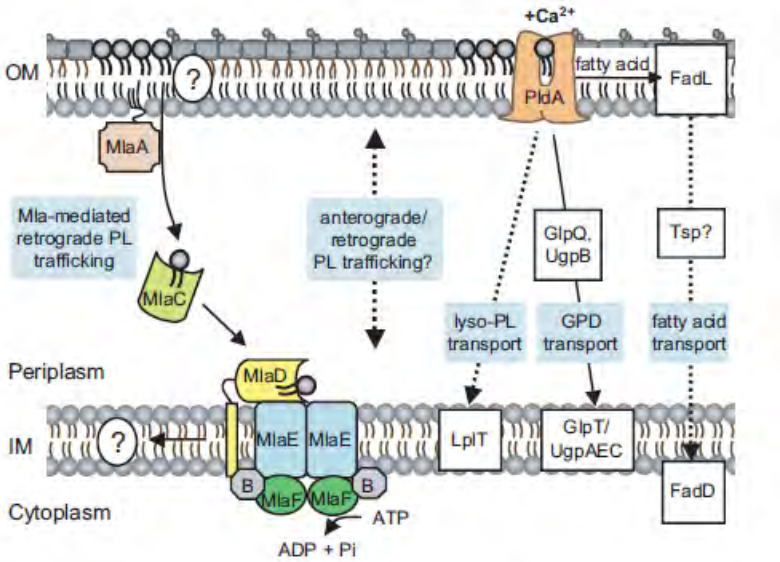


Figure 1.10 The Mla pathway in *E. coli* (Figure taken from Malinverni and Silhavy 2009). The Mla pathway removes phospholipids from the OM and delivers them to MalBEDF complex in the IM. The fate of the phospholipids are unknown, whether they are chemically cleaved or inserted into the IM.

1.7.1.5 Concluding remarks

The maintenance of the asymmetry of the OM is crucial for barrier function. Conditions which compromise LPS interaction, such as the presence of EDTA, can lead to the presence of lipids in the outer leaflet of the OM, which greatly affects barrier function.

Three major players involved in this process have been identified each utilising different methods. PldA serves to completely remove phospholipids from the OM, PagP modifies lipid A to patch compromised regions of the OM and the Mla pathway is involved in retrograde trafficking of lipids. These processes serve to maintain OM asymmetry.

The process in which lipids are sent and assembled to the OM (anterograde trafficking) is still unknown, but the identification of factors involved would contribute greatly in understanding OM biogenesis.

1.8 Lipoproteins

Lipoproteins are the other major protein species in the OM, although they are not exclusive to this membrane like OMPs, as they are also observed on the IM. Their defining feature is that they contain a lipid moiety attached to their N-terminus, which serves the purpose of anchoring them to the membrane. Generally lipoproteins in Gram-negative bacteria protrude into the periplasmic space from either the OM or IM, but surface exposed lipoproteins also exist (Tokuda, 2009, Tokuda et al., 2007, Schulze and Zuckert, 2006, Cowles *et al*., 2011).

E. coli has been shown to express over 90 lipoproteins (Miyadai *et al*., 2004) which have been shown to participate in a variety of functions, such as OM biogenesis (OMPs and LPS) (Wu et al., 2005, Wu et al., 2006, Malinverni et al., 2006, Paradis-Bleau et al., 2010, Typas et al., 2010), secretion (Collin *et al*., 2011), signalling (Laubacher and Ades, 2008), cell division (Uehara *et al*., 2009), transport of substrates and drug efflux (Bernadac *et al*., 1998, Clavel *et al*., 1998, Ehrmann *et al*., 1998, Nikaido, 1998). Additionally, lipoproteins may be of medical importance. For example, it has been shown in *Borrelia burgdorferi*, the Lyme disease spirochete, that surface exposed lipoproteins cause an inflammatory response in human host cells (Fraser *et al*., 1997, Ramesh *et al*., 2003, Scragg *et al*., 2000). Lipoproteins may also interact with periplasmic proteins, OMPs, other lipoproteins, the murein layer and can also be part of vital membrane complexes to perform their functions.

Lipoproteins are initially synthesised in the cytoplasm and are translocated through the IM by virtue of their N-terminal signal sequence, which is followed by lipid modification at their N-terminal cysteine residue on the periplasmic surface of the IM (Sankaran and Wu, 1994). Localisation of the mature lipoproteins is dependent on the Lol (localisation of lipoproteins) system, which guides lipoproteins to OM or IM. The major player in localisation is the IM ATP-dependent translocator complex, LolCDE, which localises lipoproteins based on the +2 residue (i.e the residue next to the N-terminal cysteine residue). Localisation to the OM requires both a periplasmic chaperone (LolA) and an OM lipoprotein (LolB) for full maturation.

It should be noted membrane anchoring of lipoproteins may not be necessarily a criterion for absolute function. Multiple essential lipoproteins of *E. coli* have been shown to function properly even when their signal sequences are deleted, so long as they are expressed in excess (Tsukahara *et al*., 2009a). This suggests that lipidation of proteins is a matter of economy rather than overall functionality. However, mislocalisation of certain envelope lipoproteins can abolish their function and in some cases, contribute to toxicity. Examples such as localising the OM-associated PBP (penicillin-binding protein) cofactor LpoA to the IM, renders it unable to be involved in peptidoglycan biosynthesis (Paradis-Bleau *et al*., 2010) and IM localisation of the peptidoglycan-associated lipopotein, Lpp, leading to cell lysis (Yakushi *et al*., 1997). Because of the essential roles that lipoproteins play in cells, proper localisation and assembly of them, is critical for growth and viability.

1.8.1 Biogenesis of lipoproteins

Lipoproteins are synthesised in the cytoplasm as precursors with a signal sequence known as a lipobox. Lipoboxes have a consensus sequence of Leu-(Ala/Ser)-(Gly/Ala)-Cys (Hayashi and Wu, 1990). Once synthesised, these precursors are then transported to the periplasmic side of the IM, via the SEC machinery, where processing into the mature lipoprotein begins (Tokuda, 2009, Sankaran and Wu, 1994).

Three well conserved enzymes are involved in the processing steps that ultimately lead to cleavage of the lipobox sequence and lipidation of the N-terminal cysteine residue (**Figure 1.11**). Firstly, phophatidylglycerol/prolipoprotein diacylglyceryl transferase (Lgt) adds diacylglycerol to the N-terminal cysteine by a thioether linkage. Then the signal peptide is cleaved by LspA (lipoprotein signal peptidase/signal peptidase II). Lastly, the N-terminal cysteine residue is N-acetylated by Lnt (phospholipid/apolipoprotein transacylase). These three enzymes are strongly conserved in Gram-negative bacteria and are essential in *E. coli*.

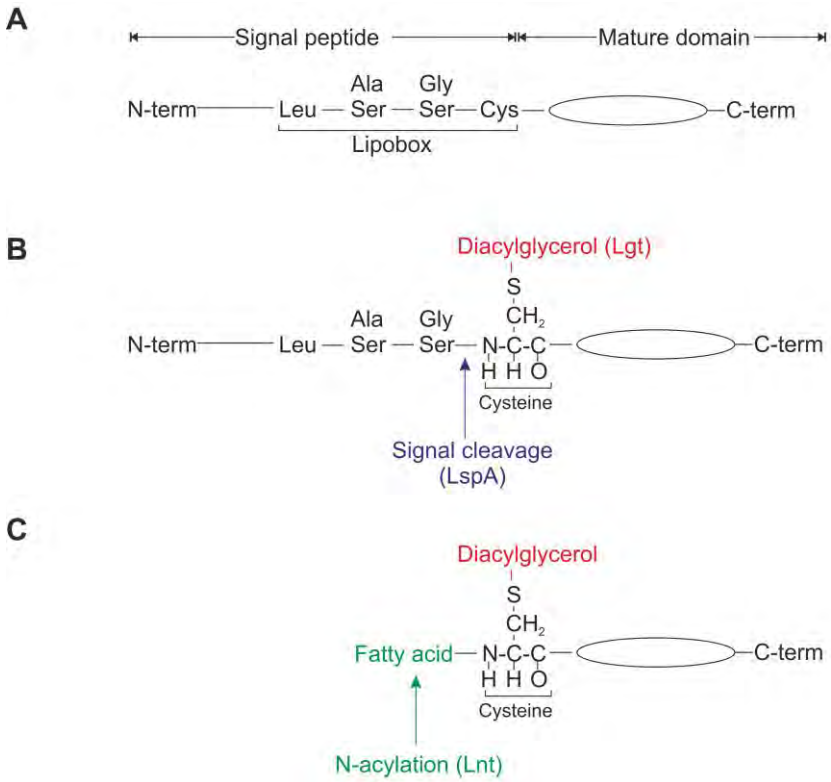


Figure 1.11 Initial lipoprotein processing steps. **A** Initially the lipoprotein exists with a lipobox signal sequence near the N-terminus. **B** The enzyme Lgt catalyses the addition of diacylglycerol to N-terminal +1 cysteine residue (red) whilst LspA cleaves the lipopbox sequence (blue). **C** Further processing occurs by Lnt, which N-acylates the protein (green).

1.8.2 Sorting of lipoproteins

Lipoprotein sorting to the IM or OM is largely dependent on the N-terminal second residue (+2 residue) (Yamaguchi *et al*., 1988). The presence of an Asp residue at +2 causes localisation of a lipoprotein to the IM whereas any other residue at the +2 position leads to OM localisation. This general localisation effect is called the "+2 rule". It should be noted that Phe, Trp, Tyr, Gly, and Pro residues at +2 can also act as IM retention signals, although no lipoproteins in *E. coli* have any of these five residues at their +2 positions (Seydel *et al*., 1999).

Sorting of lipoproteins is not entirely dependent on the +2 residue. The +3 residue has been shown to play an important role in localisation as the presence of His or Lys at +3 causes only partial retention of lipoproteins with a +2 Asp to the IM (Gennity and Inouye, 1991, Yokota *et al*., 1999). Additionally, the presence of Asp, Glu or Gln at +3 with a +2 Asp leads to strong IM retention (Terada *et al*., 2001). It should be noted that the presence of Ser at +3, leads to OM localisation, unless there is +2 Asp, strongly indicating the importance of the +2 Asp for IM retention of lipoproteins.

Overall, it seems that a negative charge or amide group at +3 with Asp at +2 leads to strong IM retention. Asn at +2 can also act in IM retention, providing that the +3 is Asp, as in the case of the *E. coli* protein AcrE (Klein *et al*., 1991, Seiffer *et al*., 1993). The +2 rule is fairly well conserved in the *Enterobacteriaceae* family (Babu *et al*., 2006, Lewenza *et al*., 2006) although there are known exceptions such as MexA of *Pseudomonas aeruginosa*, which uses a +2 Gly (Narita and Tokuda, 2007).

The method by which this IM retention signal from the +2 Asp residue works is that the ATP-dependent IM translocator, LolCDE (discussed below), fails to interact with lipoproteins that contain these residues (Masuda *et al*., 2002). This process, termed "Lol avoidance", has been demonstrated by modifying OM lipoproteins with IM retention signals and observing localisation to the IM (Sakamoto *et al*., 2010). LolCDE only recognises the N-acylated group of lipoproteins rather than the +2 residue (Hara *et al*., 2003). Rather, the IM contains 75% PE, which is positively charged, steric and electrostatic interactions occur between PE and the +2 Asp

residue, which forms a phospholipid-lipoprotein complex that does not interact with LolCDE and remains in the IM.

1.8.3 The Lol system

So far, only IM lipoprotein biogenesis has been discussed. OM lipoprotein sorting requires 5 additional proteins, LolA through LolE, which are all essential for the growth of *E. coli*. LolCDE forms an ABC (ATP-Binding Cassette) transporter in the IM, which is responsible for the extraction of lipoproteins from the IM. LolA is a periplasmic chaperone which guides the lipoprotein to the OM. The final member is LolB, which is an OM lipoprotein responsible for insertion.

1.8.3.1 The LolCDE-Lipoprotein complex

LolCDE forms an ABC transporter in the IM (Yakushi *et al*., 2000). This complex is made up of LolC, LolD and LolE subunits in a 1:2:1 stoichiometry. LolC and LolE are membrane subunits, with each protein having four membrane spanning helices and one large periplasmic loop between the first and second periplasmic helices (Yasuda *et al*., 2009). Additionally, the sequences of LolC and LolE are very similar to each other, explaining the similar organisation in the IM. LolD is a nucleotide binding protein, which has conserved ABC signature, Walker A and Walker B motifs (Yakushi *et al*., 2000).

The crystal structure of MJ0796, a LolD homologue in *Methanococcus jannasachii* with 43.7% identity to *E. coli*, was solved to 2.7 Å resolution (Yuan *et al*., 2001). The

structure shows a fold similar to the ATPase subunits of other ABC transporters (Smith *et al*., 2002).

1.8.3.2 LolA

LolA is a periplasmic chaperone that has been shown to bind and release several OM lipoproteins from the IM (Matsuyama *et al*., 1995). LolA is able to bind to lipoproteins with a 1:1 stoichiometry to form a water-soluble complex and has the important task of transporting hydrophobic lipoproteins through the periplasm.

The crystal structure of LolA has been solved to 1.65 Å (Takeda *et al*., 2003a) (**Figure 1.12A**). The structure indicates a fold reminiscent of an incomplete β-barrel consistent of 11 anti-parallel β-strands with a lid of 3 α-helices. The lid and incomplete β-barrel form a hydrophobic cavity, which is a likely binding site for the N-acylated N-termini of lipoproteins. The hydrophobic cavity of LolA is mainly made up of aromatic residues.

Mutational analysis has revealed that Arg43 of LolA is important for function (Miyamoto *et al*., 2001). This residue closes the lid of the hydrophobic cavity by forming a hydrogen bond between the β2-strand the α-helical lid. An R43L mutant has been shown to display interactions with LolCDE, but fails to transfer lipoproteins to LolB which result in periplasmic accumulation of lipoproteins. This mutant has increased hydrophobic interactions between LolA and lipoproteins, thereby causing inefficient transfer to LolB (Taniguchi *et al*., 2005).

Structural analysis of the R43L mutant further confirmed that R43 has a critical function in closing the lid (Miyamoto and Tokuda, 2007), and that the hydrophobic cavity of LolA undergoes opening and closing upon the binding and release of lipoproteins respectively (Pastukhov and Ropson, 2003). Additionally, it was found that opening of the LolA hydrophobic cavity is essential for growth in *E. coli*, as a mutant with a permanently closed cavity led to cell envelope stress and activated Cpx levels (Raivio and Silhavy, 2001, Tao *et al*., 2010).

1.8.3.3 LolB

LolB is an OM lipoprotein that is the receptor for OM lipoproteins and has a role of inserting them into the OM (Matsuyama *et al*., 1997). Like LolA, LolB is essential for growth and depletion of LolB leads to the periplasmic accumulation of OM targeted lipoproteins (Tanaka *et al*., 2001).

The crystal structure of mLolB, a soluble LolB derivative lacking the N-terminal acyl chain, demonstrated that it had a similar structure to LolA, despite their amino acid sequences being dissimilar (Takeda *et al*., 2003b) (**Figure 1.12B**). However, although similar, there are some differences between the structures of LolA and LolB. LolA has an extra loop region containing a short helix, that is thought to prevent retrograde transfer of lipoproteins (Okuda *et al*., 2008). Also, LolA has a twelfth β-strand at the C-terminus. Finally, the nature of the hydrophobic cavity is quite different, with LolB mainly harbouring Leu and Ile residues and LolA mainly having aromatic residues. This changes the affinity for lipoproteins, as the side

chains that are present in LolB have more flexibility than the aromatic rings, which increases the intensity of the hydrophobic interactions when compared to LolA.

The soluble mLolB is functional, despite lacking an N-terminal anchor, and is able to replace LolB providing that it is expressed at a higher level (Tsukahara *et al*., 2009b). This suggests that OM targeting is an intrinsic feature of LolB rather than its possible mislocalisation. The method of lipoprotein insertion by LolB is largely unknown, but it thought that residue L68 is important for this function, as it is part of a hydrophobic loop that could protrude into the membrane (Takeda *et al*., 2003a).

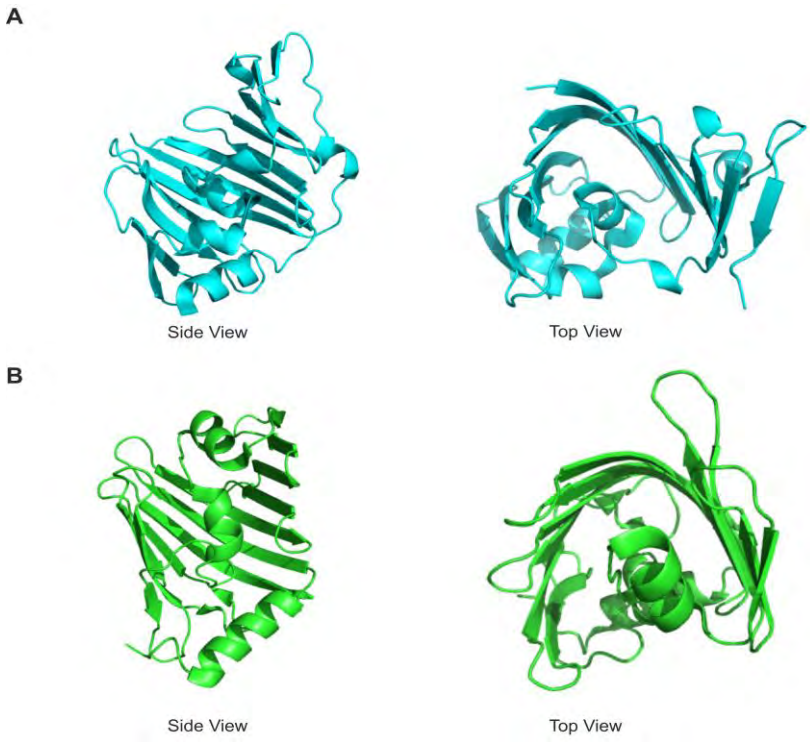


Figure 1.12 Crystal structures of LolA and LolB. A LolA (PDB: 1IWL) and **B** LolB (PDB: 1IWM), exhibit a remarkably similar fold despite their dissimilar amino acid sequences (~10% homology). Both structures contain a hydrophobic cavity formed from an unclosed β-barrel and an α-helical lid.

1.8.3.4 Lipoprotein Transfer

Co-purification experiments with LolCDE, ATP and lipoproteins have revealed the molecular events that are involved with LolA-dependent lipoprotein release (Ito *et al*., 2006, Taniguchi and Tokuda, 2008). Firstly, LolCDE binds to an OM targeted lipoprotein which leads to an increase in the affinity of LolD for ATP. The ATP binding to LolD then weakens the hydrophobic interaction between the lipoprotein and LolCDE. Finally, ATP hydrolysis causes the transfer of lipoproteins to LolA and opens the hydrophobic cavity of LolA (**Figure 1.13A**).

In vivo cross-linking experiments have revealed how LolA interacts with LolCDE (Okuda and Tokuda, 2009). Although LolC and LolE have similar sequences (26% identity) and topologies (Narita *et al*., 2002, Yasuda *et al*., 2009), only LolC is observed to bind to LolA. The binding of lipoproteins to LolCDE increases the interaction between LolC and LolA and that this interaction requires the use of the entrance of the hydrophobic cavity of LolA. Additionally, these experiments have revealed that only LolE specifically interacts with lipoproteins, rather than LolC. It should be noted that *in vitro*, minimum lipoprotein releasing activity is observed from LolDE complex without LolC (Kanamaru *et al*., 2007). This result suggests that LolC may function as a scaffold for LolA and that LolE is involved in the recognition of OM targeted lipoproteins.

The periplasmic regions of LolC and LolE exhibit sequence similarity to LolA and LolB, which suggests that these proteins may also have hydrophobic cavities and

that lipoprotein transfer could be mediated by the interaction of these cavities (Okuda and Tokuda, 2009).

The same cross-linking experiments suggested that lipoprotein transfer between LolA and LolB may be mediated in a "mouth-to-mouth" manner (**Figure 1.13B**). Indeed this has now been shown by NMR, the hydrophobic cavities of LolA and LolB directly interact with each other and the lipoprotein is transferred between the two. LolB has a higher affinity for lipoprotein than LolA (Taniguchi *et al*., 2005), thus once this cavity connection is made, the transfer of lipoprotein could be driven by the direction of higher affinity.

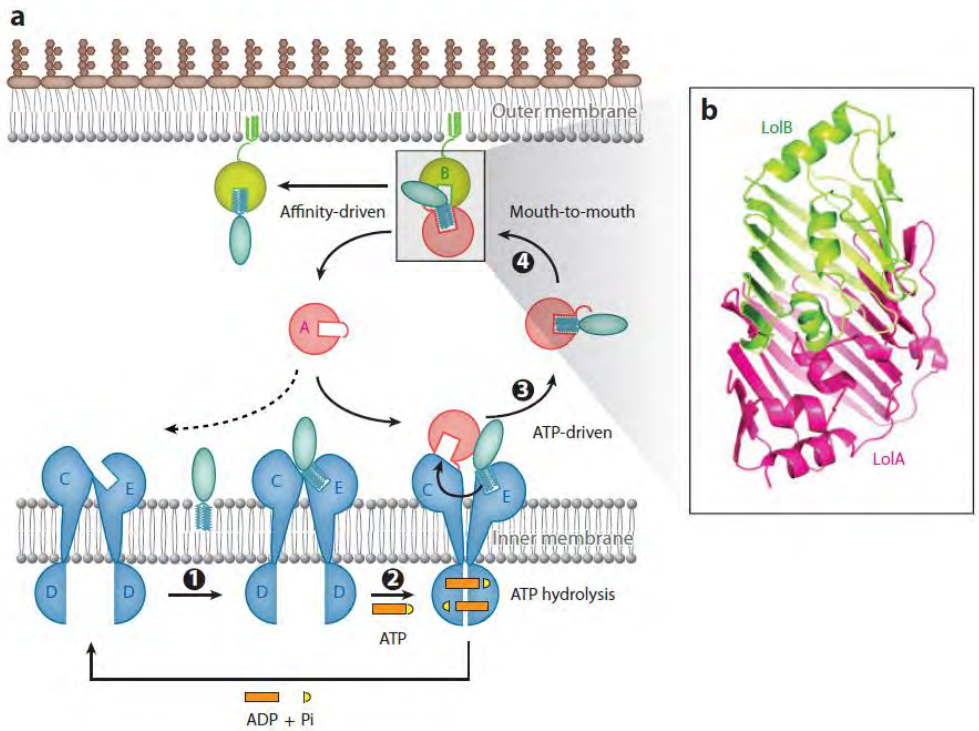


Figure 1.13 Molecular events underlying lipoprotein sorting by the Lol system (taken from S. Okuda and H. Tokuda, 2011). **A** Lipoproteins first bind to LolE (step1) and are then transferred to LolA with ATP mediated release from the LolCDE complex (step2). LolA then forms a hydrophilic complex with the lipoprotein (step3), which is then transferred to LolB by an affinity-driven interaction (step4). **B** A model displaying mouth-to-mouth interactions between the cavities of LolA and LolB.

1.8.4 Concluding remarks

The conserved Lol system has been shown to be critical for OM biogenesis and viability of *E. coli*. Many lipoproteins have been shown to play essential roles in *E. coli* and the correct localisation of them is important for their function.

The molecular mechanisms underlying transfer of lipoproteins from the IM to OM have been clarified. IM lipoproteins are retained in the IM by Lol avoidance, whereas OM lipoproteins are recognised by the LolCDE complex and are transferred to the periplasmic chaperone LolA in an ATP-dependent manner. LolA then transfers the lipoprotein to the OM lipoprotein receptor LolB in a mouth-to-mouth manner. However, the exact nature of transferring lipoproteins from LolB to the OM is still unknown nor is it known if the same mechanism is used for surface exposed lipoproteins.

1.9 Outer membrane beta barrels

Outer membrane proteins (OMPs) form the major class of proteins present in the outer membrane of Gram-negative bacteria. Their presence solely within this membrane is one of the defining features that distinguish the OM from the IM. OMPs are arranged in a unique β-barrel topology in the OM, which enables the cell to interact and mediate exchange with its environment.

Due to their contact with the external milieu, OMPs have been implicated in a variety of functions in *E. coli*. The most abundant OMP species serve as non-specific, passive transporters that allow the diffusion of ions, sugars and small hydrophilic molecules (generally smaller than 700 Da). Examples of these are PhoE, OmpC and OmpF. OMPs can also serve as specific transporters, such as the sucrose transporter ScrY. Apart from metabolite transportation, OMPs are involved in diverse functions, such as adhesion (e.g. Ag43), ATP-mediated export (TolC), secretion (Type V secretion proteins) and OMP biogenesis (BamA) (Tamm *et al*., 2004).

Although OMPs have diverse functions, they all share a similar structure and topology. All OMPs are arranged as a β-barrel, which in most cases consist of an even number of β-strands. The strands are arranged in an anti-parallel manner, which produces a central hydrophilic pore (**Figure 1.14**). Main chain hydrogen bonding between the strands allows this structure to be maintained. The β-strands are tilted relative to the axis and each strand of the protein is equal in length and the

barrel is closed by hydrogen bonding between the first and last β-strand, with both the N and C-terminus present in the periplasm. β-barrel proteins are very stable and *in vitro*, temperatures close to boiling point are needed to denature them. It should be noted that there are many OMPs that are composed of trimers, with each monomer contributing to the β-barrel (Seshadri *et al*., 1998).

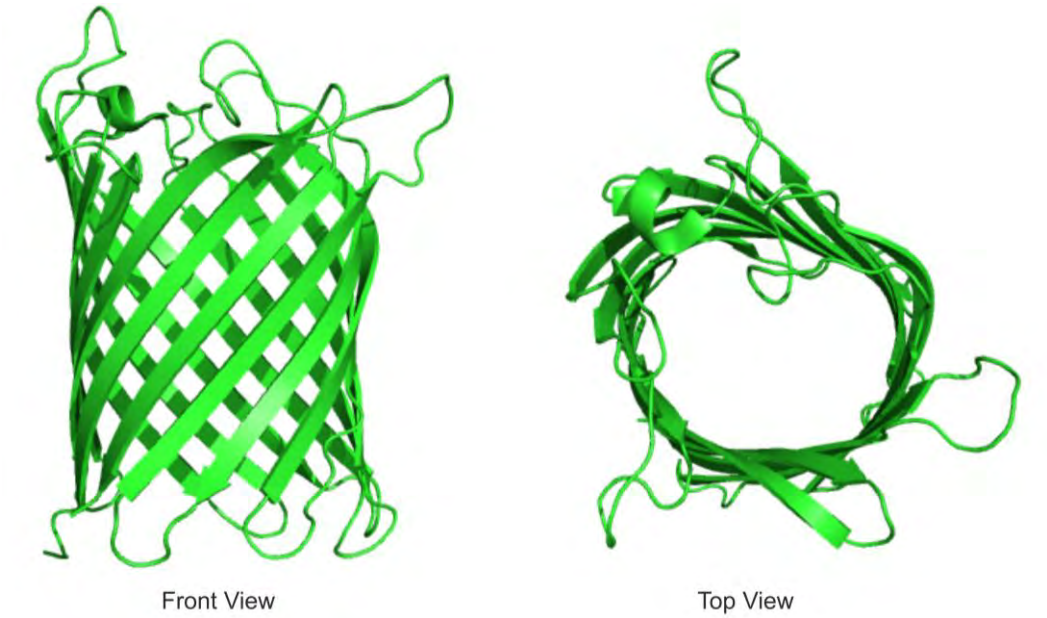


Figure 1.14 Structure of OmpG. A β-barrel structure is formed from an extensive hydrogen bonding network between extended β-strands. With OmpG (PDB: 2X9K), 14 β-strands are used to form a hydrophilic channel.

As β-barrels reside in the OM and produce a hydrophilic pore (Tamm *et al*., 2004), they have a distinct arrangement of amino acids. Residues involved in the water filled pore are generally hydrophilic and/or charged, whereas those that are in contact with the lipid bilayer of the OM tend to be short, hydrophobic residues. It should be noted that aromatic residues are particularly prominent in OMPs and can have roles in other functions (discussed below) (Wimley, 2003).

Apart from just providing the hydrophilic pore for exchange, other parts of the OMP may be involved in function. Aromatic side chains may be orientated around the barrel rims, which leads to the production of „aromatic girdles" (Nikaido, 2003), which are presumed to influence the stability of the folded OMP. β-hairpins are also another common feature, which are formed from the β-strands, and are also thought to be involved in structural maintenance (Koebnik, 1996). OMPs may also contain sizeable periplasmic domains, such as POTRA domains (discussed later) (Wu *et al*., 2005), that may be involved in the function of the OMP or act as a scaffold to other proteins. α-helices can also protrude from the β-barrel into the periplasm and produce smaller barrel-like structures (Koronakis *et al*., 2000). Results from *in silico* modelling suggests that β-barrels may also contain extracellular loops, which may act as plugs involved with substrate specificity (Delattre *et al*., 2010, Kramer *et al*., 2000, Adamian *et al*., 2010) and the presence of helical „out-clamps", which may reinforce weakly stable β-strands at the membrane-solvent interface (Naveed *et al*., 2009).

OMPs are an essential part of the OM and confer various functions to the cell. Biogenesis of OMPs is now widely accepted to be dependent on the evolutionary conserved Bam complex (Wu *et al*., 2005, Anwari *et al*., 2010, Anwari *et al*., 2012, Gatsos *et al*., 2008, Kim *et al*., 2007, Sklar *et al*., 2007a, Volokhina *et al*., 2009). Although some OMPs have been shown to insert and fold themselves spontaneously into phospholipid bilayers, the process is much more efficient with the Bam complex (Burgess *et al*., 2008, Kleinschmidt, 2003, Tamm *et al*., 2001).

There are some other features of OMP biogenesis that need to be discussed, such as transport through the IM and periplasm, before addressing the Bam complex.

1.9.1 Transport through the IM

OMPs are able to be transported through the IM using either the Sec or SRP pathway by virtue of their N-terminal signal sequence. Generally, OMPs are transported using the SecB-dependent post-translational pathway. However, with increasing hydrophobicity of signal sequences of certain OMPs (e.g. OmpA, LamB), the co-translational SRP is used (Bowers *et al*., 2003, Lee and Bernstein, 2001).

1.9.2 Late translocation steps

Nascent precursor OMPs emerge from the Sec translocon in an N to C-terminal manner (du Plessis *et al*., 2010). Since unfolded OMPs contain many hydrophobic regions, there is the possibility of these proteins aggregating in the periplasm. To prevent this, numerous periplasmic factors have been implicated in preventing this and aiding release of OMPs from the translocon.

SecD and SecF are two accessory components of the Sec translocon. The crystal structure of the SecDF complex suggests that they are involved in mediating the late steps of translocation in a manner dependent on the PMF (Tsukazaki *et al*., 2011). There is also evidence that the IM associated periplasmic chaperone, PpiD, influences transport of OMPs through the translocon by interactions with emerging proteins from the translocon (Antonoaea *et al*., 2008).

Skp, a periplasmic chaperone, has been shown to interact with translocation intermediates of OMPs at the periplasmic face of the IM. This interaction may aid dissociation of nascent OMPs from the IM, as has been observed for OmpA (Schafer *et al*., 1999).

Once released from the IM, the N-terminal signal sequence used to lead translocation is cleaved by periplasmic proteases. Mutational analysis has indicated that signal sequence cleavage is a key stage for successful OMP assembly, as this can interfere with the targeting and assembly of the OMP (Carlson and Silhavy, 1993).

1.9.3 Interactions with periplasmic chaperones

Interactions with periplasmic chaperones are crucial for OM biogenesis. Without these chaperone interactions, a toxic accumulation of aggregates occur in the periplasm, as targeting to the OM does not take place. This in turn leads to activation of stress responses (Carlson and Silhavy, 1993). Furthermore, the depletion of critical OMPs in the OM (e.g. LptD), results in a compromised OM barrier function. Periplasmic chaperones have a major role in keeping precursor OMPs in a folding-competent state as well as guiding and targeting OMPs to the OM (Silhavy *et al*., 2010). The current model in OMP assembly suggests that there are two parallel chaperone pathways for OMPs, one involving SurA and another with Skp and DegP (Sklar *et al*., 2007b, Rizzitello *et al*., 2001). This section details the known periplasmic chaperones and gives some functional insight into how they function.

1.9.3.1 SurA

SurA is regarded as the major periplasmic chaperone in *E. coli* (Sklar *et al*., 2007b) and is proposed to have a role in extracytoplasmic protein folding. Evidence for this comes from the observation that Δ*surA* strains show OM defects, delays in OMP maturation, and activate the σ^E regulon and that overexpression of SurA can suppress OM permeability defects (Lazar and Kolter, 1996, Rouviere and Gross, 1996, Missiakas *et al*., 1996). SurA has also been shown to interact specifically with unfolded OMPs (Behrens *et al*., 2001, Bitto and McKay, 2004) and with peptides containing motifs present in OMPs (Bitto and McKay, 2003, Hennecke *et al*., 2005, Stymest and Klappa, 2008, Webb *et al*., 2001, Xu *et al*., 2007).

SurA contains two peptidyl-prolyl-isomerase domains (PPIase), referred to as P1 and P2 (Lazar and Kolter, 1996, Missiakas *et al*., 1996) (**Figure 1.15**). PPIases are able to catalyse the *cis* to *trans* isomerisation of peptidyl-proline bonds. SurA also contains chaperone activity, which has been localised to a domain that is produced from the association of the N and C-terminal domains (Behrens *et al*., 2001, Bitto and McKay, 2002, Watts and Hunstad, 2008).

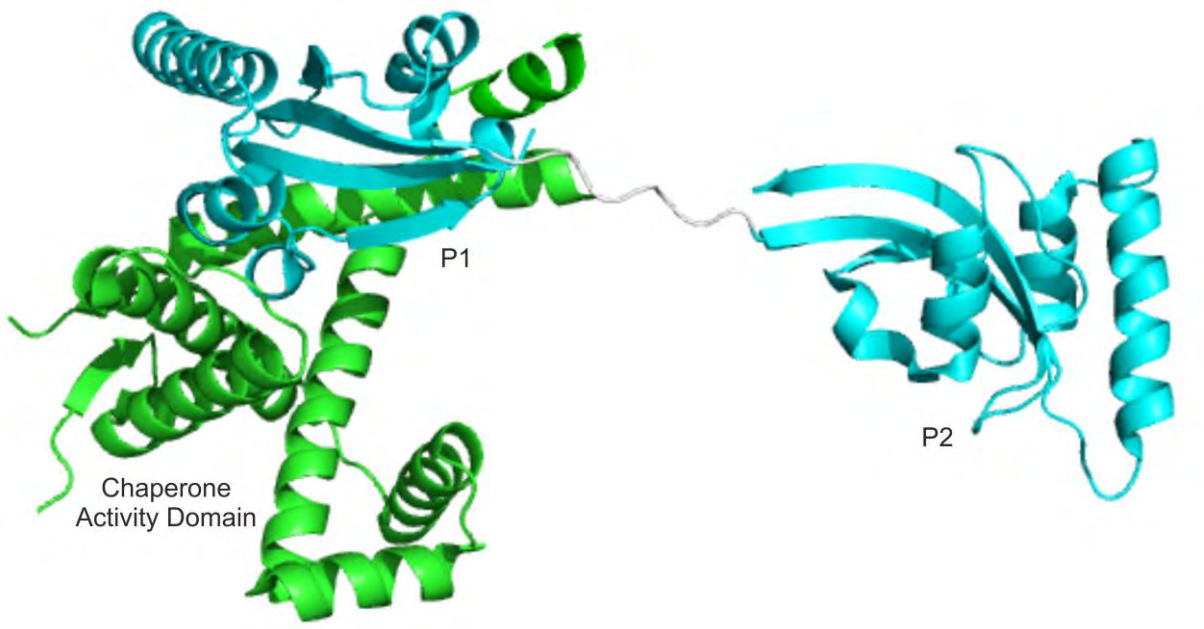


Figure 1.15 SurA exhibits a modular architecture. Structure of SurA (PDB: 1M5Y), with P1 and P2 domains highlighted (cyan). The chaperone activity of SurA has been localised to a domain formed between the N and C-termini.

It is not immediately apparent what the purpose of the P1 and P2 domains in SurA is. It has been demonstrated that deletion of either of these domains causes novobiocin sensitivity in UPEC strains of *E. coli* (Watts and Hunstad, 2008), but this may be a species specific effect. It seems conceivable that the use of PPIase activity could aid folding of OMPs. However, mutations in the catalytic site of P2, can abolish PPIase activity, without affecting the chaperone function of SurA. In addition to this, the P1 domain is conserved in several homologues, but lacks PPIase activity (Giuseppe *et al*., 2010). These data suggest that the PPIase activities of P1 and P2 may contribute very little to overall SurA function. It should be noted that SurA is able to bind to peptides rich in aromatic residues via the P1 domain (Xu *et al*., 2007), which may suggest that they are involved in substrate binding.

SurA has been shown enhance the rate of folding of OMP monomers (Ureta *et al*., 2007, Lazar and Kolter, 1996, Rouviere and Gross, 1996) and depletions or deletions of *surA* strongly induces the σ^E regulon (Rouviere and Gross, 1996), strongly suggesting a role in OMP biogenesis. Further evidence for this comes from the demonstrations that SurA increases OmpT assembly in an *in vitro* β-barrel assembly system (Hagan *et al*., 2010), there are changes in OM density following SurA depletion (Sklar *et al*., 2007b) and SurA being involved with various OM components, such as autotransporters IcsA (Purdy *et al*., 2007), Hbp (Sauri *et al*., 2009), EspP (Ieva and Bernstein, 2009, Ruiz-Perez *et al*., 2009) and the chaperone usher/pilus assembly systems PapC and FimD (Watts and Hunstad, 2008, Vertommen *et al*., 2009). Although SurA is thought to be the major periplasmic chaperone, a proteomic study was performed to determine the range of substrates of SurA (Vertommen *et al*., 2009). From this study, they found that 15 of the 23 OMPs studied were not affected by the absence of SurA. This could imply that SurA exhibits substrate specificity and that these effects are due to the absence of specific SurA-dependent OMPs used in this study. However, it clearly highlights that other chaperone systems are used for the transport of OMPs to the OM.

Crystallographic studies have demonstrated the modular nature of SurA (Bitto and McKay, 2002, Giuseppe *et al*., 2010, Clantin *et al*., 2009). Structural analysis of *E. coli* SurA showed the presence of a core chaperone domain that was tightly associated with the P1 domain and that the P2 domain was connected by a flexible linker region (Bitto and McKay, 2002, Abe *et al*., 2003). Following this result, another

structure of SurA lacking the P2 domain, demonstrated peptide binding with a conformational rearrangement, whereby the P1 domain dissociated from the core chaperone domain, to bind to the aromatic rich peptide and that the chaperone domains of each SurA monomer interacted with each other to produce a dimer (Xu *et al*., 2007). A similar dimer has been crystallised in a structural homologue Par27 of *Bordetella*, although the dimer interface is different (Clantin *et al*., 2009). Whether these conformational changes are physiologically relevant has yet to be proven.

1.9.3.2 Skp

Skp is another periplasmic chaperone that is thought to have a role in OMP biogenesis. Although implicated in the latter steps of IM translocation (Thome *et al*., 1990, Thome and Muller, 1991), further evidence suggests other functional roles. Skp was demonstrated to bind specifically to unfolded OMPs (Chen and Henning, 1996) and form soluble periplasmic intermediates of OmpA (Schafer *et al*., 1999). Additionally, the absence of *skp*, caused increased activation of the σ^E response. These findings suggest a role for Skp in OMP biogenesis.

Another finding, which highlighted the importance of Skp, is that it is required for viability in a Δ*surA* background. This synthetically lethal combination has suggested that SurA and Skp may have overlapping pathways with substrates common to both chaperones and that this pathway is the major one involved in periplasmic handling of OMPs en route to the OM (Sklar *et al*., 2007b, Rizzitello *et al*., 2001). Proteomic analysis has indicated that individual levels of OMPs did not decrease in a Δ*skp* mutant, but levels of all OMPs decreased in a mutant with *surA* and *skp* disrupted, suggesting redundant substrate specificity (Vertommen *et al*., 2009). It should be noted that in *N. meningitidis*, a Δ*surA* and Δ*skp* pair is not synthetically lethal, suggesting that this chaperone pathway is not universal.

The structure of Skp has indicated that it is a homotrimer, which has a core domain, where trimerisation takes place, and three α-helices projecting from it (Walton and Sousa, 2004, Korndorfer *et al*., 2004) (**Figure 1.16**). This structural feature bears some resemblance to chaperones in other organisms, such as mitochondrial

intermembrane chaperone Tm9.10 and the archaeal chaperone perfoldin (Abe *et al*., 2003, Webb *et al*., 2006, Walton and Sousa, 2004). Substrate binding is thought to be mediated by the α-helices which are able to form a hydrophobic pit. Electrostatic and hydrophobic interactions keep the substrate in place in this pit (Qu *et al*., 2007). Studies performed using site-directed fluorescence spectroscopy and NMR have indicated that the β-barrel domain of OmpA is buried in the pit from the projected helices and that the hydrophilic periplasmic domain resides outside of the pit and adopts its native conformation (Qu *et al*., 2009, Walton *et al*., 2009). The flexible nature of the α-helices allows Skp to theoretically bind to diverse sizes of OMPs (Walton and Sousa, 2004). Indeed this seems to be the case as Skp has been shown to bind to broad range of OMPs that include porins (Harms *et al*., 2001, Chen and Henning, 1996, Qu *et al*., 2007, Jarchow *et al*., 2008), OmpA (Bulieris *et al*., 2003), autotransporters (Ieva and Bernstein, 2009) and intimin (Bodelon *et al*., 2009). Further roles of Skp have been reported outside of *E. coli*, such as being implicated in virulence in *Salmonella* (Purdy *et al*., 2007) and has a possible role in cell-cell spread in *Shigella* (Rowley *et al*., 2010).

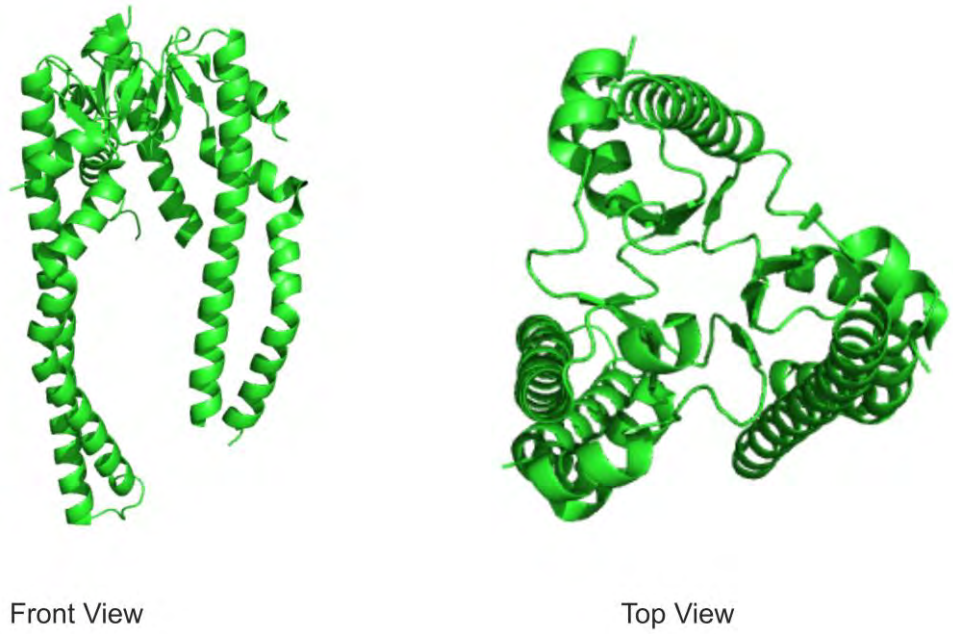


Figure 1.16 Trimeric structure of Skp. Skp (PDB: 1U2M), is a homo-trimer with extended α-helices that form a 'jellyfish' structure. The α-helices are thought to be involved in substrate binding.

1.9.3.3 DegP

DegP is a periplasmic ATP-independent serine protease. DegP has the major function of being a housekeeping protein in the cell envelope, promoting the refolding and proteolysis of unfolded or aggregated proteins (Meltzer *et al*., 2009). DegP also plays an important role in coping with extracytoplasmic stresses and is an essential gene under high temperature conditions (Subrini and Betton, 2009). DegP has also been shown to be specifically induced by the presence of precursor OMPs in the periplasm and responds to those proteins that have fallen off the route to OM biogenesis degrading them in a process that requires oligomerisation to form molecular cages (Krojer *et al*., 2010) (**Figure 1.17**).

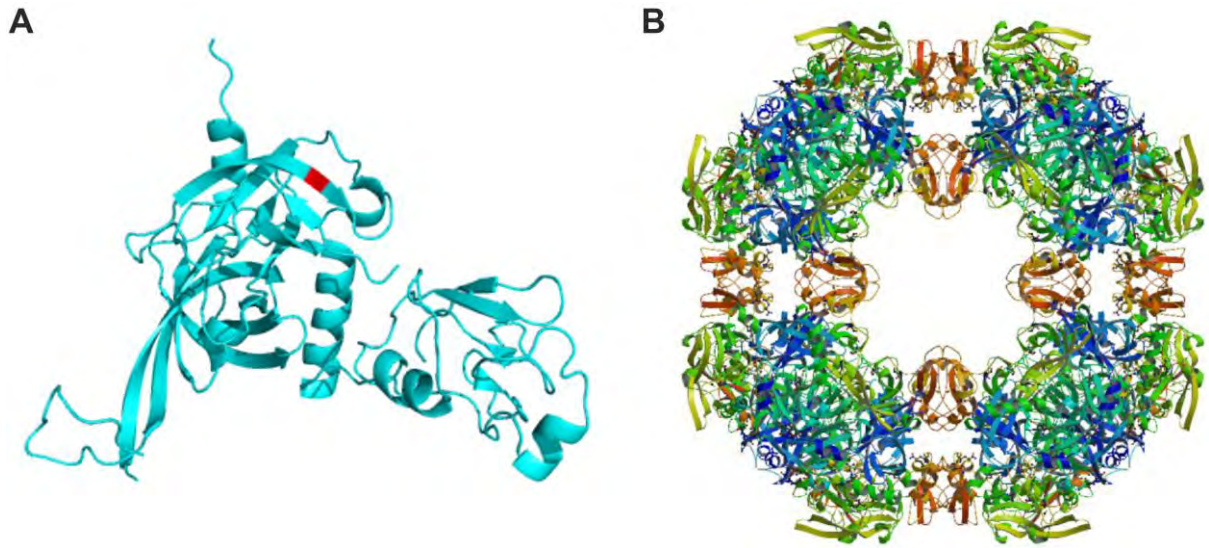


Figure 1.17 Structure of DegP. **A** Monomeric structure of DegP (PDB: 1KY9), with S210 coloured in red. **B** Crystal structure of the DegP 24mer (PDB: 3CS0).

DegP also contains chaperone activity alongside its protease ability, even when the active serine is mutated (Subrini and Betton, 2009). Protease activity is thought to predominate at higher temperatures and chaperone activity at lower temperatures, which is monitored by a temperature dependent conformational switch (Krojer *et al*., 2008). A role for DegP as a chaperone in OMP biogenesis was determined by a simultaneous deletion of *surA* and *degP*, which produced a synthetic lethal phenotype. This result suggested the presence of a secondary OMP chaperone pathway involving Skp and DegP (Sklar *et al*., 2007b) and SurA being the major pathway.

The exact role that DegP contributes to the OMP assembly pathway is not apparently clear. Overproduction of a protease deficient mutant of DegP ($DegP^{S210A}$) has shown to suppress the lethal effects of expression of unfolded OMPs (CastilloKeller and Misra, 2003, Misra *et al*., 2000), but this mutant can only bind

and sequester these misfolded intermediates in the periplasm. The return of these intermediates back to the OMP folding pathway has not been observed (Misra *et al*., 2000). This data suggests that DegP may function as a dead-end pathway, which degrades misfolded OMPs in the periplasm that have fallen off the OMP assembly pathway.

1.9.3.4 Other periplasmic chaperones

It has been suggested that there are other periplasmic proteins that may have a chaperone role involved with OM biogenesis, but may not be part of the major pathways involving SurA, DegP or Skp that have been discussed previously. Although they are not as essential, it is worth briefly discussing them.

FkpA is a dimeric periplasmic chaperone that also exhibits PPIase activity (**Figure 1.18**) (Horne *et al*., 1997, Saul *et al*., 2004). Like SurA, FkpA has chaperone activity that is independent of its PPIase domains (Bothmann and Pluckthun, 2000, Ramm and Pluckthun, 2000, Arie *et al*., 2001, Hullmann *et al*., 2008). FkpA has been implicated in a variety of functions for cell invasion and virulence in Gram-negative bacteria (Helbig *et al*., 2003, Horne *et al*., 1997, Moro *et al*., 1995), although the relative contribution of FkpA to these processes is still under question (Humphreys *et al*., 2003). FkpA has been demonstrated to be essential for the toxicity of colicin M, which requires the PPIase activity (Helbig *et al*., 2010) and is needed for the maturation of the autotransporter EspP (Ruiz-Perez *et al*., 2011). Although the chaperone and PPIase activities of FkpA are important for these functions, no

evidence is currently present that demonstrates the direct involvement of FkpA in the OMP biogenesis pathway.

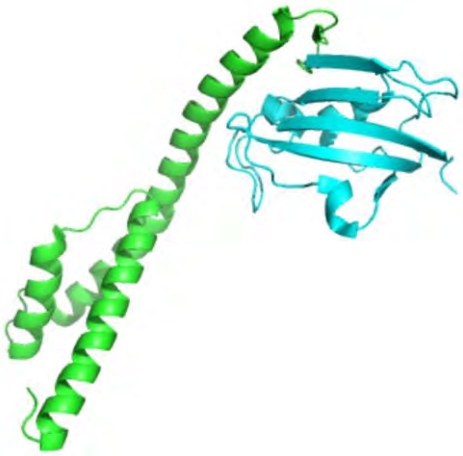


Figure 1.18 Monomeric structure of FkpA. FkpA (PDB: 1Q6U) also contains a PPIase domain (cyan), but its chaperone activity is independent on this domain.

PpiD is another periplasmic chaperone that is thought to contribute to OMP biogenesis. Overexpression of PpiD can suppress the OM defects associated with a Δ*surA* strain and that a double deletion of *ppiD* and *surA* yields a synthetic lethal phenotype (Dartigalongue and Raina, 1998). Evidence suggests that PpiD may be a periplasmic chaperone, but its substrate may not be precursor OMPs (Justice *et al*., 2005, Matern *et al*., 2010, Weininger *et al*., 2009). PpiD may act as a "gate-keeper", interacting with emerging membrane proteins as they enter the periplasm (Antonoaea *et al*., 2008) and utilises its chaperone ability for this purpose (Ureta *et al*., 2007).

1.9.4 OM substrate/target recognition

OMPs are driven through the IM by the Sec translocon via their N-terminal signal sequence, which is then cleaved by the periplasmic protease signal peptidase I (SPase I) and releases them from the IM (Silhavy *et al*., 2010, Bos *et al*., 2007a, Paetzel *et al*., 2002). With OMPs, there is another signal sequence present at the C-terminus. Data suggests that the terminal amino acid is the crucial part of this signal (de Cock *et al*., 1997) and that phylogenetic analysis of OMPs show a conserved aromatic residue at this position (Struyve *et al*., 1991). Deletion or mutation of this aromatic residue can block assembly into the OM (Misra *et al*., 2000). However, it should be noted that these mutants are able to fold into bilayers *in vitro*, indicating that this conserved phenylalanine residue is present for signal recognition rather than folding competency (Jansen *et al*., 2000). The OMP substrates that are recognised by the Bam complex have a consensus C-terminal sequence of X-Z-X-Z-X-Z-Tyr-Z-Phe/Trp, where X is hydrophobic and Z is any amino acid (Struyve *et al*., 1991, Robert *et al*., 2006, de Cock *et al*., 1997, Lehr *et al*., 2010).

1.9.5 The Bam Complex

Once precursor OMPs have been guided by periplasmic chaperones, they are delivered to the Bam complex, which has the task of inserting β-barrels into the OM. This molecular complex has been shown to be essential for OMP biogenesis in every organism tested (Wu *et al*., 2005, Gentle *et al*., 2004, Anwari *et al*., 2010, Voulhoux *et al*., 2003, Arnold *et al*., 2010, Bullmann *et al*., 2010, Schleiff *et al*., 2011) as well as being heavily conserved, with structural and sequential homologues of the most conserved Bam complex member, BamA (Omp85/YaeT)

being identified in chloroplasts (Toc75), mitochondria (Sam50/Tob55) and in all diderm bacteria (Bullmann *et al*., 2010). This suggests that the Bam complex is involved in a fundamental biological process and that β-barrel insertion by Bam is a universal mechanism across phylogenies. It should be noted that there are some OMPs that have been shown to be able to assemble independently of the Bam complex, such as PulD (which lacks the conserved C-terminus aromatic residue) and Wza (Collins and Derrick, 2007), however it remains to be seen whether under normal conditions folding via the Bam complex occurs.

In *E. coli*, the Bam complex consists of 5 units; BamA, a β-barrel itself, and four accessory lipoproteins, BamBCDE, that protrude into the periplasm (Ruiz *et al*., 2005, Wu *et al*., 2005, Anwari *et al*., 2010, Gatsos *et al*., 2008, Kim *et al*., 2007, Sklar *et al*., 2007a, Volokhina *et al*., 2009). The complex is organised into two major subcomplexes, BamAB and BamACDE (Malinverni *et al*., 2006, Kim *et al*., 2007), with interactions with precursor OMPs being mediated by BamA.

A wealth of information is available on the individual components and how the complex may function. Each of the components will be discussed in turn and then how the complex functions.

1.9.5.1 BamA

BamA is an essential and nearly ubiquitous member of the Bam complex. It is an essential gene in *E. coli* and is predicted to form a β-barrel. Depletion studies have highlighted its importance, with a range of OM defects associated with these mutant strains. Although the actual structure of BamA is still not yet fully determined, there is a wealth of information from, bioinformatic, biochemical, electrophysical and genetic characterisation.

BamA possesses an N-terminal periplasmic domain and a predicted C-terminal β-barrel. Bioinformatic analysis predicted that the N-terminus was divided into 5 distinct domains and structural studies indicated that BamA contains 5 POTRA (polypeptide translocation associated) domains (labelled P1-P5) (Gentle *et al*., 2005, Sanchez-Pulido *et al*., 2003, Gatzeva-Topalova *et al*., 2008). These structures show that each POTRA domain displays a β-α-α-β-β topology as its characteristic fold. Despite the lack of sequence homology between the individual POTRA domains, the structures of the domains are homologous and can be superimposed with very little deviation (<1.80 Å RMSD) (**Figure 1.19**).

Figure 1.19 The N-terminal domain of BamA contains POTRA domains. Crystal structure of POTRA1-4 (PDB: 3EFC), with each POTRA domain exhibiting a fold containing 3 β-strands and 2 α-helices, despite little sequence homology between them. Solution structure analysis has revealed that there is flexibility between the domains.

There are some notable differences between P3 and the other POTRA domains despite the structural similarity. P3 contains a loop between the two α-helices that is significantly longer (10 extra residues) and that its second β-strand contains a "β-bulge", which exposes additional β-strands that could be used for binding (Gatzeva-Topalova *et al*., 2008, Kim *et al*., 2007).

Numerous structural studies have been performed on the POTRA domains with some contrasting results. A POTRA domain dimer has been crystallised, with the β-bulge of P3 binding to a short segment of P5 as the dimerisation interface. The nature of this binding is β-augmentation, which is thought to be how POTRA domains bind to OMP substrates. β-augmentation is a mode of protein-protein interaction whereby a β-strand of one protein is added to an existing β-sheet of

another protein (Remaut and Waksman, 2006). Analysis by NMR has indeed shown the ability of POTRA domains to bind unfolded OMPs (Knowles *et al*., 2008).

From the two structural studies performed on this dimer, differing orientations of the β-augmentation interaction have been observed in both a parallel and anti-parallel manner (Kim *et al*., 2007, Gatzeva-Topalova *et al*., 2008). This may suggest that the POTRA domains may tolerate these orientations in order to accommodate the wide variety of substrates. Further analysis using a variety of methods such as SAXS, NMR and PELDOR have confirmed flexibility between the POTRA domains (Knowles *et al*., 2008, Ward *et al*., 2009) and have also suggested that dimerisation of the POTRA domains was due to crystal packing and is not representative of the wildtype conformation. It does seem feasible that the POTRA domains could re-orientate themselves to accommodate a variety of substrates.

The importance and roles of each of the POTRA domains remains enigmatic, although subsets of domains are essential and are most probably involved in a crucial step in OM biogenesis (Bos *et al*., 2007b). Experimental evidence suggests that they serve as docking sites for the Bam complex lipoproteins and have chaperone-like function for substrates (Kim *et al*., 2007). In *E. coli*, deletion of a single domain leads to reduced viability and impaired β-barrel assembly. Cells are able to tolerate the deletion of P1 and/or P2, but the removal of P3, P4 or P5 is lethal even if P1 or P2 is present (Kim *et al*., 2007). It should be noted that the nature of these POTRA domains are different between species. *Neisseria meningitidis* is able to tolerate the removal of P1-P4, with marginal effects on

viability (Bos *et al*., 2007b). However, P5 seems to be essential, even in distantly related Gram-negatives (Malinverni *et al*., 2006). Certain other homologues also lack five POTRA domains, such as cyanobacteria having only 3 POTRA domains (Arnold *et al*., 2010) and the mitochondrial Sam50, which only has a single POTRA domain (Kutik *et al*., 2008).

Although there have been great success in the structural characterisation of the N-terminal POTRA domains, the C-terminal β-barrel has yet to be solved. Studies have shown that BamA exhibits channel activity when reconstituted in lipid membranes and that the channel conductivity increases upon OMP binding (Stegmeier and Andersen, 2006). The structure of the distantly related two-partner secretion transporter FhaC has been solved (13.5% sequence identity) and it shows 16 strands forming a β-barrel that contains a long extracellular loop (L6) (**Figure 1.20**) (Clantin *et al*., 2007). This loop is a common motif amongst the Omp85 family members and is conserved in BamA and its sequence shows a highly conserved tetra motif (VRGY) (Clantin *et al*., 2007, Jacob-Dubuisson *et al*., 2009). The structure of FhaC indicates that this loop is able to extend into the lumen of the barrel. Mutations of this loop in FhaC stops it from secreting its passenger, the FHA adhesin (Clantin *et al*., 2007, Delattre *et al*., 2010). It is predicted that the L6 loop of BamA is much longer and could extend into the periplasm, which could possibly be involved in Bam function (Jacob-Dubuisson *et al*., 2009).

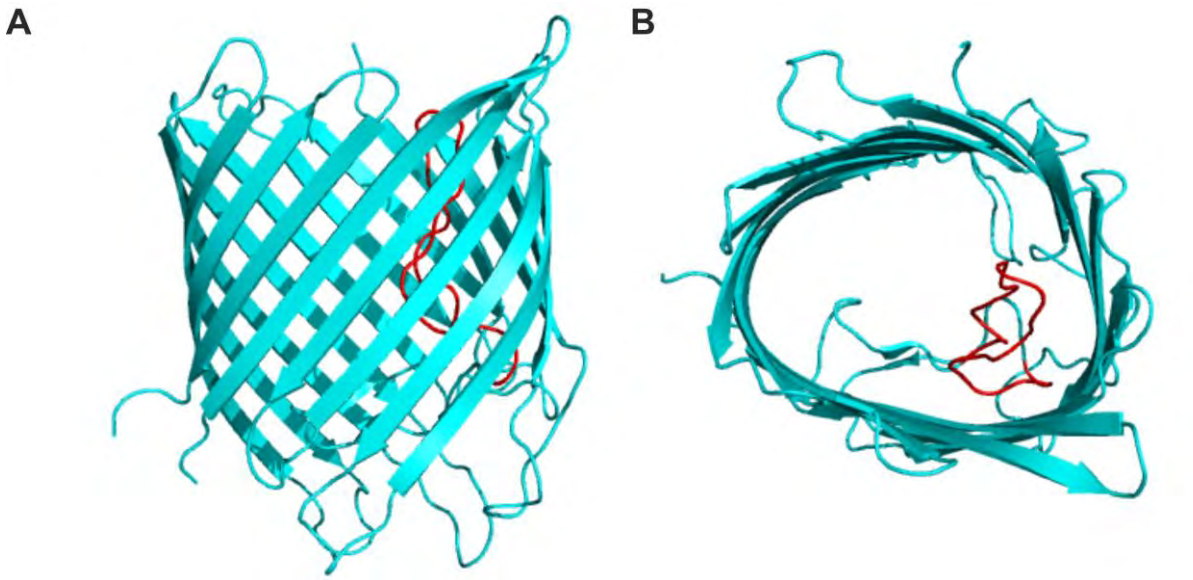


Figure 1.20 Crystal structure of FhaC. FhaC (PDB: 2QDZ) has 13.5% sequence identity to BamA and has a C-terminal β-barrel. A long extracellular loop L6 is present (red), which BamA also has, but is much longer.

1.9.5.2 BamB

BamB is an accessory lipoprotein that binds to BamA independently of BamC, D and E. Strains harbouring knockouts of *bamB* display a range of OM defects, such as hypersensitivity to antibiotics, increased stress responses and decreased OMP levels, indicating that BamB plays a role in OMP biogenesis (Ruiz *et al*., 2005, Wu *et al*., 2005, Charlson *et al*., 2006, Vuong *et al*., 2008).

Four independent crystal structures of BamB have been solved and all of them show very similar architecture (Vuong *et al*., 2008, Gatsos *et al*., 2008, Albrecht and Zeth, 2011, Heuck *et al*., 2011) (**Figure 1.21**). BamB has a β-propeller fold containing 8 blades. Each blade is comprised of four anti-parallel β-strands and the blades are joined together by interconnecting loops (ILs). A highly electronegative

groove has also been identified, which is thought to aid in protein-protein interactions (Kim and Paetzel, 2010).

Hydrophobic pockets have also been observed between the blades and are thought to possibly be an OMP binding site (Heuck *et al*., 2011). Although it has been reported that BamB increased OMP assembly efficiency (Hagan *et al*., 2010), there is no evidence of direct binding of OMPs to BamB.

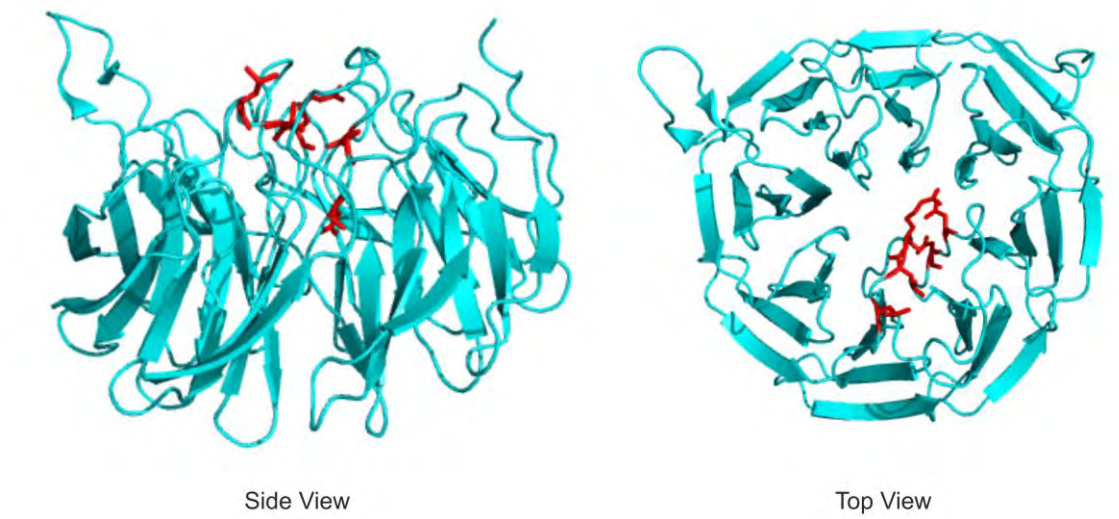


Figure 1.21 Structure of BamB. BamB (PDB: 3PIL) contains a β-propeller fold consisting of 8 blades. The BamA interaction site is shown in red.

1.9.5.3 BamC

BamC is another accessory lipoprotein of the Bam complex, although its function is not apparent. Knockouts of *bamC* display mild OM defects and have very slight decreases in OMP levels (Hagan and Kahne, 2011). In addition to this, BamC has also been shown to be surface exposed, indicating that it has regions present both in the periplasm and cell exterior (Webb *et al*., 2012).

Two structures of BamC have been published (Albrecht and Zeth, 2011, Warner *et al*., 2011), although it is difficult to gauge functional insights from them. Both of these structures report a disordered N-terminus (spanning the first ~70-100 residues) (**Figure 1.22**). Also present are two structurally homologous helix grip domains, despite the low sequence identity (12%), that are connected by a highly flexible linker (Kim *et al*., 2011b, Albrecht and Zeth, 2010, Knowles *et al*., 2009).

Despite knowing that BamC is a member of the Bam complex and that deletion of the gene causes some OM defects, very little is known about BamC. Structural analysis has not revealed much information. Further experimentation is needed to determine the function of BamC.

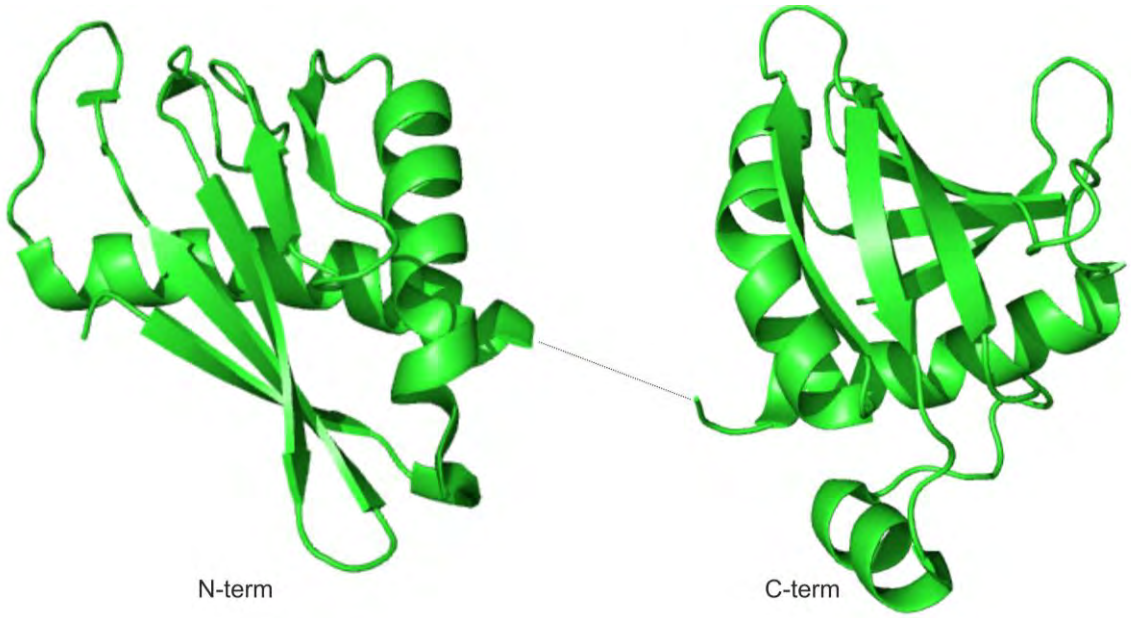


Figure 1.22 Structure of BamC. Crystal structures of the N and C-termini of BamC (PDB: 2LAE, 2LAF). The dashed line between the structures represents a predicted α-helix.

1.9.5.4 BamD

BamD is a lipoprotein accessory member of the Bam complex and is an essential gene in *E. coli*. Deletion of this gene yields a lethal phenotype and depletions of *bamD* are similar to *bamA* depletions, with OM defects, increased permeability of the OM and stalling of OMP assembly (Wu *et al*., 2006, Malinverni *et al*., 2006). The nearly identical effects of *bamA* and *bamD* depletions suggest that BamD plays an essential, but possibly overlapping role, in OMP biogenesis. BamD binds directly to BamA in a manner that is dependent on BamC and BamD (Malinverni *et al*., 2006, Hagan *et al*., 2010, Kim *et al*., 2007, Sklar *et al*., 2007a).

Two crystal structures of BamD have been reported from *E. coli* and *Rhodothermus marinus* (**Figure 1.23**) (Albrecht and Zeth, 2011, Sandoval *et al*., 2011). Both of these structures show that BamD is organised into a long superhelical structure and is composed of multiple tetratricopeptide (TPR) repeats. These TPRs stack against each other and stabilise each repeat. A major difference between the structures is that in *E. coli*, the interaction surface of the terminal TPR is capped off by an α-helix, whereas in *R. marinus*, this helix is missing entirely (Albrecht and Zeth, 2011, Sandoval *et al*., 2011). It should be noted that the C-terminal region of BamD is critical for function, as truncations of this region, lead to the loss of the associations between BamD and other members of the Bam complex (Malinverni *et al*., 2006).

Figure 1.23 Structure of BamD from *E. coli*. BamD (2YHC) contains long superhelical structure with multiple TPRs.

Because of this structure, BamD is predicted to serve as a protein scaffold (Sandoval *et al*., 2011). Structural homologues that have roles as chaperone receptors have been shown to bind to substrates by virtue of their C-terminal region into a conserved binding pocket (Kriechbaumer *et al*., 2011). As the C-terminus of unfolded OMPs is needed for assembly (de Cock *et al*., 1997), BamD could possibly recognise and bind precursor OMPs in the periplasm (Sandoval *et al*., 2011, Albrecht and Zeth, 2011).

1.9.5.5 BamE

BamE is a non-essential lipoprotein that is a member of the Bam complex (Sklar *et al*., 2007a). Like *bamC*, deletions of *bamE* yield mild defects, such as sensitivity to antibiotics and slightly decreased OMP levels.

Several groups have reported the structure of BamE using both NMR and crystallography (Albrecht and Zeth, 2010, Kim *et al*., 2011c, Knowles *et al*., 2011).

Overall, the structure of BamE consists of an anti-parallel β-sheet packed against a pair of α-helices with an α-α-β-β-β topology (**Figure 1.24**).

BamE is thought to stabilise the interaction between BamD and BamA with the aid of BamC (Malinverni *et al*., 2006, Sklar *et al*., 2007a). Although the exact functional role BamE is not exactly known, BamE has been shown to bind to phosphatidylglycerol (PG) (Knowles *et al*., 2011) and possibly modulate the conformation of BamA (Rigel *et al*., 2011). Further discussion of BamE is present in Chapter 3, where the functional aspects of BamE are discussed from the Knowles *et al.* 2011 study.

Figure 1.24 Structure of BamE. NMR structure of BamE (2KM7), which contains a fold reminiscent of the POTRA domain fold.

1.9.6 Interactions of Bam complex members

Previous studies have indicated that the Bam complex is organised into two major subcomplexes, BamAB and BamACDE and that each of the subunits are essential for efficient OMP assembly (Wu *et al*., 2005, Gatsos *et al*., 2008, Malinverni *et al*., 2006, Anwari *et al*., 2010, Kim *et al*., 2007, Sklar *et al*., 2007a, Volokhina *et al*., 2009). The stoichiometry of the complex is thought to be a 1:1:1:1 ratio for BamA, B, C and D, although it is hard to determine how many subunits of BamE are present due to its small size (Hagan *et al*., 2010). However, *in vitro* reconstitution of the complex indicated that only one molecule of BamE was needed for activity of the full complex (Hagan *et al*., 2010, Gatsos *et al*., 2008). It should be noted that BamA tetramers have been observed (Robert *et al*., 2006). Due to this subcomplex organisation, it is possible that BamB and BamCDE perform separate steps in OMP assembly, or that they have discrete functions.

There is a wealth of information relating to the BamAB interaction. Mutational analysis has revealed five crucial residues required for BamB function. Mutations of these residues, that cluster to two adjacent β-blade interconnecting loops (IL4 and IL5), disrupt the physical association of BamB to BamA (Vuong *et al*., 2008). These mutations essentially yield a Δ*bamB* phenotype, although stable BamB is produced, suggesting that this interaction with BamA is needed for proper OMP assembly. Analysis of the BamB crystal suggests that these residues specify a BamA binding site (Albrecht and Zeth, 2011, Heuck *et al*., 2011). In addition to this, deletion of the propeller blade connected to IL5 abolishes BamB function (Ruiz *et al*., 2005).

A re-ordering of the residues in the β-bulge of P3 has been shown to disrupt the BamA-B association, suggesting that there is a BamB binding site at the edge of the P3 β-sheet. Simulated protein docking has also suggested the binding of BamB IL4 and Bam P3 β-bulge by β-augmentation (Noinaj *et al*., 2011), which agrees with the data that demonstrates that both IL4 and the β-bulge are needed for the interaction (Vuong *et al*., 2008, Kim *et al*., 2007, Gatzeva-Topalova *et al*., 2008). It should also be noted that this interaction may be maintained electrostatically as P3 has been shown to be electropositive and the surface of IL4 and IL5 to be electronegative (Noinaj *et al*., 2011). However, these are not the only BamA-B associations as this interaction requires most of the POTRA domains for stable association. Removal of any POTRA domain except for P1 disrupts the BamA-B interaction.

The interaction of BamA-CDE seems to be dependent on P5 although from the current data, it is difficult to determine the nature of the interactions between BamA and BamC,D and E (Kim *et al*., 2007, Hagan *et al*., 2010). Data suggests that BamC and BamE interact with the complex through BamD in a manner that requires the C-terminus of BamD, as truncations in this region compromise the BamA-D interaction (Sklar *et al*., 2007a, Malinverni *et al*., 2006).

It has been suggested that BamC and BamE stabilise the BamA-D interaction. A co-crystal structure of the unstructured region and N-terminal domain of BamC and full length BamD revealed that the conserved unstructured region is important for the formation of the BamCD subcomplex (Kim *et al*., 2011a). Additionally, the

unstructured region of BamC was shown to obscure the C-terminus of BamD, which is thought to bind to OMP substrates (Kim *et al*., 2011a). This data from this structure suggests the possibility of BamC playing a regulatory role in OMP assembly (Kim *et al*., 2011a, Sandoval *et al*., 2011, Albrecht and Zeth, 2010).

Genetic experiments can also give insight into Bam complex function. It has been shown that deletions of the individual non-essential lipoproteins yields OM defects (Hagan *et al*., 2011, Sklar *et al*., 2007a, Wu *et al*., 2005). However double knockout data can yield further information. Individual knockouts of *bamC* or *bamE* yield minor OM defects, however a double deletion of *bamC* and *bamE*, generates severe OM defects (Sklar *et al*., 2007a). Interestingly, a double *bamB* and *bamE* deletion is lethal (Sklar *et al*., 2007a). These genetic interactions may imply some functional redundancy between Bam complex components, but also highlight critical roles for the non-essential lipoproteins as part of the complex despite their individual dispensability.

Binding of the periplasmic chaperone SurA to BamA has been observed and is thought to be a transient interaction (Bennion *et al*., 2010). Although the nature of this interaction is not apparent, this data could suggest that it is involved in substrate handover from SurA to BamA, and this interaction is used as a docking site to facilitate this exchange. There is further biochemical evidence that suggests that P1 is crucial for this transient docking step (Bennion *et al*., 2010).

The SurA-BamA interaction seems to be independent of BamB, despite some functional overlap between SurA and BamB. Both SurA and BamB depletions have shown altered efficiency of assembly of a subset of OMPs (Ureta *et al*., 2007, Vertommen *et al*., 2009) and that BamB is important for the assembly of Bam substrates delivered by SurA (Hagan *et al*., 2010). Additionally, the simultaneous deletion of *bamB* and *surA* leads to a synthetic lethal phenotype (Ureta *et al*., 2007). Since the loss of BamB does not affect the SurA-BamA interaction, it seems unlikely that SurA binds directly to BamB. BamB may somehow promote the assembly of SurA substrates in a manner that is independent of binding (Noinaj *et al*., 2011). It should also be noted that simultaneous removal of *bamB* and *degP* is also synthetically lethal (Charlson *et al*., 2006). These results suggest that BamB is involved in the earlier steps of OMP assembly.

So far, interactions of Skp or DegP to the Bam complex have not yet been observed (Sklar *et al*., 2007a). This could be due to the transient nature of these possible interactions, or that Skp and DegP are able to deliver Bam substrates in a manner that does not require docking at the Bam complex.

1.9.7 Concluding remarks

Although the individual contributions and interactions of the components of the Bam complex have been discussed, the actual process of OMP assembly still remains enigmatic. It is known that the periplasmic chaperones deliver Bam complex substrates that may bind to either the POTRA domains by β-augmentation or by the C-terminal of BamD (Kim *et al*., 2007, Albrecht and Zeth, 2011, Sandoval *et al*.,

2011). BamB, BamC and BamE may have roles in regulation and generally increasing the efficiency of the process. Substrates could then be passed to the BamA β-barrel domain, which could act as a scaffold to aid the formation and assembly of β-barrels. During this process, BamC could act as a substrate binding regulator, by using its unstructured region to bind to BamD. BamB could provide substrate binding surfaces and BamE may recruit PG to increase folding efficiency.

The least understood aspect is the molecular nature of OMP assembly by the Bam complex. Data from *in vitro* studies suggest that the insertion of OMPs takes place in a concerted manner (Burgess *et al*., 2008, Kleinschmidt, 2003). Three different models of OMP assembly have been suggested.

The first model suggests that the substrate OMP is translocated through the β-barrel domain of BamA into the extracellular space in an unfolded form before being assembled into the OM. This model is based on structural studies of FhaC. FhaC has a pore diameter of ~3 Å, which is too small to accommodate unfolded polypeptides (Clantin *et al*., 2007). However, upon binding of the substrate (FHA), the pore diameter can increase to 16 Å (Clantin *et al*., 2007, Delattre *et al*., 2010), which could be enough to accommodate the substrate. A similar observation has been made for BamA, where its electrical conductivity has increased with substrate binding (Stegmeier and Andersen, 2006). It is feasible that this conformational change could allow translocation. A major caveat with this model is that it assumes that OMP folding and assembly would take place on the extracellular surface which could be an inefficient process without the direct aid of folding factors.

The second model suggests that unfolded OMPs use the outer wall of BamA as a template, whereby the substrate is inserted between the BamA-lipid interface. This could lead to secondary structure formation, aided by the template of BamA and ultimately leading to formation of β-barrels. This model also takes into account the possible BamA tetramer, where the substrate could be contained in the space between BamA subunits. This enclosed space could drive the formation of β-barrels, which would then be released laterally into the bilayer.

The third model suggests that substrates are able to enter the β-barrel domain of BamA and that BamA is able to fold them into β-barrels and release them laterally into the bilayer (Bos and Tommassen, 2004). The major problems with this model are that the channel of BamA is not large enough to house a folded OMP and that lateral insertion would require breaking the hydrogen bonds of the β-barrel domain of BamA, which is very costly energetically. An additional model has been suggested by (Kim *et al*., 2012) based on this template. Instead of the substrate folding within the β-barrel of BamA, the N and C-terminal β-strands of BamA (the two strands that hydrogen bond with each other to close the β-barrel), could act as folding templates of OMP substrates. Hydrogen bonds could form between these BamA terminal strands and the substrate, which leads to the formation of β-sheets in the substrate. This process is repeated which leads to the formation of more β-strands, until the β-barrel of the substrate is formed, which is then released laterally. This model requires the breakage of hydrogen bonds of the β-barrel of BamA, but

the formation of hydrogen bonds between BamA and the substrate β-strands by β-augmentation is much more energetically favourable.

Within the next forthcoming years, the molecular events leading to assembly will hopefully be solved. Questions such as the structure of BamA, the role its periplasmic loops and the assembly of the complex would fill major gaps in understanding OMP biogenesis.

1.10 Aims of this Study

The field of OM biogenesis is very diverse with several pathways contributing to overall maintenance. Although these pathways have been mostly deciphered, the molecular mechanisms leading to assembly and the functional relevance of the individual protein components have not been determined. This study is focussed on performing both a structural and functional analysis of two proteins that are involved in OM biogenesis.

The first aim of this project is to study the Bam complex member, BamE. Using NMR, the solution structure of BamE will be determined. Additionally, by exploiting the OM defects associated with Δ*bamE* cells a complementation assay will be produced which will be used to gauge BamE function. Finally, using a comprehensive mutagenesis strategy along with the complementation assay, critical residues required for BamE function can be identified.

The second part of this project focuses on YraP. *yraP* was initially identified as a gene which may have a role in OM biogenesis from a preliminary screen performed by the Henderson lab. Currently there is very little information on *yraP*. Our aim is therefore to solve the structure using NMR and characterise its function. Using a range of molecular biology, biophysical and bacteriological techniques, we aim to deduce precisely what role YraP plays in the field of OM biogenesis.

Chapter 2

Materials and Methods

2.1 Working concentrations of antibiotics used in this study

Ampicillin	100 µg/ml
Kanamycin	50 µg/ml
Vancomycin	100 µg/ml
Tetracycline	10 µg/ml

All antibiotics were made up to 1000 x their working concentrations in sterile distilled water (SDW), except for tetracycline which was made up in 100% methanol. Antibiotics were filtered through a 0.2 µm pore (Millipore) prior to use and stored at -20^{o}C until required.

2.2 Regaents used in this study

2.2.1 Media Recipes

Luria Bertani (LB) Broth

Per litre

10 g	Bactotryptone
5 g	Yeast extract
10 g	NaCl

15 g bacteriological agar (Agar No. 1) was added to the above recipe to make LB-Agar plates

M9 Minimal media (pH 7.2)

Per litre

4.75 g	Na_2HPO_4
6 g	KH_2PO_4
1 g	NaCl

Nutrient mix

1 ml	Thiamine (20 mg/ml)
400 µl	3 mM $FeCl_3$
2 ml	1 M $MgSO_4$
2 ml	50 mM $CaCl_2$

2 g Glucose
1 g NH_4Cl

All media was prepared in SDW and autoclaved at 121^{o}C, 15 psi for 20 minutes. With M9 media, the media was autoclaved prior to the addition of the nutrient mix using a 0.2 µm filter.

2.2.2 Reagents Used in this Study

Table 2.1 Reagents and chemical names used in this study

Chemical Name	Supplier
Ampicillin	Sigma-Aldrich
Kanamycin	Sigma-Aldrich
Vancomycin	Sigma-Aldrich
Tetracycline	Sigma-Aldrich
Bactotryptone	Oxoid
Yeast Extract	Oxoid
Agar No. 1	Oxoid
Salt (NaCl)	Sigma-Aldrich
Sodium dihydrogen phosphate	Sigma-Aldrich
Sodium phosphate dibasic	Sigma-Aldrich
Monopotassium phosphate	Sigma-Aldrich
Thiamine	Sigma-Aldrich
Iron(III) chloride	Sigma-Aldrich
Magnesium sulfate	Sigma-Aldrich
Calcium chloride	Sigma-Aldrich
D-Glucose	Sigma-Aldrich
^{13}C D-Glucose	Cambridge Isotope Laboratories
Ammonium chloride	Sigma-Aldrich
^{15}N Ammonium chloride	Cambridge Isotope Laboratories
Glycerol	Sigma-Aldrich
Phusion PCR Polymerase	New England Biolabs
NdeI Restriction Enzyme	New England Biolabs
EcoRI Restriction Enzyme	New England Biolabs
XhoI Restriction Enzyme	New England Biolabs
BsiWI Restriction Enzyme	New England Biolabs
Calf Intestinal Phosphotase	New England Biolabs

DNA Primers	Alta Biosciences
Agarose	Sigma-Aldrich
6x DNA Loading Dye	New England Biolabs
Hyperladder I	Bioline
Acetic Acid	Sigma-Aldrich
EDTA	Sigma-Aldrich
Tris Base (Trizma)	Sigma-Aldrich
SYBR Green	Promega
Laemmli Buffer	Sigma-Aldrich
4-12% Bis Tris Pre-Cast Gels	Bio-Rad
Precision Plus Protein Markers	Bio-Rad
XT-MES	Bio-Rad
Instant Blue Protein Stain	Gentaur
IPTG	Sigma-Aldrich
Imidazole	Sigma-Aldrich
MWCO Spin Columns	GE Healthcare
EDTA-free σ-Complete Protease Inhibitor	Roche
Potassium chloride	Sigma-Aldrich
Tween-20	Sigma-Aldrich
Anti-YraP antibody	Eurogentec
Anti-rabbit HRP conjugated antibody	GE Healthcare
EZ-ECL Development Solution	GeneFlow
Amersham Hyperfilm	GE Healthcare
BCIP®/NBT-Purple Liquid Substrate	Sigma-Aldrich
Deuterium oxide	Sigma-Aldrich
Phosphatidylcholine (PC)	Avanti
Phosphatidylglycerol (PG)	Avanti
Phosphatidylehtnaolamine (PG)	Avanti
Lysozyme	Sigma-Aldrich
Triton X-100	Sigma-Aldrich
Poly-L-Lysine	Sigma-Aldrich
Goat Anti-Rabbit Alexa Fluor® 448 Antibody	Invitrogen
Triethanolamine	Sigma-Aldrich
Sucrose	Sigma-Aldrich
Dithiothreitol	Sigma-Aldrich
Zwittergent ZW3-14	Calbiochem
Dithiobis(succinimidyl) propionate (DSP)	Sigma-Aldrich
Formaldehyde	Sigma-Aldrich
Protein A-Sepharose	Sigma-Aldrich
n-Dodecyl β-D-Maltopyranoside (DDM)	Sigma-Aldrich
Silver Stain Plus Kit	Bio-Rad
Potassium ferricyanide	Sigma-Aldrich

Sodium thiosulphate	Sigma-Aldrich
Ammonium bicarbonate	Sigma-Aldrich
Iodoacetamide	Sigma-Aldrich
Acetonitrile	Sigma-Aldrich
Trypsin	Promega
Formic acid	Sigma-Aldrich
Proteinase K	Qiagen
ProteoSilver™ Plus Silver Stain Kit	Sigma-Aldrich
CHAPS detergent	Sigma-Aldrich
Trifluoroacetic acid	Sigma-Aldrich

2.3 Bacterial strains

The bacterial strains used in this study are listed in **Table 2.2**. All strains were stored in 15% LB glycerol at -20^{o}C.

Table 2.2 Strains used in this study

Strain	Relevant characteristics	Source
BW25113	*lacI*^{q}*rrnB*$_{T14}$Δ*lacZ*$_{WJ16}$*hsdR514*Δ*rhaBAD*$_{LD78}$	(Baba *et al*., 2006)
Δ*bamE*	BW25113 derivative with gene substituted with kanamycin resistance cassette, *bamE*::*aph*	(Baba *et al*., 2006)
Δ*yraP*	BW25113 derivative with gene substituted with kanamycin resistance cassette, *yraP*::*aph*	(Baba *et al*., 2006)
SL1344	Parental strain	(Wrey *et al .,* 1978)
SL1344 Δ*bamE*	SL1344 derivative with gene substituted with kanamycin resistance cassette, *bamE*::*aph*	Dr. D. Squire
SL1344 Δ*yraP*	SL1344 derivative with gene substituted with kanamycin resistance cassette, *yraP*::*aph*	Dr. D. Squire
XL-1 Blue	*recA*1*endA*1gyrA*96*thi-*1*hsd*R17*sup*E*44*rel*A1*lac*	Stratagene
DH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Hanahan *et al*., 1985)
BL21 (DE3)	F– *omp*T *hsd*SB(rB–, mB–) *gal dcm* (DE3)	(Studier *et al.,* 1986)

2.4 Cell manipulations

2.4.1 Bacterial growths

All growths were performed in LB or M9 at 37^{o}C and 180 rpm for aerobic growth, with the exception of growth for protein expression (**Section 2.6**).

2.4.2 Generation of competent cells

BW25113, Δ*yraP* and Δ*bamE* cells were streaked onto selected antibiotic LB Agar plates to single colonies. A colony was grown in 5 ml LB antibiotic medium overnight. 1 ml of the overnight culture was used to inoculate 100 ml LB Kanamycin culture, which was grown at 37^{o}C, 180 rpm, to OD_{600} = 0.5. Cells were placed in ice for 10 minutes and then centrifuged at 6000 x *g* for 10 minutes at 4^{o}C to recover them. The pellet was resuspended in 50 ml of ice cold 0.1 M Calcium chloride ($CaCl_2$) and kept on ice for 90 minutes. Cells were then centrifuged again at 6000 x *g* for 10 minutes and resuspended in 4 ml of ice cold 0.1 M $CaCl_2$ with 15 % glycerol. Cells were snap frozen and stored at -80^{o}C.

2.4.3 Transformation by heat shock

Transformations were performed according to (Hanahan, 1983)50 µl of competent cells were kept on ice for 5 min. 50-200 ng of plasmid was added to the cells and kept in ice for a further 30 min. Cells were then heat shocked at 42^{o}C for 30 s and then placed on ice for 5 min. 0.5 ml LB was added to the cells and were grown at 37^{o}C 180 rpm for 1 hour. 80 µl cells were then plated onto LB-Agar antibiotic plate to select for transformants and incubated overnight at 37^{o}C.

2.5 DNA manipulations

2.5.1 Polymerase chain reactions (PCR)

PCRs were carried out using Phusion® High Fidelity DNA polymerase (New England Biolabs) according to the manufacturer's recommendations. Between 3-5 ng of plasmid or genomic DNA was used as a template, with 0.5 µM primers 200 µM dNTPs for a 50 µl reaction. Reactions were made up to volume with SDW.

A standard thermal cycling protocol was carried out, based on (Chiang *et al*., 1993) with an initial denaturing step at 98^{o}C for 30 s, and 30 cycles of 98^{o}C for 30 s (denaturing), X^{o}C for 30 s (annealing) and 72^{o}C/[1 min/kb](extension) and a final extension step at 72^{o}C for 10 mins. X is the annealing temperature of the primers. **Section 2.5.4** shows the list of primers used in this study.

2.5.2 Gel electrophoresis of DNA

DNA samples were prepared by mixing in a 5:1 ratio in 6x DNA loading dye (NEB) and analysed using 1% agarose dissolved in 1x Tris-acetate-EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.4). Gels were ran in 1x TAE buffer at 80-120 V (constant voltage) for 30-60 mins (Norton *et al*., 1986). Hyperladder I DNA maker (Bioline) was used to determine DNA size. Gels were visualised with SYBR green and a UV transilluminator with the gel-documentation system (Bio-Rad).

2.5.3 Cloning of genes

Primers were designed to amplify required genes from K-12 genomic DNA. Generally, forward primers contained an *NdeI* restriction site and reverse primers with an *EcoRI* site for full length proteins and *XhoI* site for His-tagged proteins. Any other restriction enzyme sites used are noted.

PCRs were performed and analysed by gel electrophoresis as described above. Bands corresponding to the correct size of amplification were excised and purified using a Qiagen Gel Extraction Kit.

Vector and purified insert were digested using the appropriate NEB restriction enzymes according to the manufacturer's recommendation. Digested vector was then heat inactivated by incubation at 65^{o}C and then treated with Calf Intestinal Phosphatase (CIP) for 1 hour at 37^{o}C. Digested products were purified using a Qiagen PCR cleanup kit.

Ligations were performed using an Invitrogen T4 Ligation Kit, with a 3:1 insert to vector ratio, at 4^{o}C overnight. A vector only ligation was performed as a negative control. Ligations were transformed into DH5α or XL1-Blue competent cells using heat shock transformation and plated onto LB Agar antibiotic plates. Colonies were screened using colony PCR and were then grown O/N in LB antibiotic cultures.

Plasmids were extracted using a QiagenMiniprep Kit and sequenced using an ABI 3730 Sequencer (Functional Genomics, University of Birmingham).

2.5.4 Primers used in this study

2.5.4.1 Primers used for cloning *bamE*

Table 2.3 Primers used for cloning *bamE*

Name	Sequence	Tm (^{o}C)
bamE NdeI For	CTGCTCGTACATATGCGCTGTAAAACGCTGAC	65
bamE EcoRI Rev	CCGGAATTCTTAGTTACCACTCAGCGC	62
bamE XhoI Rev	CCGCTCGAGAGTTACCACTCAGCGCAGG	68

2.5.4.2 Primers used for cloning *yraP*

Table 2.4 Primers used for cloning *yraP*

Name	Sequence	Tm (^{o}C)
yraP NdeI For	GGGAATTCCATATGAAGGCATTATCGCCAATCG	63
yraP EcoRI Rev	CCGGAATTCCTATTTAATAAACGTAAACGC	64
yraP XhoI Rev	CCGCTCGAGTTTAATAAACGTAAACGCCGT	64

2.5.4.3 Primers used to confirm the Δ*yraP* strain

Table 2.5 Primers used to confirm the Δ*yraP* strain

Primer Name	Sequence	Comments
yraP For	GGCATTGACCGGCTATGACG	Anneals 200 bp upstream of start
yraP Rev	TCGGATGCGGCGTAAACGCC	Anneals 100 bp downstream of end
Kan For	CCTGCAAAGTAAACTGGATG	Internal primer in kanamycin cassette of *aph* gene
Kan Rev	CATGCTCTTCGTGCACATCA	Internal primer in kanamycin cassette of *aph* gene

2.5.4.4 Primers used for YraP mutagenesis

Table 2.6 Primers used for YraP mutagenesis

Name	Primer sequence	Comments
C19A_For	CCGCGCTGCTGTTGCAAGGTGCTGTTGCCGCTG	Substitution to alanine
C19A_Rev	CAGCGGCAACAGCACCTTGCAACAGCAGCGCGG	Substitution to alanine
R112*_For	ACGGTGCCAACGAAGTGTATAACGAGATTTAGTAGG GCCAGCCGATTG	Substitutes residue for 2 stop codons
R112*_Rev	CAATCGGCTGGCCCTACTAAATCTCGTTATACACTTC GTTGGCACCGT	Substitutes residue for 2 stop codons
Q113*_For	GAAGTGTATAACGAGATTCGTTAGTAGCAGCCGATT GGTCTGGGCGAA	Substitutes residue for 2 stop codons
Q113*_Rev	TTCGCCCAGACCAATCGGCTGCTACTAACGAATCTC GTTATACACTTC	Substitutes residue for 2 stop codons
G114*_For	GTGTATAACGAGATTCGTCAGTAGTAGCCGATTGGT CTGGGCGAAG	Substitutes residue for 2 stop codons
G114*_Rev	CTTCGCCCAGACCAATCGGCTACTACTGACGAATCT CGTTATACAC	Substitutes residue for 2 stop codons
Q115*_For	ACGAGATTCGTCAGGGCTAGTAGATTGGTCTGGGCG AAG	Substitutes residue for 2 stop codons
Q115*_Rev	CTTCGCCCAGACCAATCTACTAGCCCTGACGAATCT CGT	Substitutes residue for 2 stop codons
P116*_For	AACGAGATTCGTCAGGGCCAGTAGTAGGGTCTGGG CGAAGCATCTAAC	Substitutes residue for 2 stop codons
P116*_Rev	GTTAGATGCTTCGCCCAGACCCTACTACTGGCCCTG ACGAATCTCGTT	Substitutes residue for 2 stop codons
I117*_For	ACGAGATTCGTCAGGGCCAGCCGTAGTAGCTGGGC GAAGCATCTAACGATAC	Substitutes residue for 2 stop codons
I117*_Rev	GTATCGTTAGATGCTTCGCCCAGCTACTACGGCTGG CCCTGACGAATCTCGT	Substitutes residue for 2 stop codons
G118*_For	CGAGATTCGTCAGGGCCAGCCGATTTAGTAGGGCG AAGCATCTAACGAT	Substitutes residue for 2 stop codons
G118*_Rev	ATCGTTAGATGCTTCGCCCTACTAAATCGGCTGGCC CTGACGAATCTCG	Substitutes residue for 2

		stop codons
L119*_For	CGTCAGGGCCAGCCGATTGGTTAGTAGGAAGCATCTAACGATACGTGG	Substitutes residue for 2 stop codons
L119*_Rev	CCACGTATCGTTAGATGCTTCCTACTAACCAATCGGCTGGCCCTGACG	Substitutes residue for 2 stop codons
V142*_For	AGCTCTTAACCAGCGACCTGTAGTAATCGTCCAACGTGAAAGTG	Substitutes residue for 2 stop codons
V142*_Rev	CACTTTCACGTTGGACGATTACTACAGGTCGCTGGTTAAGAGCT	Substitutes residue for 2 stop codons
K143*_For	TTAACCAGCGACCTGGTGTAATAGTCCAACGTGAAAGTGAC	Substitutes residue for 2 stop codons
K143*_Rev	GTCACTTTCACGTTGGACTATTACACCAGGTCGCTGGTTAA	Substitutes residue for 2 stop codons
S144*_For	AGCTCTTAACCAGCGACCTGGTGAAATAGTAGAACGTGAAAGTGACC	Substitutes residue for 2 stop codons
S144*_Rev	GGTCACTTTCACGTTCTACTATTTCACCAGGTCGCTGGTTAAGAGCT	Substitutes residue for 2 stop codons
S145*_For	CCAGCGACCTGGTGAAATCGTAGTAGGTGAAAGTGACCACCGAAAAC	Substitutes residue for 2 stop codons
S145*_Rev	GTTTTCGGTGGTCACTTTCACCTACTACGATTTCACCAGGTCGCTGG	Substitutes residue for 2 stop codons
N146*_For	GCGACCTGGTGAAATCGTCCTAGTAGAAAGTGACCACCGAAAACG	Substitutes residue for 2 stop codons
N146*_Rev	CGTTTTCGGTGGTCACTTTCTACTAGGACGATTTCACCAGGTCGC	Substitutes residue for 2 stop codons
G83V_For	CAAAGTGCTGCTGGTTGTGCAGTCACCAAATGCTG	Individual substitution
G83V_Rev	CAGCATTTGGTGACTGCACAACCAGCAGCACTTTG	Individual substitution
G160V_For	GAAGTGTTCCTGATGGTGCTGGTGACTGAACGT	Individual substitution
G160V_Rev	ACGTTCAGTCACCAGCACCATCAGGAACACTTC	Individual substitution
W127A_For	GAAGCATCTAACGATACGGCGATCACCACCAAAGTGCG	Individual substitution
W127A_Rev	CGCACTTTGGTGGTGATCGCCGTATCGTTAGATGCTTC	Individual substitution
I128A_For	GCATCTAACGATACGTGGGCCACCACCAAAGTGCGTTC	Individual substitution
I128A_Rev	GAACGCACTTTGGTGGTGGCCCACGTATCGTTAGAT	Individual

	GC	substitution
K131A_For	TACGTGGATCACCACCGCAGTGCGTTCGCAGCTC	Individual substitution
K131A_Rev	GAGCTGCGAACGCACTGCGGTGGTGATCCACGTA	Individual substitution
R133A_For	GGATCACCACCAAAGTGGCTTCGCAGCTCTTAACCA	Individual substitution
R133A_Rev	TGGTTAAGAGCTGCGAAGCCACTTTGGTGGTGATCC	Individual substitution
Q135A_For	CACCAAAGTGCGTTCGGCGCTCTTAACCAGCGAC	Individual substitution
Q135A_Rev	GTCGCTGGTTAAGAGCGCCGAACGCACTTTGGTG	Individual substitution
L137A_For	GTGCGTTCGCAGCTCGCAACCAGCGACCTGGT	Individual substitution
L137A_Rev	ACCAGGTCGCTGGTTGCGAGCTGCGAACGCAC	Individual substitution
S144A_For	TTAACCAGCGACCTGGTGAAAGCGTCCAACGTG	Individual substitution
S144A_Rev	CACGTTGGACGCTTTCACCAGGTCGCTGGTTAA	Individual substitution
W127-K131 Ala For	GGCGAAGCATCTAACGATACGGCGGCCACCACCGC AGTGCGTTCGCAGCTCTTAAC	Multiple mutatations in PG binding helix
W127-K131 Ala Rev	GTTAAGAGCTGCGAACGCACTGCGGTGGTGGCCGC CGTATCGTTAGATGCTTCGCC	Multiple mutatations in PG binding helix
R133-L137 Ala For	GTGGATCACCACCAAAGTGGCTTCGGCGCTCGCAA CCAGCGACCTGGTGAAA	Multiple mutatations in PG binding helix
R133-L137 Ala Rev	TTTCACCAGGTCGCTGGTTGCGAGCGCCGAAGCCA CTTTGGTGGTGATCCAC	Multiple mutatations in PG binding helix
W127-K131 Charged For	GGGCGAAGCATCTAACGATACGCAGGAGACCACCG AGGTGCGTTCGCAGCTCTTAACCA	Multiple mutatations in PG binding helix
W127-K131 Charged Rev	TGGTTAAGAGCTGCGAACGCACCTCGGTGGTCTCCT GCGTATCGTTAGATGCTTCGCCC	Multiple mutatations in PG binding helix
R133-L137 Charged For	TACGTGGATCACCACCAAAGTGGAGTCGGAGCTCGA TACCAGCGACCTGGTGAAATCGT	Multiple mutatations in PG binding helix
R133-L137 Charged Rev	ACGATTTCACCAGGTCGCTGGTATCGAGCTCCGACT CCACTTTGGTGGTGATCCACGTA	Multiple mutatations in PG binding helix
W127A-Q135 Ala For	TGGGCGAAGCATCTAACGATACGGCGGCCACCACC GCAGTGGCTTCGGCGCTCGCAACCAGCGACCTGGT GAAATCGT	Multiple mutatations in PG binding helix
W127A-Q135 Ala Rev	ACGATTTCACCAGGTCGCTGGTTGCGAGCGCCGAA GCCACTGCGGTGGTGGCCGCCGTATCGTTAGATGC TTCGCCCA	Multiple mutatations in PG binding helix

W127-Q135 Charged For	TGGTCTGGGCGAAGCATCTAACGATACGCAGGAGA CCACCGAGGTGGAGTCGGAGCTCGAGACCAGCGAC CTGGTGAAATCGTCCAACG	Multiple mutatations in PG binding helix
W127-Q135 Charged Rev	CGTTGGACGATTTCACCAGGTCGCTGGTCTCGAGCT CCGACTCCACCTCGGTGGTCTCCTGCGTATCGTTAG ATGCTTCGCCCAGACCA	Multiple mutatations in PG binding helix

2.6 Protein experimentation

2.6.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples were prepared by mixing in a 1:1 ratio to 2x Laemmli (Sigma-Aldrich) and boiled for 10 min. Criterion 4-12% Bis-Tris pre-cast gels (BioRad) were used. Gels were run alongside Precision Plus protein standards (BioRad), at 180 V for 40 min in 1x XT-MES gel running buffer (BioRad) (Webber *et al*., 1972). Protein bands were detected by staining gels with InstantBlue (Gentaur).

2.6.2 Protein expression and purification

Protein purifications in this study were performed with constructs that were designed with C-terminal 6x His Tag. Vectors used were derived from pET20b or pET26b (Invitrogen), either manipulated in house or synthesised (GenScript).

Vectors were transformed into BL21 (DE3) cells and colonies were picked and grown overnight in LB antibiotic (50 ml). 20 ml of the overnight culture was transferred aseptically to 2 litres of LB antibiotic. This was grown at 37^{o}C, 180 rpm until an OD_{600} = 0.4 was reached. The cells were then grown at 18^{o}C. Once grown to OD_{600} = 0.6, protein overexpression was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and left to grow overnight.

Cells were harvested by centrifugation at 5000 x *g* and resuspended in 25 ml of wash buffer (50 mM Sodium Phosphate buffer, 500 mM NaCl, 50 mM Imidazole with EDTA-free σ Complete Protease Inhibitor [Roche], pH 7.2). Cells were lysed by 8 passages through an Emulsiflex (Avastin) and the soluble fraction was recovered by centrifugation at 75000 x *g* for 1 hour. The soluble fraction was filtered through a 0.45 micron filter.

A 5 ml His-Trap Nickel affinity column (GE Healthcare) was equilibrated with 25 ml of wash buffer, using a pump. The soluble fraction was added to the column, which was then washed with 25 ml of wash buffer. The protein was eluted using elution buffer (50 mM Sodium Phosphate buffer, 500 mM NaCl, 500 mM Imidazole with protease inhibitor, pH 7.2) and 1 ml fractions were collected.

The eluate was analysed by SDS-PAGE. Fractions containing the protein were pooled and concentrated using a 10 kDa MWCO centrifugation filter (Amicon). Size exclusion chromatography was performed using a Superdex S75 PG (340 ml) column, with Gel Filtration buffer (50 mM Sodium Phosphate buffer, 300 mM NaCl, protease inhibitor, pH 7.2). Typically a 2.5 ml/min flow rate, with 1.2 column volumes of used for equilibration and elution of the column and 4 ml fractions collected in the elution. Fractions containing the protein were pooled and concentrated using a centrifugation filter.

For NMR studies, the same purification methodology was used, except that cells were grown in M9 minimal media, with ^{13}C-glucose and ^{15}N-ammonium sulphate. After size exclusion chromatography, the protein was buffer exchanged 3 times into NMR buffer (50 mM Sodium phosphate, 50 mM NaCl at pH 6.0), using a centrifugation filter.

2.6.3 Western Blotting

Either cells normalised according to OD_{600} measurements, or proteins were resuspended in 2 x Laemmli buffer, boiled for 10 minutes and SDS-PAGE was performed.

Proteins were transferred onto either PVDF or nitrocellulose membranes (Towbin *et al.*, 1979) using Bio Rad Wester Blot Transfer Apparatus. Membranes were washed twice with 0.05% Phosphate buffered saline with Tween-20 (PBS-Tween) (140 mM NaCl, 27 mM KCl, 10 mM Na2HPO4.2H2O, 2 mM KH2PO4, pH 7.2, 0.05% Tween-20 [v/v]) and blocked with 5% Milk PBS-Tween by shaking incubation at room temperature for 1 hour. Membranes were washed twice with PBS-Tween and incubated with 1:5000 dilution of primary antibody for 4 hours at room temperature. Membranes were again washed twice with PBS-Tween and incubated with either 1:10000 dilution of Anti-Rabbit alkaline phosphatase conjugated antibody (GE Healthcare) or Anti-Rabbit HRP conjugated antibody (GE Healthcare). Membranes were finally washed twice with PBS-Tween and blots were either developed with EZ-ECL kit (GeneFlow) and then developed with X-ray film (if HRP antibody was used) or with BCIP®/NBT-Purple Liquid Substrate (Sigma).

2.7 NMR Experiments

2.7.1 General overview

NMR experiments were carried out at 298 K on either a Varian Inova 600, 800 or 900 MHz spectrometers equipped with a triple-resonance 5 mm z-PFG cryogenic probes with enhanced ^{1}H, ^{13}C and ^{15}N sensitivity and *z*-axis pulse-filed gradients. Pulse sequences were loaded Biopack (Varian) set of pulse sequences. Spectrometers were calibrated to their absolute frequencies.

2.7.2 1D NMR experiments

Unlabelled YraP at a concentration of 0.4 mM in 50 mM Sodium phosphate, 50 mM NaCl, pH 6.0 (90% H_2O, 10% D_2O), was analysed in an Agilent 600 MHz direct drive spectrometer equipped with cryoprobe. The standard Biopack (Agilent) 1D experiment was used.

2.7.3 HSQC

^{15}N-labelled YraP at a concentration of 0.4 mM in 50 mM Sodium phosphate, 50 mM NaCl, pH 6.0 (90% H_2O, 10% D_2O), was analysed using an Agilent 800 MHz direct drive spectrometer equipped with cryoprobe. A standard Biopack HSQC experiment was performed (Agilent), (Bodenhausen and Ruben 1980), with 1024 and 128 points and spectral widths of 8000 Hz and 2200 Hz in the ^{1}H and ^{15}N dimensions respectively.

2.7.4 3D experiments

3D backbone experiments [HNCA, HN(CO)CA, HNCACB, CB(CACO)NH, HNCO and HN(CA)CO] were performed on ^{13}C, ^{15}N-labelled YraP sample at 1 mM in in 50 mM Sodium phosphate, 50 mM NaCl, pH 6.0 (90% H_2O, 10% D_2O), using an Agilent 600 MHz direct drive spectrometer equipped with cryoprobe (Kay *et al*., 1990, Farmer *et al*., 1992, Grzesiek *et al*., 1992).

All backbone experiments were performed with 842 points and spectral widths of 6897 Hz and 1899 Hz for the ^{1}H and ^{15}N dimensions respectively. HNCA and HN(CO)CA experiments were performed with spectral widths of 4125 Hz, HNCACB and CB(CACO)NH experiments with spectral widths of 8867 Hz and HNCO and HN(CA)CO experiments with spectral widths of 1971 Hz for the ^{13}C dimensions.

NOESY and TOCSY experiments were performed on an Agilent 900 MHz direct drive spectrometer equipped with cryoprobe (Grzesiek *et al*., 1993, Monetlione *et al*., 1992, Bax *et al*., 1990, Marion *et al*., 1989). H(C)CH-TOCSY experiments were performed with 1514 points and 9817 Hz and 5622 Hz for the direct and indirect ^{1}H and ^{13}C dimensions respectively. ^{13}C-edited NOESY-HSQC was performed with 1258 points spectral widths of 8978 Hz and 6622 Hz for the ^{1}H dimensions and ^{13}C dimension respectively. Aromatic NOESY was performed with 1082 points and spectral widths of 6862 Hz, 6099 Hz and 8000 Hz for the direct ^{1}H, indirect ^{1}H and ^{13}C dimensions respectively. ^{15}N-edited NOESY-HSQC was performed with 1434 points and spectral widths of 9113 Hz and 2500 Hz for the direct and indirect ^{1}H dimensions and ^{13}C dimension respectively.

2.7.5 Hydrogen/Deuterium (H/D) exchange

H/D exchange was performed using 400 µM of ^{15}N-YraP, which was lyophilised using a HETO PowerDry LL1500 lyophiliser (Thermo). The protein was resuspended into 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 in 100% D_2O. 400 µM of ^{15}N-YraP was prepared in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 (90% H_2O, 10% D_2O) was used as a reference.

2.7.6 Structure Calculations

NMR spectra were processed using NMRPipe (Delaglio *et al*., 1995), using parameters such as zero-filling to double the number of points, linear prediction, Fourier transformation, water signal suppression and baseline correction. Assignments of YraP were performed using the program SPARKY3 (Goddard and Kneller, 2004).

Secondary structure predictions and dihedral angles restraints were calculated using the prediction software TALOS (Cornilescu *et al*., 1999), using chemical shift data from the backbone assignment process (CA, CB, CO and NH).

Hydrogen bond restraints were added in manually and 3D restraints were generated using NOESY data, which were interpreted as proton distance constraints using SPARKY3.

All structure calculations were performed using the automated NOE assignment program CYANA (Guntert, 2004). In brief, calculations involved 7 cycles of structure calculations with NOE restrain filtering, starting from 200 random conformers, with 10,000 torsion angle steps performed for each conformer. After 7 cycles, the final structures are calculated, with the top 20 conformers with the lowest target energy functions are chosen as the representative structures of the calculations. RMSDs are calculated from the final structure using the MOLMOL program (Koradi *et al*., 1996).

2.7.7 NMR for lipid binding titrations

Di-hexanoyl Phosphatidylglycerol (PG) was solubilised in NMR buffer at a stock concentration of 100 mM and added to 300 mM ^{15}N YraP at concentrations of 0, 8, 16, 24, 32 and 40 mM, with 10% D_2O and NMR buffer. HSQCs were performed on a Agilent 800 Mhz spectrometer.

Chemical shift perturbation was determined by taking peak intensities of the first and last point of the titration and using this formula:

$$\Delta\delta_{Overall} = [(\Delta\delta_H \times 500)^2 + (\Delta\delta_N \times 50.7)^2]^{0.5}$$

2.7.8 SurA binding by NMR

His-tagged SurA was expressed and purified as described in **Section 2.6**. 700 µM SurA was added to 300 µM ^{15}N-YraP in 50 mM sodium phospahte, 50 mM NaCl at

pH 6.0 (90% H_2O, 10% D_2O). HSQCs were performed on ^{15}N-YraP with and without SurA and spectra were compared to determine binding.

2.8 Functional experiments studying YraP

2.8.1 Screening of OM defects

BW25113, Δ*yraP* and Δ*bamE* cells were transformed with pET20b or pET20b-complementing gene screened for OM defects by streaking individual transformant colonies onto either 100 µg/ml vancomycin, 4% SDS or 0.5% SDS + 0.5 mM EDTA, depending on the experiment and 100 µg/ml ampicllin LB-Agar plates, as a positive control, from a single streak. The presence or absence of confluent growth was used to determine if a strain was sensitive or resistant to the condition. Streaks were performed in replicates.

2.8.2 Analytical Ultracentrifugation Sedimentation Velocity

Purified YraP was concentrated to 92.2 µM (corresponding to an Abs_{280} = 0.789), as Abs_{280} values of 0.1 - 1.0 are needed for AUC experiments, in 50 mM Sodium Phosphate buffer, 300 mM NaCl pH 7.2. Sedimentation velocity experiments were performed using An Ti50 rotor which was spun at 40,000 RPM for 24 hrs using a Beckman ProteomeLab XL-1 ultracentrifuge. The data was analysed using the software SEDFIT.

2.8.3 Analytical gel filtration

Analytical gel filtration was performed on a Superdex S200 column (28 ml) (GE Healthcare). 500 µl of sample was loaded in gel filtration buffer (50 mM sodium phosphate, 300 mM NaCl pH 7.2). The column was equilibrated and eluted with 1.2 column volumes, with a flow rate of 0.5 ml/min and 0.5 ml elution fractions were collected.

2.8.4 Immunfluorescence Microscopy

Overnight growths of cells (BW25113 and Δ*yraP*) were diluted 1:10 into LB and grown to OD_{600} 0.3. 1 ml of cells were centrifuged (10 mins at 5000 x *g*), washed in PBS (140 mM NaCl, 27 mM KCl, 10 mM Na2HPO4.2H2O, 2 mM KH2PO4, pH 7.2), and then fixed by resuspension in 4% paraformaldehyde in PBS at pH 7.2. Cells were washed twice in PBS and finally resuspended in 1 ml of PBS.

As a positive control, cells were permeabilised after fixation as described by (den Blaauwen *et al*., 2001). This was done by collecting fixed cells by centrifugation and washing once with PBS and incubating in 0.1% Triton X-100 in PBS for 45 min at room temperature (RT). Cells were recovered by centrifugation and washed 3 times in PBS and then incubated in 100 µg/ml lysozyme and 5 mM EDTA for 45 min at RT. Finally the cells were washed 3 times in PBS.

Microscope slides were coated with 0.1% poly-L-lysine. 20 ul fixed cells were added to slides and then dried off. The slides were washed 3 times with PBS-Tween. Primary antibody was added to slides at a dilution of 1:5000 and incubated in a moisture chamber for 1 hour at room temperature. The slides were then washed 3

times with PBS-Tween and Goat Anti Rabbit Alexa Fluor® 488 conjugated antibody (Invitrogen) was added at a dilution of 1:10000 and incubated for 1 hour in darkness. The slides were finally washed 3 times with PBS-Tween and then dried. Mounting solution was used and a coverslip was placed over the cells.

Imaging was performed using Leica DMRE fluorescence microscope-DC200 digital camera system.

2.8.5 Sucrose density centrifugation

250 ml of cells (BW25113 and Δ*yraP*) were grown in LB to OD_{600} 0.8 and harvested by centrifugation (6000 rpm for 30 minutes). The pellet was resuspended in 20 ml of Buffer M (50 mM Triethanolamine, 1 mM DTT and protease inhibitor, pH 7.2). Cells were lysed by 6 passages through an Emulsiflex (Avastin), and centrifuged as before to remove unbroken cells. The lysate was centrifuged at 75000 x *g* for 1 hour, allowing the membranes to be recovered in the pellet. Membranes were washed once in Buffer M and resuspended in approximately 3.5 ml of Buffer M normalised to the OD_{600} reading.

The final membranes were transferred onto a five-step sucrose gradient to separate the outer and inner membranes (Osborn *et al*., 1972). Sucrose gradients were prepared in Buffer M by layering sucrose solutions into 14 x 95 mm ultracentrifuge tubes (Beckman) as follows (from bottom to top): 0.8 ml at 55%, 3.0 ml at 50%, 3.0 ml at 47%, 2.0 ml at 42% and 0.8 ml at 30% (all w/w). Sucrose gradients were centrifuged

(210,000 x *g*, 16 h, 4ºC) in an SW 40 Ti rotor (Beckman), and 500 µl fractions were collected by puncturing the bottom of the tube with a needle. Outer membrane fractions were determined by boiling samples with 2 x Laemmli buffer and running SDS-PAGE. Outer membrane fractions were pooled and dialysed against Milli-Q water to remove the sucrose.

2.8.6 Co-Immunoprecipitations

Bacteria for co-immunoprecipitations (Co-IPs) were prepared in the following ways according to (Phizicky and Fields, 1995).

50 ml of BW25113 and Δ*yraP* cells were grown overnight at 37^{o}C in LB. Cells were harvested by centrifugation (8000 x *g* for 10 minutes) and resuspended in 2-3 ml PBS (pH 7.2) normalised according to the OD_{600} measurements. Protein complexes were cross-linked with 0.5 mM dithiobis(succinimidyl) propionate (DSP) and shaking incubation at RT. Cells were lysed by the addition of 100 µg/ml Lysozyme, with 1% Zwittergent 3-14 (Calbiochem) and 2 µl Benzonase. Following lysis, a second roud of cross-linking was performed by adding 0.5 mM DSP (1 mM total DSP final concentration), and incubation at RT for 30 min. The cross-linking reaction was quenched by the addition of 100 µl of 1 M Tris-HCl (pH 7.4) for 10 min at RT. Membranes were solubilsed with 2% Zwittergent 3-14 and recovered by centrifugation at 130,000 x *g* for 30 min at 4^{o}C. The recovered samples were made up to 10 ml in PBS.

50 ml BW25113 and Δ*yraP* cells were grown overnight at 37°C in LB. Cells were harvested by centrifugation (8000 x *g* for 10 minutes) and resuspended in approximately 10 ml PBS (pH 7.2) normalised according to the OD_{600} measurements. Protein complexes were stabilised with 0.25% formaldehyde for 15 min at 37°C. Bacteria were recovered by centrifugation, washed once in PBS and then a second round of crosslinking was performed with 0.4% formaldehyde. Spheroplasts were generated by resuspending cells in ice cold spheroplast buffer (0.75 M sucrose, 50 mM Tris pH 7.8, 0.6 mg/ml lysozyme, 6 mM EDTA) with shaking incubation at RT for 90 min. Spheroplasts were lysed by adding 1% Triton X-100 (Sigma) with 15 mM $MgCl_2$. Unlysed cells were removed by centrifugation and the supernatant was retained.

Co-IPs were performed by adding 5 µg of antibody to samples with shaking incubation for 4 hours at RT. 50 µl of pre-washed Protein A-Sepharose® (Sigma) was added to the sample and incubated at 4°C overnight. Beads were recovered by centrifugation and washed 6 times with PBS 0.1% Zwittergent 3-14. Beads were then resuspended in 2 x Laemmli buffer.

Outer membrane vesicles were generated as descrbed in **Section 2.8.5** and solubilised in PBS with 0.5% DDM. 600 µg of protein (determined by Bradford assay) was used to perform co-immunoprecipitations using a Pierce Direct IP Kit. Samples were eluted by resuspension of beads and in 2 x Laemmli buffer.

2.8.7 Periplasmic Co-IP

BL21 (DE3) cells were transformed with either pET26b-*yra*P-His or with empty pET26b vector. Colonies were picked and inoculated into 5 ml LB, with appropriate antibiotics, and grown overnight at 37^{o}C. 1 ml of the overnight culture was used to inoculate 50 ml of LB with antibiotic, which were grown at 37^{o}C at 180 rpm. Once OD_{600} = 0.6 had been reached, cells were induced with 0.5 mM IPTG, and grown for another 3 hours.

Cells were harvested by centrifugation at 6000 x *g* for 10 minutes and resuspended in PBS. For cells that required cross-linking, cells were resuspended in 10 ml PBS, and 0.5 mM DSP was added. The cells were incubated at room temperature for 30 min. 50 mM Tris-HCl was added and cells were incubated for a further 15 mins to quench the reaction. Cells were harvested by centrifugation as before and resuspended in 5 ml of 5 mM $CaCl_2$. Cells were recovered by centrifugation and then resuspended in 8 ml Osmotic Shock buffer (33 mM Tris-HCl pH 8.0, with 5 mM EDTA and 20% sucrose) and incubated at 4^{o}C for 15 min. Cells were then harvested by centrifugation and resuspended in 5 ml of reverse osmosis water, which was incubated at 4^{o}C for 15 min. After a final centrifugation, the supernatant was concentrated using a 3 kDa MWCO filter (Millipore).

2.8.8 Identification of proteins

Eluates of Co-IPs were analysed by SDS-PAGE and stained using a Silver Stain Plus kit (BioRad) according to the manufacturers recommendation.

Gel slices were excised and washed with 300 µl ultrapure water. Silver stain was removed by the addition of 200 µl 15 mM potassium ferricyanide / 50 mM sodium thiosulphate and incubation at RT for 1 hr. Gel slices were washed again in ultrapure water and then washed twice with 50% acetonitrile / 50% ammonium bicarbonate. 50 mM DTT (dithiothreitol) in 10% acetonitrile / 50 mM ammonium bicarbonate was added and bands were incubated at 56^{o}C for 1 hour. The supernatant was removed and 100 mM iodoacetamide in 10% acetonitrile / 50 mM ammonium bicarbonate was added and incubated at RT in the dark. Bands were washed 3 times with 10% acetonitrile / 50 mM ammonium bicarbonate and bands were dried down using a Speedvac (Eppendorf). Gel slices were then rehydrated with 12.5 µg/ml Trypsin (Promega) in 10% acetonitrile / 50 mM ammonium bicarbonate and incubated at 37^{o}C overnight. The supernatant was collected and 3% formic acid was added to the gel slice and incubated at 37^{o}C for 1 h. This step was repeated and all supernatants were pooled together. A final concentration step was performed using a Speedvac, to produce a final volume of approximately 20 µl.

Peptides were separated using MALDI-TOF (Brucker) and identified using MASCOT database.

2.8.9 Generation of suppressor mutants

BW25113 Δ*yraP* cells were grown on 100 µg/ml vancomycin and 0.5% SDS + 0.5 mM EDTA LB-Agar plates until single colonies were observed. Colonies were picked and selected again on the test conditions. A further 2 selections were

performed and on the final selection, additional streaks were performed on 50 µg/ml kanamycin LB-Agar.

Genomic DNA was extracted from the suppressors using a QiagenQiaAmp DNA Blood Mini Kit, according to manufacturer"s recommendations. Whole genome sequencing was performed by Dr. David Rasko (University of Maryland) and Dr. Trevor Lawley (Sanger Institute, Cambridge). Analysis was performed by comparing the mutant genomic data to either WT BW25113 or Δ*yraP* genomic data.

2.8.10 Mutagenesis of genes

Mutations were performed in three ways. Firstly with the use of PCR with mutations designed in the primers (Alta Biosciences). Genes with the mutations were amplified from these primers, purified using a Qiagen PCR cleanup kit and digested with appropriate restriction enzymes alongside with vectors. These inserts were then ligated into the vectors with a 3:1 insert to vector ratio, using an Invitrogen T4 Ligation Kit at 4^{o}C overnight.

Mutations were also introduced using a QuikChange® Lightning Kit (Agilent), using custom oligos containing mutations (Invitrogen), designed by using the Agilent QuikChange Primer Design tool. PCR were setup according to the manufacturers recommendations (5 µl of 10x reaction buffer, 1 µl of pET20b-*yraP* [100 ng of double stranded template DNA], 1.25 µl of each primer [125 ng of primer], 1 µl dNTP mix, 1.5 µl QuikSolution reagent, distilled H_2O to a final volume 50 µl and 1 µl of QuikChange Lightning Enzyme). The thermal cycling protocol was carried out

with an initial denaturing step at 95^{o}C for 2 mins, 18 cycles of 95^{o}C for 20 secs, 60^{o}C for 10 secs and 68^{o}C for 2.5 mins, with a final extension step at 68^{o}C for 5 mins. Template DNA was digested with 2 µl of *DpnI* enzyme at 37^{o}C for 5 mins. 2 µl of *DpnI* treated DNA was transformed into 45 µl of XL1-Blue ultracompetent cells using heat shock transformation described in **Section 2.4.3**.

Finally, a transposon-mediated mutagenesis kit was used for random mutagenesis of YraP using an MGS™ Mutation Generation System Kit, (Thermo). This kit would randomly insert a 15 bp in-frame insertion into YraP (XXXXXTGCGGCCGCA, where XXXXX is the randomly selected 5 bp target site for transposition). Reactions were carried out using 1.5 µl pET20b-*yraP* (150 ng DNA), 4 µl 5x MuA Transposase Reaction Buffer, 1 µl Entranceposon M1-Kan^{R} (100 ng), 1 µl MuA Transposase and 12.5 µl distilled H_2O to make a final volume of 20 µl. The reaction was mixed and incubated at 30^{o}C for 1 hour. The MuA Transposase was inactivated by incubation at 75^{o}C for 10 mins. Reactions were diluted ten-fold and 5 µl was heat shock transformed into 50 µl XL1-Blue cells and plated onto 50 µg/ml kanamycin LB-Agar plates.

2.8.11 LPS visualisation

Overnight cells (BW25113, Δ*yraP*, Δ*bamE*, SL1344, SL1344 Δ*yraP* and SL1344 Δ*bamE*) were grown in 5 ml LB-antibiotic and OD_{600} readings were recorded. A normalised volume of each growth ([1000/OD_{600}] µl) was used and cells were harvested by centrifugation (14,000 x *g* at 10 min). Supernatants were removed and cells were resuspended in 100 µl 2 x Laemmli buffer (Sigma-Aldrich). Cells were

lysed by boiling for 5 min, freezing at -80^{o}C for 5 min and then boiling again for 10 min. Lysates were centrifuged and 80 µl of the supernatant was collected. 5 µl of 5 mg/ml Proteinase K (in 2 x Laemmli buffer) was added to the supernatants and kept at 60^{o}C for 1 h. Samples were heated at 95^{o}C for 5 min and SDS-PAGE was performed with 10 µl of each sample. LPS was visualised by staining gels using a ProteoSilver™ Plus Silver Stain Kit (Sigma-Aldrich) (Zhu *et al*., 2012).

2.8.12 Population difference identification of OM phospholipids

OM vesicles were generated using sucrose density centrifugation as described in **Section 2.8.5** and solubilised in ultrapure water. 6 different BW25113 and 6 different BW25113 Δ*yraP* OM vesicles, picked from individual colonies, were used in this study. Lipids were extracted from the OM vesicles using a two-step methanol:water:chloroform protocol (Wu *et al*., 2008), performed by Dr. Jennifer Kirwan. The nonpolar extracts were dried under nitrogen and then made up in their original volume of methanol:chloroform (3:1). Fourier transform ion cyclotron resonance (FT-ICR) mass was performed using a LTQ FT Ultra instrument (Thermo Scientific). Experiments were performed in triplicates for each sample. Data was processed (Southam *et al*., 2007) and principal components analysis (PCA) was used to assess the differences in phospholipid populations between BW25113 and Δ*yraP* groups using the software MATLAB.

2.8.13 Relative quantification of outer membrane proteins

Outer membrane fractions were generated using sucrose density centrifugation as described in **Section 2.8.5** and solubilised in PBS with 0.5% DDM. 200 µg of protein, determined by Bradford assay, were denatured with 8 M urea, 50 mM Ammonium bicarbonate, 25 mM DTT and 1% CHAPS and then treated with 50 mM iodoacetamide. Samples were then retained in a 30 kDa MWCO filter (Millipore) by centrifugation at 15000 x *g* for 5 minutes and washed 5 times with 50 mM Ammonium bicarbonate. Proteins were digested with 5 µg of trypsin and incubated at 37^{O}C overnight. Peptides were obtained by centrifugation through the filter, treated with 0.1% Trifluoroacetic acid (TFA) and desalted using a Disocovery® DSC-18 SPE column (Sigma-Aldrich). Peptides were eluted in 600 µl 50% Acetonitrile and 0.1% TFA. Acetonitrile was removed using a Speed-Vac (Eppendorf).

Peptides were labelled as described by (Hsu *et al*., 2003). In brief, peptides from BW25113 cells were labelled with 4% formaldehyde whereas peptides from Δ*yraP* cells were labelled with 4% deuterated formaldehyde. Overall, this would ensure that peptides from Δ*yraP* cells would be 4 Da heavier than peptides from BW25113 cells. ESI-QTOF was performed using a Bruker Micro-TOF to separate labelled peptides (performed by Dr. D. Ward and Dr. N. Shimwell). Data was analysed by comparing the elution profiles of the peptides using ProteinScape software (Bruker).

2.8.14 Screening of phenotypes using BioLog™

Wild-type *E. coli* BW25113 and an isogenic BW25113ΔbamE mutant were analysed using Phenotype Microarrays (BioLog™) for chemical sensitivity for bacteria (PM plates 11-20. Drug content of each plate can be found at ww.biolgo.com/pdf/PM11-PM20.pdf). Arrays were conducted in triplicate over a 24 hour period and growth rates compared between WT and the mutant. Identified conditions where the WT strain grew at a faster rate than the mutant were indicated on the microarray as green data points, were replicated under laboratory conditions.

Chapter 3

Functional and Structural Study of BamE

3.1 Introduction

BamE (formerly known as SmpA) has previously been shown to be member of the Bam complex (Sklar *et al*., 2007a). The gene itself encodes a small (12.3 kDa) lipoprotein which serves as an accessory factor to BamA. Previous studies have shown that BamE forms a subcomplex with BamC and BamD, which binds to BamA in a manner that is independent of BamB (Sklar *et al*., 2007a, Gatsos *et al*., 2008, Malinverni *et al*., 2006). Although BamE is shown to be a member of the Bam complex, the exact stoichiometry of BamE to the other subunits is still unknown, though studies by (Hagan and Kahne, 2011) have shown that a 1:1 stoichiometry to the other members of the Bam machinery is needed to produced an active complex.

The gene *bamE* is found across all proteo-bacteria (Gatsos *et al*., 2008) and multiple sequence alignments have confirmed that *bam*E is well conserved (Knowles *et al*., 2011, Kim *et al*., 2011c). Although fairly ubiquitous, *bamE* is a non-essential gene. Δ*bamE* strains are viable, however they do show growth defects as well as having moderate OM defects, including inefficient OMP assembly and increased sensitivity to various antibiotic agents (e.g. vancomycin) and OM disrupting agents (e.g. SDS and EDTA) (Sklar *et al*., 2007a, Rigel *et al*., 2011). These defects have also been reported in homologous Δ*bamE* strains in other organisms, such as *N. meningitidis, S. typhimurium* and *C. crescentus* (Volokhina *et al*., 2009, Lewis *et al*., 2008, Ryan *et al*., 2009). These observations suggest that *bamE* has a key role in OM biogenesis.

The exact function of BamE remains enigmatic. It is widely accepted that BamE is involved in stabilising the interaction of BamD to BamA with the aid of BamC. Although Δ*bamC* and Δ*bamE* strains have similar phenotypes, it has been suggested that they do not have redundant functions. (Rigel *et al*., 2011) have suggested that BamE might function as a modulator of BamA, possibly altering the conformation of BamA through its interactions with BamD. It has also been reported that BamE has a role in bacterial stalk growth in *Caulobacter crescentus* as well as maintaining membrane integrity as well as a role in phospholipid binding and genetic interactions with other OM biogenesis proteins (Knowles *et al*., 2011, Ryan *et al*., 2009, Bennion *et al*., 2010). Whatever the exact role of BamE is, it plays an important role in OM homeostasis.

The structure of BamE has been solved by several groups (Knowles *et al*., 2011, Kim *et al*., 2011c, Albrecht and Zeth, 2010). The results of these studies show that the overall fold of BamE consists of an anti-parallel β-sheet packed against a pair of α-helices with an ααβββ topology (**Figure 3.1**). This particular fold is reminiscent of the fold of a POTRA domain (Kim *et al*., 2007, Knowles *et al*., 2008), further suggesting a role in OM biogenesis and possibly in interacting with precursor OMPs, as is the case with POTRA domains. Both the (Kim *et al*., 2011c, Albrecht and Zeth, 2011) studies demonstrated the possibility of BamE dimerisation, but this was suggested to be an artefact due to cytoplasmic expression in the Knowles *et al*. 2011 study.

The Knowles *et al*. 2011 study demonstrated some further functional features of BamE, demonstrating its binding to the OM phospholipid phosphatidylglycerol (PG). The identification of the residues involved in lipid binding were also subjected to a functional analysis, which involved generating point mutations and observing growth on vancomycin. These residues mapped onto a conserved core region of BamE. This study also demonstrated BamD and BamE binding and identification of residues involved in this interaction.

This chapter details some of the functional work performed in the Knowles *et al*. 2011 study, which led to a comprehensive structural and functional study of BamE. Prior to this publication, the structure of BamE had not yet been determined.

Figure 3.1 Structure of BamE. Solution structure of BamE (taken from Knowles *et al*., 2011) (PDB:2KM7). The fold contains three β-strands overlaid with a pair of α-helices, reminiscent of the fold of a POTRA domain.

3.2 Results

3.2.1 OM defects associated with Δ*bamE* strains

Previous work (Sklar *et al*., 2007a) has shown that BamE is part of the of the BAM complex and that knockouts of the genes associated with the BAM complex have associated OM defects. We wanted to elucidate the defects associated with Δ*bamE* strains. These could then be used to allow complementation to further understand the function of *bamE* in regards to OM homeostasis.

In order to determine possible conditions which Δ*bamE* cells are sensitive to, a BioLog™ screen was performed (**Figure 3.2**). From the initial hits that were produced, we decided to replicate them on LB Agar plates to confirm the sensitive conditions.

We decided to choose conditions that were readily available and that were not difficult to reproduce. The results from these experiments suggested the use of 4% SDS and 100 μg/ml vancomycin as test conditions, where Δ*bamE* strains show immediate sensitivity.

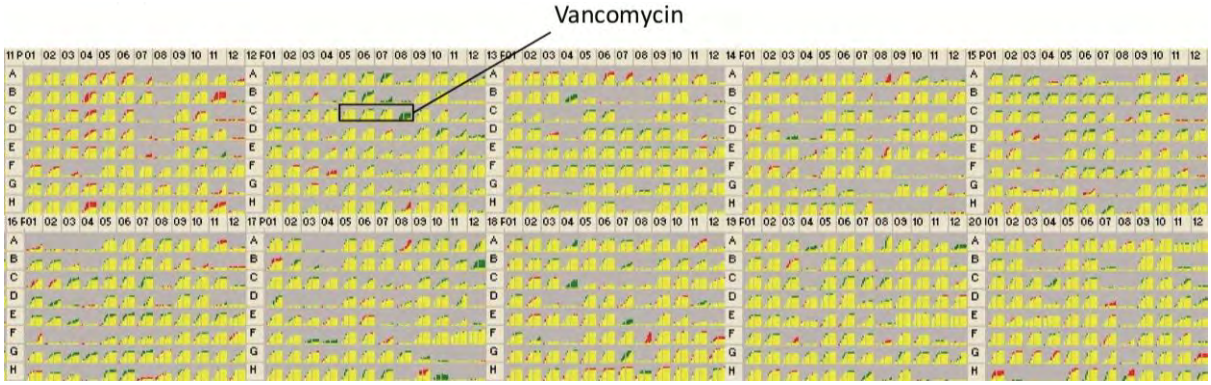


Figure 3.2 Δ*bamE* cells display sensitivity to vancomycin. WT BW25113 and Δ*bamE* cells were analysed using a BioLog™ screen to determine sensitivity to a range of compounds. Arrays were performed in triplicate over a 24 hour period. The data displays growth rates of bacteria derived from one replicate. Growth rates are gauged by colour, with yellow curves indicating a similar growth rate and green and red curves indicating faster rates for WT and Δ*bamE* respectively. The drug content of each well can be found at www.biolgog.com/pdf/PM11-PM20.pdf. The position of vancomycin is displayed, with higher concentrations of vancomycin demonstrating that Δ*bamE* cells fail to grow as fast as WT. Due to availability and reproducibility, vancomycin was used as a test condition where Δ*bamE* would not outperform WT cells.

3.2.2 Cloning of *bamE* into pET20b

In order to study *bamE*, the full length gene was cloned into expression vector pET20b. This expression vector lacks the *lacI* gene, which allows leaky expression in non-T7 polymerase containing strains. Primers were designed to amplify the gene from BW25113 genomic DNA, with *NdeI* and *EcoRI* restriction enzyme sites. The size of the *bamE* gene is 342bp. Agarose gel electrophoresis shows that PCR products between the 200 and 400 bp markers (**Figure 3.3**), indicating that the PCR has worked.

The fragments were then gel excised, purified and cloned into pET20b and confirmed by DNA sequencing (Functional Genomics, University of Birmingham).

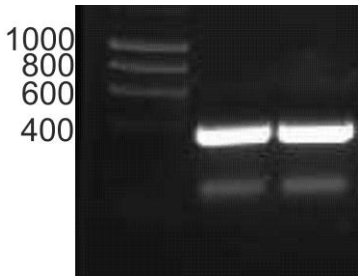


Figure 3.3 PCR of *bamE* used for cloning into pET20b. Using the primers mentioned in **Table 2.5.4.1**, a PCR product is observed at just under the 400bp marker. The PCR reactions are pure, with no other products seen, except for free nucleotides at the bottom of the gel.

3.2.3 Complementing the defective phenotype associated with Δ*bamE* strains

As we had determined two phenotypes that could be tested with BW2113 Δ*bamE* strains, we wanted to see if the pET20b-*bamE* plasmid was able to complement the OM defects. WT BW25113 cells were transformed with pET20b, to act as a control, whilst Δ*bamE* strains were transformed with pET20b and pET20b-*bamE*. Individual transformant colonies were picked and a streak assay was performed onto 4% SDS and 100 µg/ml vancomycin LB-Agar plates, where the presence of confluent growth could be observed under these test conditions (**Figure 3.4A**). A single pick was used to streak onto each of the test conditions, with a final streak onto 100 µg/ml ampicilin LB-Agar plates. This was repeated an additional 4 times for reproducibility. Our results show that complementation was achieved as Δ*bamE* cells with pET20b-*bamE* could grow similar to WT cells, unlike those with pET20b, which showed

sensitivity to these conditions, suggesting that the leaky expression of pET20b-*bamE* was sufficient to rescue the Δ*bamE* phenotype.

To confirm that complementation was due solely to the effect of leaky expression from the pET20b-*bamE* plasmid, LB overnight cultures of the colonies were grown and western blotting was performed on cell lysates using anti-BamE antibody. The blot result shows that WT BW25113 cells produce BamE, and that the Δ*bamE* with pET20b strain is deficient for BamE as expected (**Figure 3.4B**). With the Δ*bamE* pET20b-*bamE* complemented strain, BamE is seen, but at a slightly elevated level than that of WT BW25113 expression. This suggests that expression levels from the pET20b vector are greater than endogenous expression levels.

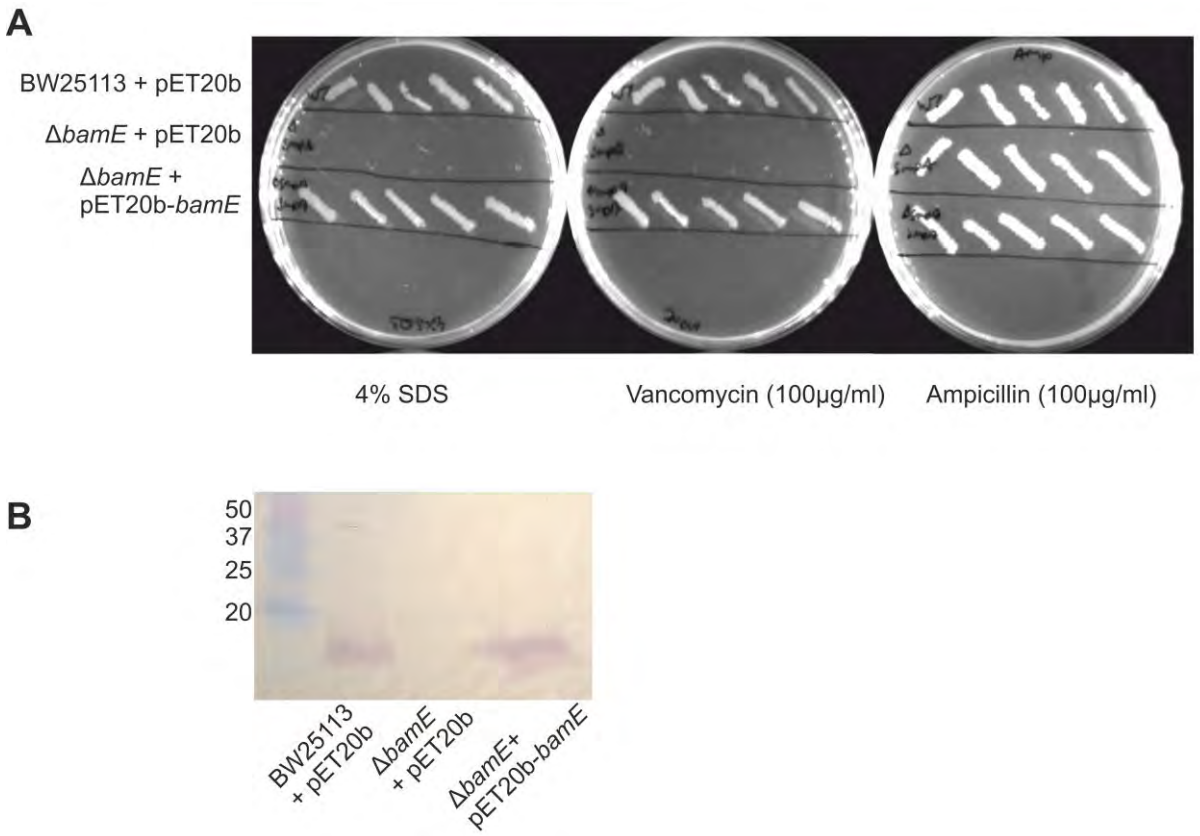


Figure 3.4 pET20b-*bamE* is able to complement the Δ*bamE* phenotype on vancomycin and SDS LB-Agar plates. A BW25113 + pET20b, Δ*bamE* + pET20b and Δ*bamE* +pET20b-*bamE* were struck onto the following test conditions: 4% SDS and 100 µg/ml vancomycin and 100 µg/ml ampicillin was used as a positive control. An individual transformant colony was streaked onto all of the plates in the same positions and streaks were repeated and additional 4 times with other colonies. Complementation is shown for the Δ*bamE* pET20b-*bamE* cells and is able to grow similar to WT cells under these test conditions. **B** Western blot performed with 20 µl of cell lysates from the selected strains and blotted against anti-BamE antibody (purified from rabbits) at a 1:5000 dilution.

3.2.4 Mutagenesis of the *bamE* gene

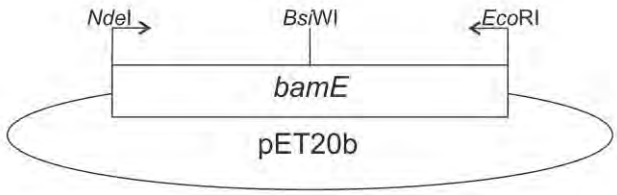
To further our understanding of the functional role of *bamE* in OM homeostasis, we performed a scanning mutagenesis screen. Our aim was to mutate each amino acid residue of BamE into either a cysteine or glycine residue and see if the particular

mutant was able to complement the defective phenotype of Δ*bamE*. This line of experimentation would indicate which residues are crucial for BamE function.

Figure 3.5A shows a brief restriction enzyme map of pET20b-*bamE*. There are three unique restriction enzyme sites that are present in the *bamE* gene, *NdeI* at the start, *EcoRI* at the end and *BsiWI* at 167bp (**Figure 3.5B**). Designing primers containing one of these restriction enzyme sites, with an in-frame substitution of a residue, would allow us to achieve this aim. A redundant codon was used (KGC), which allowed the generation of both cysteine and glycine residues from a primer pair. The primers used in this study are shown in Appendix I. By amplifying PCR products using these primers which contain the mutant substitution from pET20b-*bamE* and digesting both the insert and pET20b-*bamE* vector with appropriate enzymes, mutants could be generated by ligating the digested PCR product into digested pET20b-*bamE*.

To simplify matters, residues 1 - 19 were omitted as these are cleaved by signal peptidase II during lipoprotein maturation. Using primers with the above sites would allow coverage of residues 45 to 113 of BamE. For residues that were too far away from the restriction enzyme site, megaprimer PCR was performed by Dr. Douglas Browning.

A



B

CATATGCGCTGTAAAACGCTGACTGCTGCAGCAGCAGTACTATTGATGTTGACCGCAGGCTGT
*Nde*I

TCCACTCTGGAGCGAGTGGTTTACCGTCCTGACATCAACCAGGGGAACTATCTGACCGCTAAC

GACGTATCCAAAATACGTGTTGGCATGACGCAACAACAAGTTGCGTACGCATTGGGTACACCG
*Bsi*WI

CTGATGTCCGATCCATTTGGTACGAATACCTGGTTCTATGTCTTCCGCCAGCAACCAGGTCAT

GAAGGTGTAACGCAGCAAACGCTGACGCTGACCTTTAACAGTAGCGGTGTGTTGACCAATATT

GATAACAAACCTGCGCTGAGTGGTAACTAAGAATTC
*Eco*RI

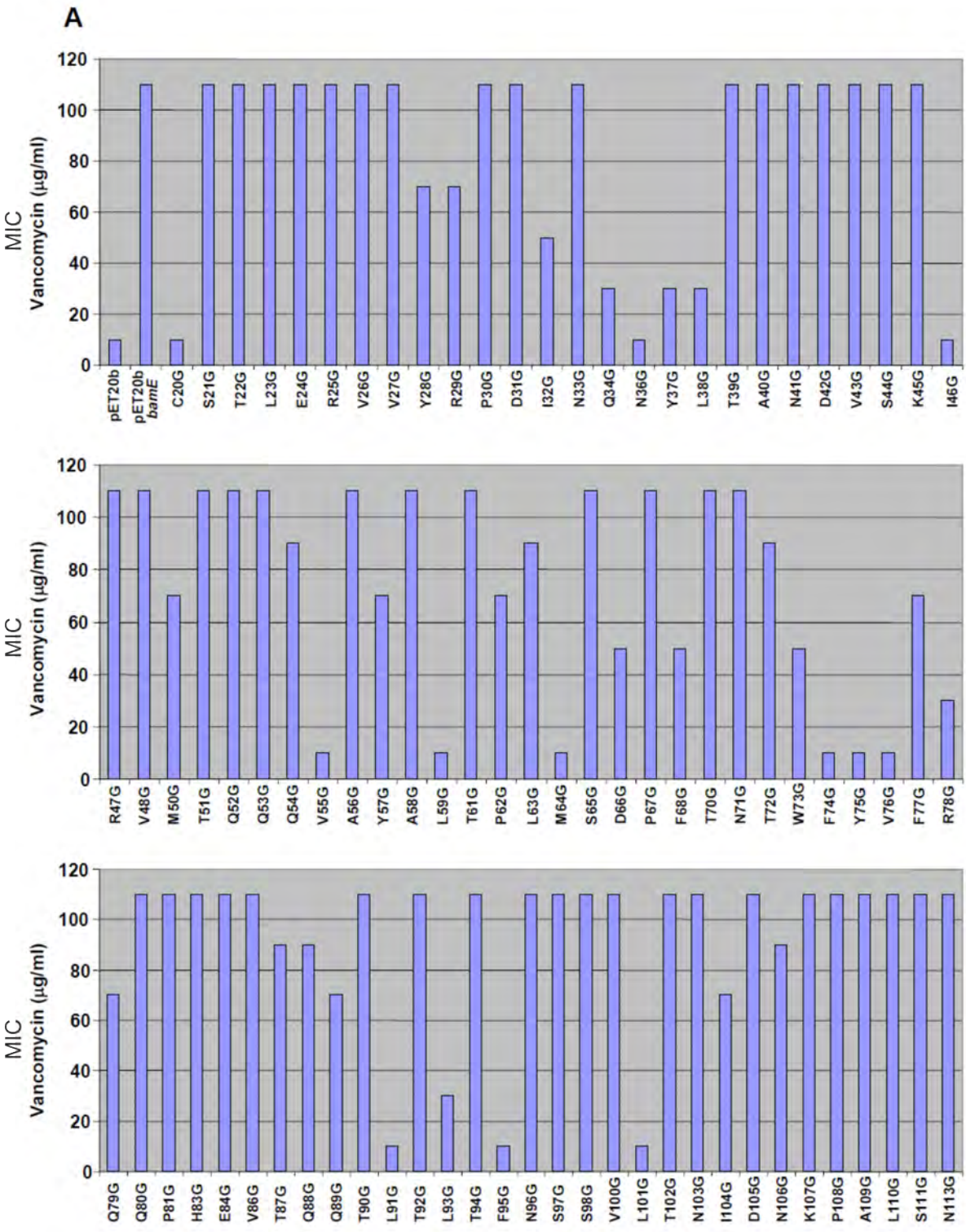
Figure 3.5 Restriction enzyme sites present in *bamE* allow primer based mutations to be performed. A Plasmid map of pET20b-*bamE*. *bamE* was cloned into pET20b using the *Nde*I and *Eco*RI sites. The orientation of the primers is indicated in the diagram. A *Bsi*WI site is located near the middle of the gene at 167bp of the 342bp *BamE* gene. **B** Sequence of the *bamE* gene with the restriction enzyme sites used in this study highlighted in red.

3.2.5 Screening of mutants

Mutants were screened onto the 4% SDS and 100 µg/ml vancomycin LB-Agar plates and a set of functionally relevant residues were identified. However this screening method has a major drawback in that only absence or presence of growth can be scored. To obtain quantitative data and gauge the severity of the mutants another screening method was employed. A range of vancomycin concentrations

were used (10, 30, 50, 70, 90 and 110 µg/ml), with a limit of 50 µg/ml vancomycin as a criterion to select for mutants (performed by Dr. D. Squire). SDS was not used in this screen due to problems with reproducibility and scoring mutants.

A list of mutants and their associated vancomycin inhibitory concentrations were generated using this screening condition (**Figure 3.6**). The data shows that there were some mutants where either the glycine or cysteine mutant was more sensitive to vancomycin than its counterpart. With the glycine mutants, due to the flexibility of this residue, there is a chance of mis-folded or unstable BamE being produced. To test this, western blotting was performed with anti-BamE antibody with these mutants, as unstable protein will results in degradation and will be detected at much lower levels (**Figure 3.7**). I64G, V55G, Y75G, L91G and F95G produce significantly reduced levels of BamE, suggesting that these mutants are unstable. From this approach, residues that are important in BamE function was identified (**Table 3.2**).


A
MIC
Vancomycin (µg/ml)
120
100
80
60
40
20
0
pET20b
pET20b bamE
C20G
S21G
T22G
L23G
E24G
R25G
V26G
V27G
Y28G
R29G
P30G
D31G
I32G
N33G
Q34G
N36G
Y37G
L38G
T39G
A40G
N41G
D42G
V43G
S44G
K45G
I46G
MIC
Vancomycin (µg/ml)
120
100
80
60
40
20
0
R47G
V48G
M50G
T51G
Q52G
Q53G
Q54G
V55G
A56G
Y57G
A58G
L59G
T61G
P62G
L63G
M64G
S65G
D66G
P67G
F68G
T70G
N71G
T72G
W73G
F74G
Y75G
V76G
F77G
R78G
MIC
Vancomycin (µg/ml)
120
100
80
60
40
20
0
Q79G
Q80G
P81G
H83G
E84G
V86G
T87G
Q88G
Q89G
T90G
L91G
T92G
L93G
T94G
F95G
N96G
S97G
S98G
V100G
L101G
T102G
N103G
I104G
D105G
N106G
K107G
P108G
A109G
L110G
S111G
N113G

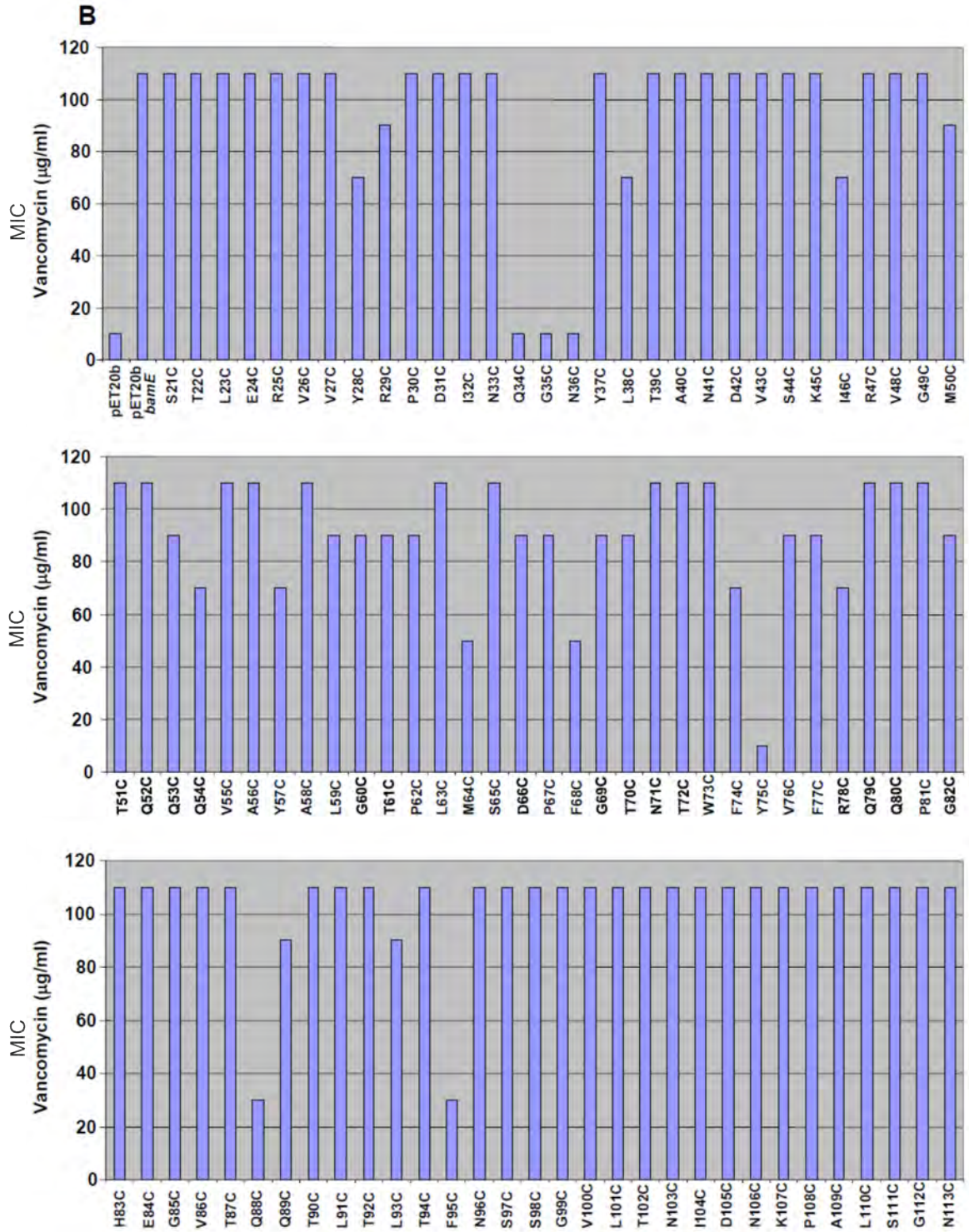


Figure 3.6 Vancomycin sensitivity of Δ*bamE* cells with the mutant plasmids. The panels show the minimum inhibitory concentration (MIC) of vancomycin that is needed to prevent growth of Δ*bamE* cells expressing BamE mutants with either **A** glycine or **B** cysteine substitutions, on LB-Agar plates supplemented with 100 µg/ml ampicillin. Each value is the median concentration of vancomycin obtained from at least three independent experiments.

Table 3.2 Summary of BamE mutations that affect *E. coli* sensitivity to vancomycin.

Mutation	Sensitivity to vancomycin (μg/ml)
WT	100
C20G	10
I32G	50
Q34G/C	30/10
G35C	10
N36G/C	10/10
Y37G	30
L38G	30
I46G	10
V55G	10
L59G	10
M64G/C	10/50
D66G	50
F68G/C	50/50
W73G	50
F74G	10
V75G/C	10/10
V76G	10
R78G	30
Q88G	30
L91G	10
L93G	30
F95G/C	10/30
L101G	10

Mutants that exhibit increased sensitivity to vancomycin (≤ 50 μg/ml, at least half the concentration needed to block growth of the wild type strain) are listed with their minimal inhibitory concentrations (MIC) shown. In cases where both glycine and cysteine mutants display sensitivity, both MICs are shown.

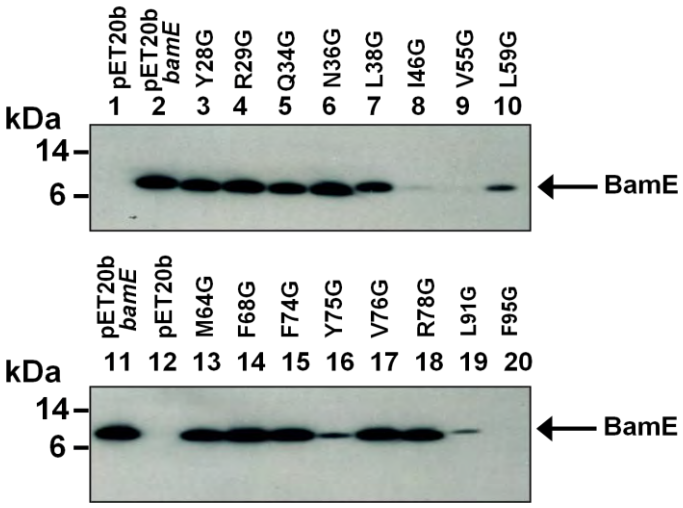


Figure 3.7 Mutants generated with point mutation substitutions to glycines can produce unstable protein. Western blot performed with 20 µl of cell lysate blotted against anti-BamE serum generated from rabbit.

3.3 Discussion

BamE is a major component of the Bam complex and its absence in Δ*bamE* cells gives rise to various OM defects such as hypersensitivity to antibiotics and inefficient OMP assembly (Sklar *et al*., 2007a). The exact role of BamE in OMP assembly is not completely understood. This study in part has aimed to address this problem by performing a structural and functional analysis of BamE.

Exploiting the property that Δ*bamE* cells show sensitivity to various compounds, we have been able to produce a complementation screen that gauges *bamE* function. By cloning WT *bamE* into the expression vector pET20b and using plasmid borne expression, we have been able to observe growth on vancomycin and SDS LB-Agar plates in Δ*bamE* strains. Western blotting has confirmed that this effect is solely due to expression of BamE from the vector.

A comprehensive mutation screen of BamE has been performed. Almost every residue of *bamE* has been mutated to either a cysteine or glycine, except for residues 1-19, which is part of the lipobox signal sequence and is cleaved during lipoprotein maturation. By screening these mutants, we have been able to identify a set of residues that are involved in BamE function (**Table 3.2**). We predicted that there would be three major classes of mutants. The first class of mutations are those that would occur in non-functional regions of BamE. This type of mutant would not produce any observable effects and the BamE produced from these mutants would be comparable to WT BamE. These mutants will not be discussed further. The second type of mutations would compromise the structural integrity of BamE

and hence result in protein degradation. These mutants would severely affect function, due to decreased levels of BamE, and most likely produce a phenotype similar to Δ*bamE*. The third class are mutations that would affect the function of BamE and produce observable results in the complementation assay.

One particular mutant C20G, was unable to complement the Δ*bamE* OM defect. This would be expected as in the WT protein, this cysteine is the site of acylation which is needed for lipoprotein maturation and localisation to the OM (Narita *et al*., 2004). In the glycine mutation however, this acylation cannot take place and hence localisation cannot occur.

Residues Q34, G35 and N36 (QGN motif) are conserved in BamE homologues. The mutagenesis of each of these residues to both cysteine and glycine has displayed poor growth on the test conditions. A similar situation is observed with the conserved residue F74. F74 is located between two aromatic residues, W73 and Y75, which are all next to the hydrophobic residue V76. The introduction of mutations in this region could disrupt the hydrophobic network and cause structural defects of BamE. Interestingly, the effect of introducing a glycine mutant at residues W73, Y75 and V76 is much more severe than a cysteine mutant, further suggesting hydrophobic interactions in this region potentially placing this region within the core of the protein. **Figure 3.8** shows the functionally important residues that are conserved.

A set of mutants consists of hydrophobic core residues (I64, V55, L59, Y75, L91, L93, F95 and L101). Glycine mutants of these residues were unable to complement the defective phenotype and from western blot analysis indicated that these mutants produced diminished levels of BamE. However cysteine mutants of these residues were not as debilitating, possibly due to a bulkier side chain group being able to maintain the hydrophobic interactions required for stability.

Another class of mutants (I32G, Q34G/C, G35C, N36G/C, Y37G, L38G, M64G/C, D66G, F68G/C, W73G, F74G, V76G, R78G and Q88C) were stably expressed, as estimated by western blotting, but failed to complement. These residues are more likely to be required for BamE function and not involved in the structural integrity of the protein.

The screening analysis also displayed a set of mutations that had slight defects in barrier function. These residues may be involved in BamE function, but are not as crucial as the other mutants identified (**Figure 3.5**).

The NMR structure of BamE was solved over the course of this study (Dr. Tim Knowles), along with the identification of residues involved in PG and BamD binding (Y37, L38, T61, L63, D66, F68, T70, N71, F74, Y75 and V76) (Malinverni *et al*., 2006, Kim *et al*., 2007). Some of the mutants identified in the vancomycin screen are the same ones that are involved in ligand binding (Y37, L38, D66, F68, F74 and V76), suggesting that a major role of BamE function is the ability to bind either PG or BamD. By mapping the functionally relevant mutants onto the structure of BamE,

it becomes apparent that all are surface exposed and form a single surface, providing strong evidence that this region has a functional role (**Figure 3.9**). Additionally, the F74A mutant completely abolished PG binding. This residue is located centrally in the PG binding site and is exposed, which could possibly disrupt the local hydrophobic network that could be required for PG binding.

In conclusion, a comprehensive screening assay to gauge BamE function has been developed. Mutants which correspond to functionally relevant residues have been identified from this approach. Combining this with the structure of BamE has identified a surface patch with functional importance. This patch was later shown to match a PG and BamD binding site identified by NMR, with mutation in this patch shown to abolish function. However, the exact role of BamE in OMP insertion is still unknown and if PG binding is involved in this process. Further experimentation is needed to confirm this, but the data generated in this study has shed some light on the function of BamE.

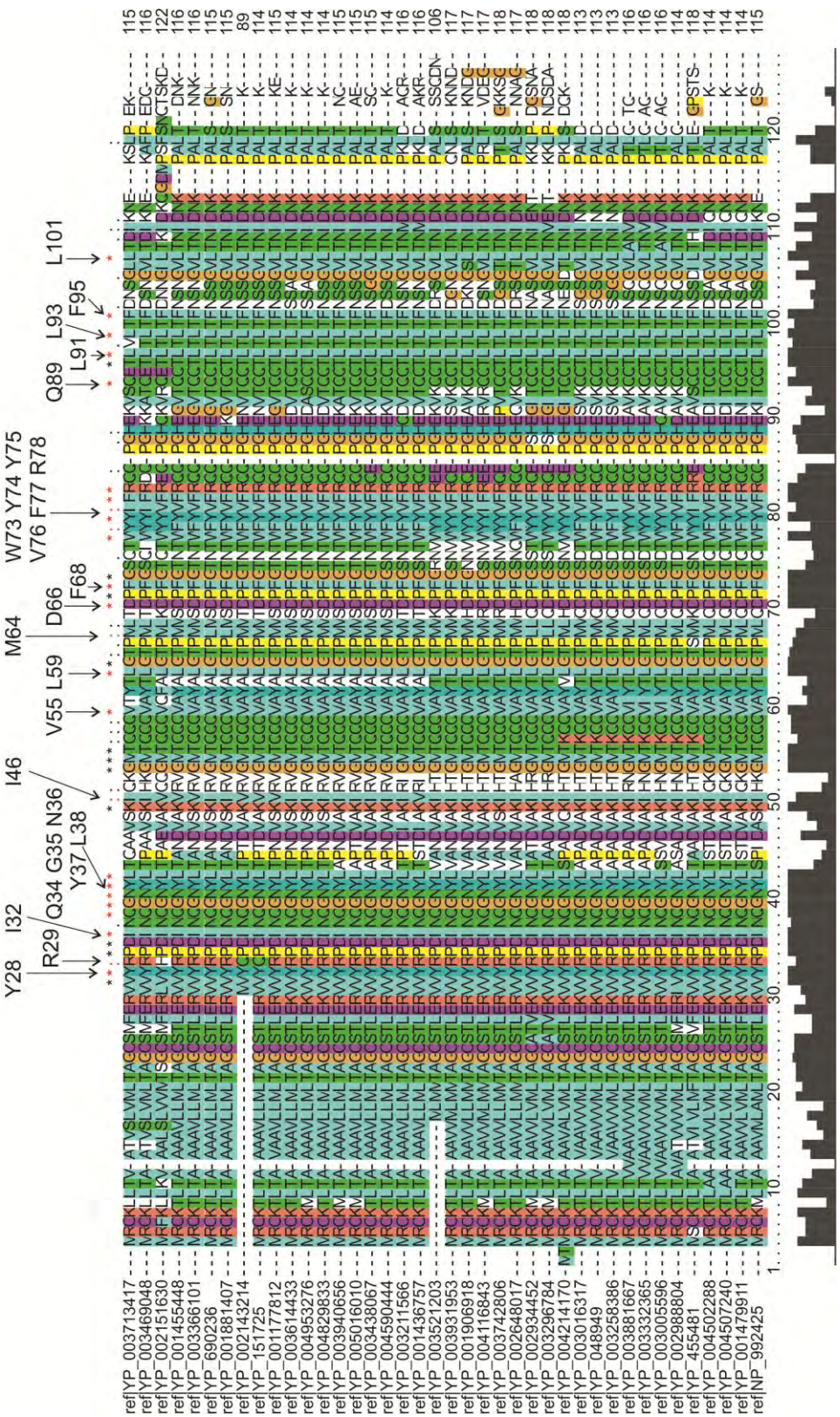


Figure 3.8 Functionally important residues found in the mutagenesis analysis are conserved. Multiple sequence alignment of several BamE proteins from other organisms based on amino acid similarity. Residues that have been identified to be functionally critical are highlighted at the top of the alignment.

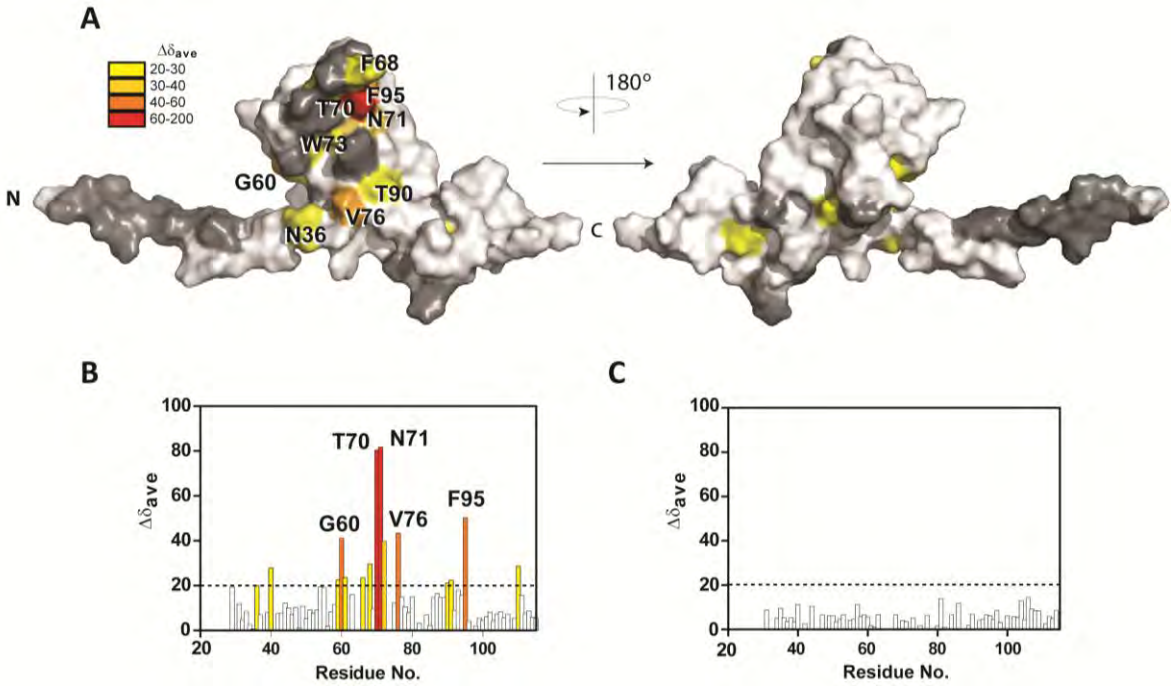


Figure 3.9 BamE binds phosphatidylglycerol (PG). **A** Surface representation of BamE with residues highlighted displaying significant chemical shift (CS) perturbations. Functionally relevant residues from the mutagenesis screen are indicated. **B** Histogram showing CS perturbations induced in ^{15}N-BamE (100 μM) on addition of DHPG (40 mM), coloured according to **A**. **C.** Histogram showing CS perturbations induced in labelled BamE (F74A; 100 μM) on addition of DHPG (40 mM). This particular mutant completely abolishes PG binding.

Chapter 4

Bioinformatic analysis of YraP

4.1 Introduction

The OM is vitally important and a distinguishing feature of Gram-negative bacteria. The OM is composed of phospholipids and LPS, as well as integral and peripheral outer membrane proteins. Perturbations in the incorporation of these components into the OM may result in defects in the barrier function of the OM. Such defects may confer of the bacterial cell susceptibility to a variety of antibiotics and other chemicals to which intact cells are normally resistant (Ruiz *et al*., 2005). Building on work presented in the previous chapter, and in an effort to identify all the genes required for OM homeostasis, the Henderson group screened the KEIO library for mutants unable to grow in the presence of SDS and Vancomycin. Many of the genes identified in this screen have previously been implicated in OM homeostasis e.g. *bamE* described in the previous chapter. However, several genes of unknown function were identified. One such gene was *yraP*. The major aim of the remainder of this project was to obtain structural and functional data of the protein YraP from *E. coli*.

YraP is a predicted lipoprotein and analyses of the signal sequence, which lacks an aspartate residue at position +2, indicated that it is most likely an OM-associated lipoprotein targeted by the Lol pathway (Onufryk *et al*., 2005, Narita *et al*., 2004). The Onufryk *et al*. 2005 study demonstrated that *yraP* is a non-essential gene as Δ*yraP* strains grew at a range of temperatures in LB and M9 media. However, this study revealed that ΔyraP strains lack normal OM homeostasis since they were sensitive to SDS. Any disturbance in the integrity of OM homeostasis results in σ^E-mediated alteration in the expression of specific genes. Several investigations

demonstrated that *yraP* is regulated in a σ^E-dependent manner in both *E. coli* and *S. enterica* (Onufryk *et al*., 2005, Dartigalongue *et al*., 2001, Skovierova *et al*., 2006), an observation consistent with a role for YraP in maintaining OM homeostasis.

Defects in OM homeostasis normally arise from incorrect assembly of LPS, OMPs, lipoproteins or phospholipids into the OM. *yraP* was identified as a factor involved with OMP biogenesis based on a proteomic study of *Shigella dysenteriae* (Kuntumalla *et al*., 2011). However, this study was based on statistical analyses of the expression of certain proteins and did not provide any biochemical or biophysical evidence to support this observation. Interestingly, Onufryk *et al*. 2005 demonstrated that a double knockout strain of *yraP* and the periplasmic chaperone *surA* produced a synthetic lethal phenotype, an observation which is consistent with a role for YraP in OMP biogenesis. However, several investigations of the OMP assembly and the BAM complex have failed to identify YraP as a partner in OMP assembly leaving open the possibility that YraP plays a role in one of the other biogenesis pathways for OM components.

To aid in the study of YraP, and to provide direction for subsequent experimental chapters, this chapter presents a variety of bioinformatic analyses used to gain insight into the function and the possible structure of the YraP protein.

4.2 Results

4.2.1 YraP is a predicted OM lipoprotein

The *E. coli* K-12 *yraP* gene encodes a 191 amino acid lipoprotein, which has an N-terminal signal sequence. Using PSORT, predicts that YraP is located in the periplasm, with a predictive score of 9.76 (**Figure 4.1A**). Additionally, LipoP, also predicts a cleavage site between residues G18 and C19, further suggesting that it is a lipoprotein. In addition to this, LipoP also indicates that the +2 residue of YraP is a valine and that the most likely enzymatic cleavage is performed by Signal Peptidase II, suggesting that YraP is an OM localised protein (**Figure 4.1B**).

A

PSORTb Results (Click here for an explanation of the output formats)

```
SeqID: lcl|ECDH10B_3323|yraP|79276046 hypothetical protein
  Analysis Report:
    CMSVM-              Unknown                    [No details]
    CytoSVM-            Unknown                    [No details]
    ECSVM-              Unknown                    [No details]
    ModHMM-             Unknown                    [1 internal helix found]
    Motif-              Unknown                    [No motifs found]
    OMPMotif-           Unknown                    [No motifs found]
    OMSVM-              Unknown                    [No details]
    PPSVM-              Unknown                    [No details]
    Profile-            Unknown                    [No matches to profiles found]
    SCL-BLAST-          Periplasmic                [matched 15599622: conserved hypothetical protein[Pseudomonas aeruginosa PAO1]]
    SCL-BLASTe-         Unknown                    [No matches against database]
    Signal-             Non-Cytoplasmic            [Signal peptide detected]
  Localization Scores:
    Cytoplasmic             0.00
    CytoplasmicMembrane     0.06
    Periplasmic             9.76
    OuterMembrane           0.06
    Extracellular           0.11
  Final Prediction:
    Periplasmic             9.76
--------------------------------------------------------------------------------
```

B

```
# lcl__ECDH10B_3323_yraP_79276046 SpII score=13.0401 margin=4.68731 cleavage=18-19 Pos+2=V
# Cut-off=-3
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Best   SpII    1       1       13.0401
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Margin SpII    1       1       4.68731
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Class  SpI     1       1       8.35279
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Class  CYT     1       1       -0.200913
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavII 18      19      13.0401 # LLLQG|CVAAA Pos+2=V
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  22      23      7.41741 # GCVAA|AVVGT
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  23      24      6.21033 # CVAAA|VVGTA
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  21      22      5.04953 # QGCVA|AAVVG
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  28      29      4.42939 # VVGTA|AVGTK
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  25      26      3.73703 # AAAVV|GTAAV
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  24      25      2.63608 # VAAAV|VGTAA
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  26      27      1.74334 # AAVVG|TAAVG
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  27      28      0.653438        # AVVGT|AAVGT
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  30      31      0.217662        # GTAAV|GTKAA
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  34      35      -0.575451       # VGTKA|ATDPR
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  29      30      -0.841275       # VGTAA|VGTKA
lcl__ECDH10B_3323_yraP_79276046 LipoP1.0:Signal CleavI  20      21      -2.82582        # LQGCV|AAAVV
```

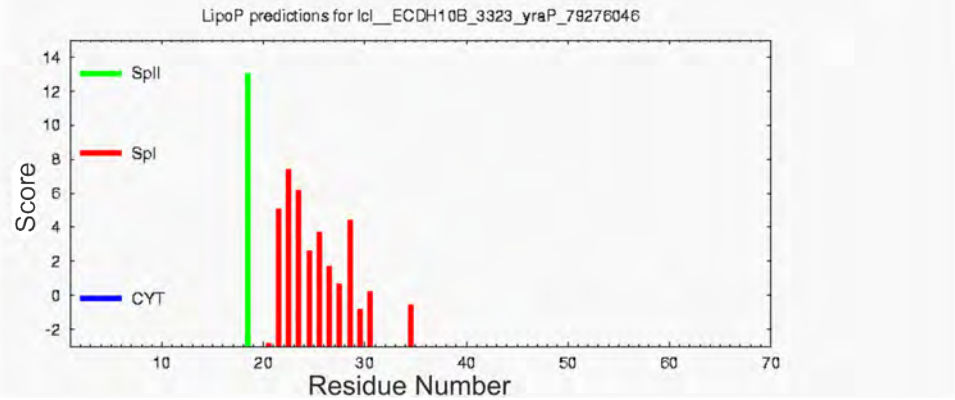
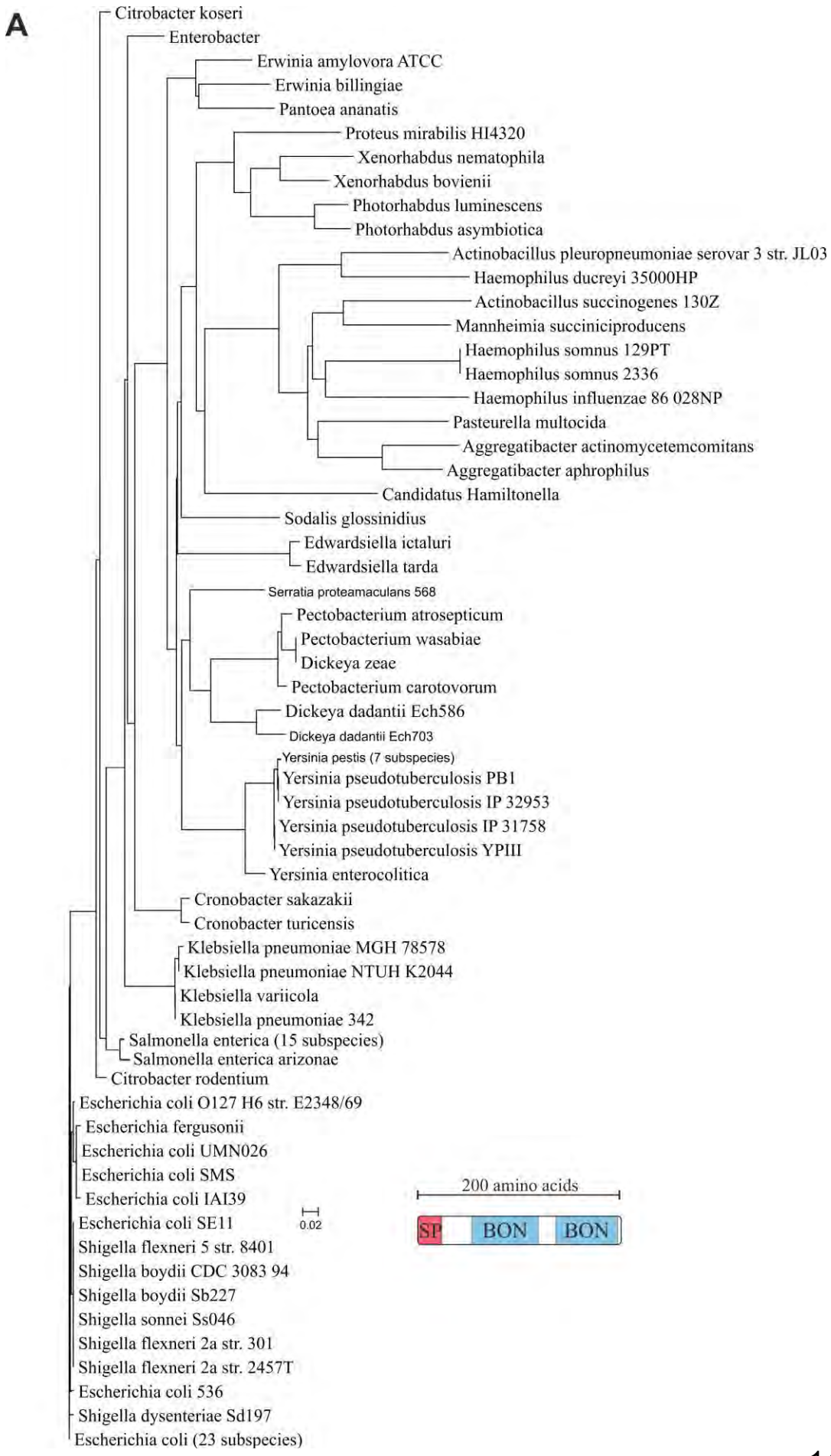


Figure 4.1 YraP is predicted to be an OM lipoprotein. A PSORT suggests that YraP is localised to the periplasm. **B** LipoP prediction of YraP strongly suggests that YraP is an OM localised lipoproten, due to its N-terminal lipobox domain.

4.2.2 YraP is widely distributed in Gram-negative bacteria

In order to identify homologous proteins of YraP in other bacteria, BLAST was used. The major hits produced from the BLAST search highlighted a variety of proteins from other bacteria with differing assigned descriptions, such as hemolysin precursor, putative lipoprotein, BON domain protein and transport protein as the most common descriptions. Because of these descriptions, there is the possibility many of these hits were redundant. To circumvent this problem, a search was performed using Protein Clusters, which would search on the basis of protein architecture as well as sequence similarity (Klimke *et al*., 2009). The result of this search produced 114 sequences, although many sequences were present in different strains of the same bacterial species. The protein sequences were obtained and aligned using Clustal X. A phylogenetic tree was then constructed, however to reduce redundancies, if there were three or more identical sequences from the same species, these condensed into one point (**Figure 4.2A)**. All of the proteins highlighted have a similar architecture and are from bacteria that are members of the gamma proteobacteria family. It should be noted that YraP is found in *Blochmannia*, which has a severely reduced genome, but it is not found in *Buchnera*, which does not produce LPS. This finding suggests that *yraP* may have a role in LPS biogenesis.

Using Pfam, a two BON domain architecture was also searched. 1329 sequences were identified, with sequences in alpha, beta proteobacteria and uncharacterised classes of bacteria, indicating that the architecture extends beyond gamma proteobacteria (**Figure 4.2B**).


A
Citrobacter koseri
Enterobacter
Erwinia amylovora ATCC
Erwinia billingiae
Pantoea ananatis
Proteus mirabilis HI4320
Xenorhabdus nematophila
Xenorhabdus bovienii
Photorhabdus luminescens
Photorhabdus asymbiotica
Actinobacillus pleuropneumoniae serovar 3 str. JL03
Haemophilus ducreyi 35000HP
Actinobacillus succinogenes 130Z
Mannheimia succiniciproducens
Haemophilus somnus 129PT
Haemophilus somnus 2336
Haemophilus influenzae 86 028NP
Pasteurella multocida
Aggregatibacter actinomycetemcomitans
Aggregatibacter aphrophilus
Candidatus Hamiltonella
Sodalis glossinidius
Edwardsiella ictaluri
Edwardsiella tarda
Serratia proteamaculans 568
Pectobacterium atrosepticum
Pectobacterium wasabiae
Dickeya zeae
Pectobacterium carotovorum
Dickeya dadantii Ech586
Dickeya dadantii Ech703
Yersinia pestis (7 subspecies)
Yersinia pseudotuberculosis PB1
Yersinia pseudotuberculosis IP 32953
Yersinia pseudotuberculosis IP 31758
Yersinia pseudotuberculosis YPIII
Yersinia enterocolitica
Cronobacter sakazakii
Cronobacter turicensis
Klebsiella pneumoniae MGH 78578
Klebsiella pneumoniae NTUH K2044
Klebsiella variicola
Klebsiella pneumoniae 342
Salmonella enterica (15 subspecies)
Salmonella enterica arizonae
Citrobacter rodentium
Escherichia coli O127 H6 str. E2348/69
Escherichia fergusonii
Escherichia coli UMN026
Escherichia coli SMS
Escherichia coli IAI39
Escherichia coli SE11
Shigella flexneri 5 str. 8401
Shigella boydii CDC 3083 94
Shigella boydii Sb227
Shigella sonnei Ss046
Shigella flexneri 2a str. 301
Shigella flexneri 2a str. 2457T
Escherichia coli 536
Shigella dysenteriae Sd197
Escherichia coli (23 subspecies)
0.02
200 amino acids
SP
BON
BON

B

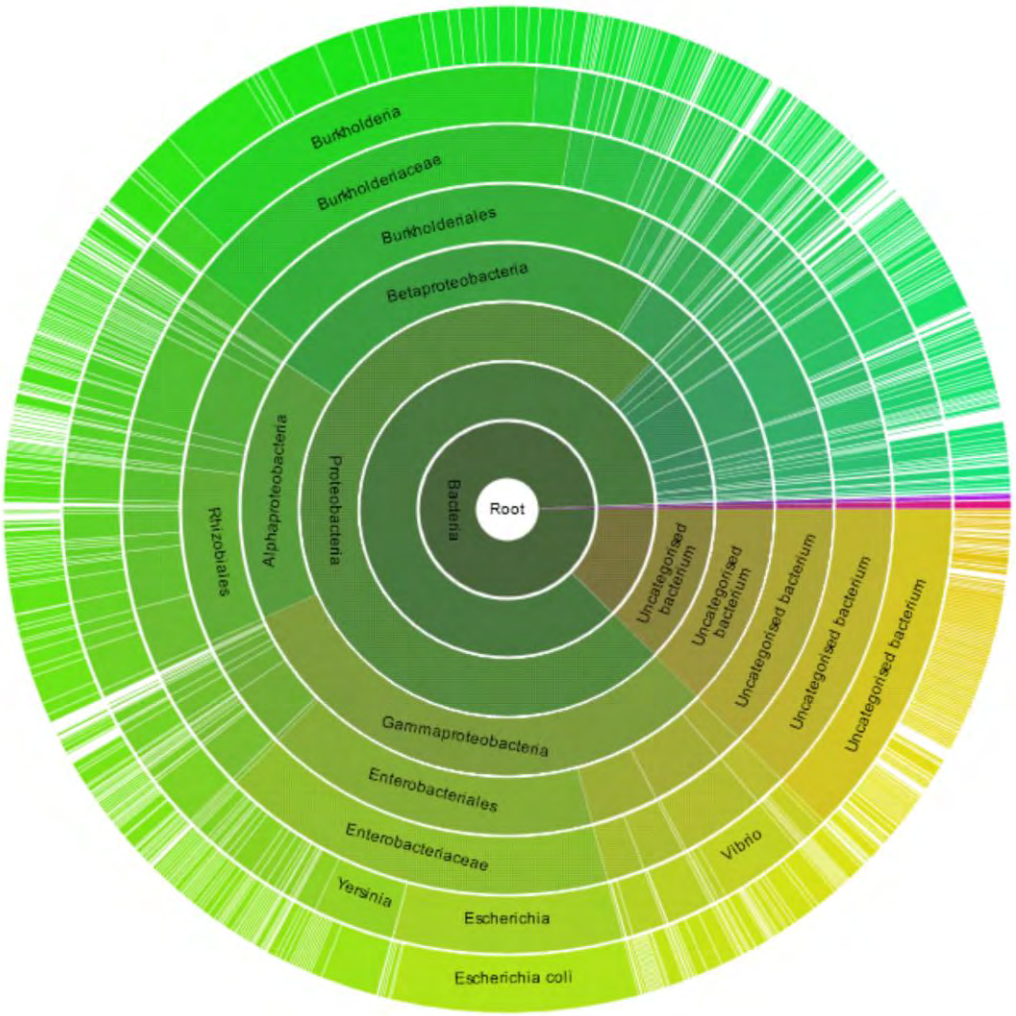


Figure 4.2A Phylogenetic tree of YraP homologues in gamma-proteobacteria from 114 identified sequences. **B** A sunburst model displaying the prevalence of 2 BON domain architecture in different organisms from 1329 identified sequences. These proteins may not direct homologues of YraP.

4.2.3 The YraM-P module is conserved in Gram-negative bacteria

To determine whether homologs of *yraP* were in the same genomic location, Xbase was used to examine genes flanking yraP. The identification of genes around *yraP* can yield some potential information about its function, such as belonging to an operon with genes of similar function. This approach revealed that YraP is flanked by genes of either unknown or putative functions (**Figure 4.3)**.

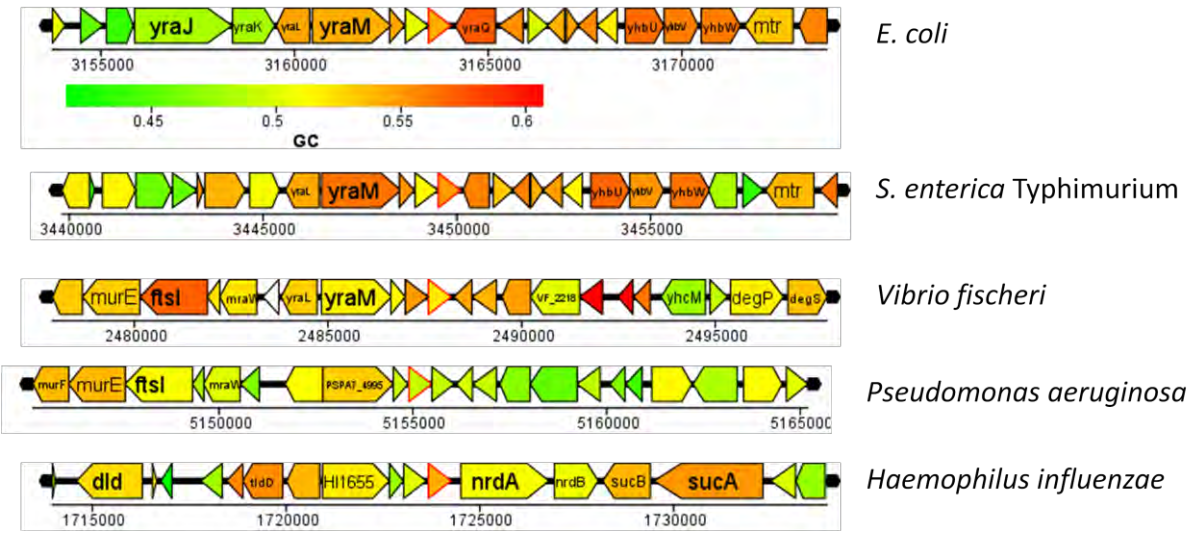


Figure 4.3 Synteny of the flanking genes of YraP. A search was performed using Xbase and aligning the synteny of *E. coli* to other bacteria. YraP is highlighted by a red outline.

yraL (known as *rsmI*), functions with *rsmH* (also known as *mraW*) in methyl modification of 16S rRNA (Kimura and Suzuki, 2009).

yraH-K encodes a fimbrial operon. The absence of the locus from *S. enterica* indicates it was acquired after the divergence of *Salmonella* from *E. coli* (Korea *et al*., 2010).

yraM (*lpoA)* is an OM lipoprotein has been shown to bind and aid the function of the penicillin binding protein PBP1A, one of the major peptidoglycan synthases (Typas *et al*., 2011).

yraN is a gene of unknown function, but has a domain similar to Type II endonucleases/Holliday junction resolvases and is thus implicated in DNA binding.

yraO (also known as *diaA*) is a DnaA binding protein that is required for the initiation of chromosomal replication during the cell cycle (Keyamura *et al*., 2007).

yraQ is predicted to be a permease. However, it is absent in *Salmonella* and is unlikely to function with YraP.

yraR is a predicted nucleoside-diphosphate-sugar epimerase.

yhbO is a member of the DJ-1/ThiJ/Pfp1 superfamily, which includes proteases and chaperones. It has a central role in stress regulation as its absence in *E.coli* leads to increased sensitivity to thermal, UV, pH and oxidative stresses (Abdallah *et al*., 2007).

yhbP is a gene of unknown function. *yhbQ* is a predicted endonuclease, *yhbS* is predicted acyltransferase and *yhbT* is a predicted lipid carrier protein.

mtr is an IM protein that functions as a tryptophan permease (Sarsero and Pittard, 1995).

yraP homologues might perform similar function in other organisms. If this is the case, the synteny of the genes would be conserved, indicating that they are true homologues of YraP. Conservation could also suggest that the surrounding genes may be needed for YraP function. To test this, synteny alignments were performed using Xbase against *S. enterica*, *Vibrio fischeri*, *Pseudomonas aeruginosa* and *Haemophilus influenzae*.

The synteny of these genes is fairly well conserved between these bacteria. Most of the genes are present in similar locations, although *yraQ* is shown to be further downstream. As the functions of the flanking genes of YraP are varied, it is difficult to assign a function using this analysis. However, this analysis has shown that *yraL*, *yraM*, *yraN*, *yraO* and *yraP* are maintained as a module.

4.2.4 YraP contains 2 BON domains

To analyse the protein architecture and predict functions, the protein sequence was placed through online tools including BLAST (Basic Local Alignment Search Tool) the Pfam (protein family) database and SMART (Simple modular architecture research tool).

All three programs indicate that YraP contains a two domain protein architecture, with the presence of two identical BON (bacterial and OsmY nodulation) domains within the sequence (**Figure 4.4A**). The BON domain superfamily is found in a variety of periplasmic osmotic protection proteins and is predicted to have the ability to bind to membrane phospholipids (Yeats and Bateman, 2003).

Table 4.1 shows proteins that contain BON domains in *E. coli*. ZP_04003228.1 contains a single BON domain, but it is a hypothetical predicted protein, so it may not be functionally relevant. However the other 3 proteins, YraP, OsmY and YgaU have had some characterisation and are confirmed to contain BON domains with some additional architecture. YraP contains 2 BON domains, OsmY a predicted 6 BON domains and YgaU has a slightly more complex architecture, with a BON domain and LysM domain. These three proteins are all predicted to be located in the periplasm, which could be a common characteristic amongst proteins containing BON domains.

Table 4.1 Proteins in *E. coli* that contain BON domains.

Protein	Length (Amino acids)	Architecture	Functional Information	Subcellular Location
YraP	191	2 x BON	Uncharacterised lipoprotein	Periplasm
OsmY	201	6 x BON	Osmoregulation protein	Periplasm
YgaU	149	BON + LysM	Peptidoglycan binding protein	Periplasm
ZP_04003228.1	80	BON	Hypothetical protein	Unknown

The protein OsmY in *E. coli* is the major member of the BON domain superfamily. OsmY has a role in homeostasis under conditions of hyperosmotic stress (Yim and Villarejo, 1992). OsmY has a two BON domain architecture similar to YraP. Using ClustalX, a multiple sequence alignment was performed between YraP and OsmY (**Figure 4.4B**). Despite the fact that YraP is a lipoprotein and OsmY is a periplasmic protein, the alignment demonstrated that they exhibit 29% identity and 50% similarity.

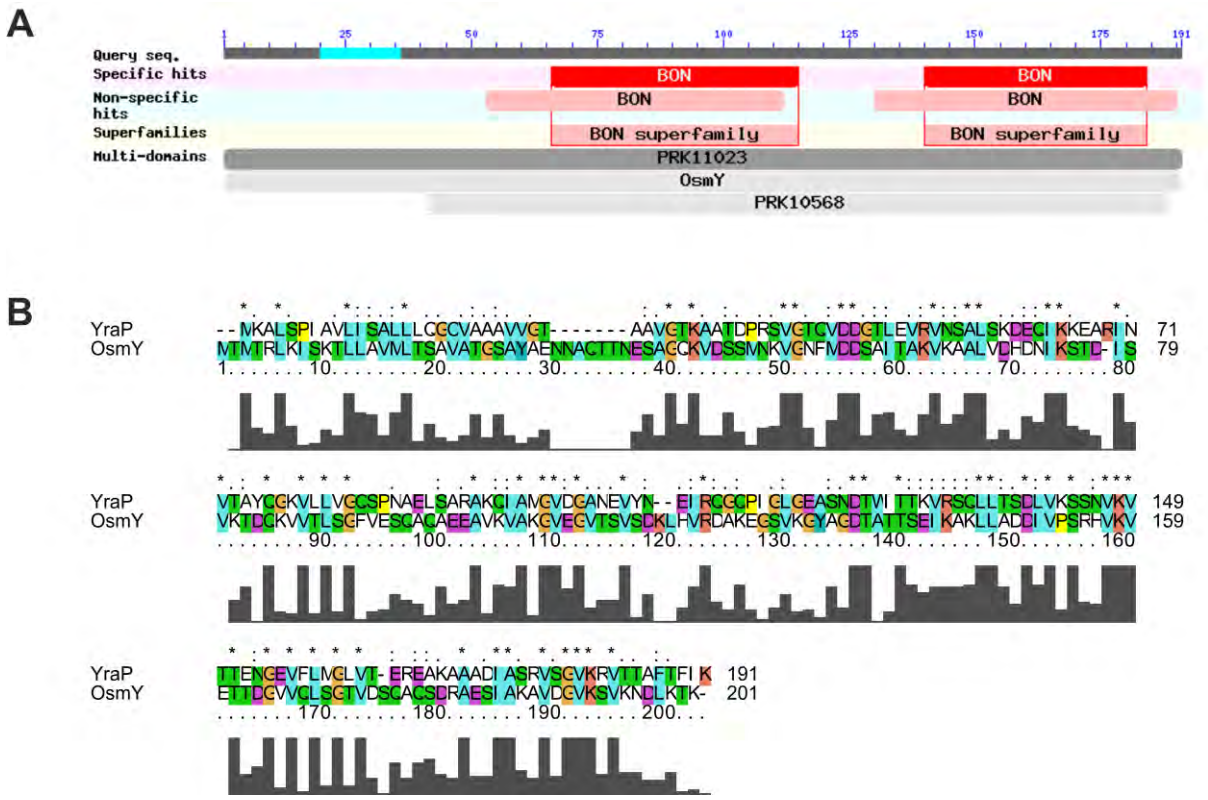


Figure 4.4 A Protein architecture of YraP predicted using BLAST. YraP contains a 2 BON domain architecture and is similar to OsmY. **B** Clustal X multiple sequence alignment using YraP and OsmY protein sequences. Regions of homology are coloured with a bar chart below the sequence to gauge the strength of homology. 29% identity and 50% similarity was calculated from this alignment.

The Onufryk *et al*. 2005 study also highlighted the similarity between the proteins and postulated that YraP and OsmY might perform redundant functions. However, they were able to produce a Δ*yraP* Δ*osmY* strain and found that this was strain was viable. This result, along with the sequence divergence between YraP and OsmY, suggest that YraP may have a different function that is not related to osmoregulation.

4.2.5 Possible interactions with *yraP*

One possible way to gauge the function of *yraP* is to determine its potential interactions with other proteins. To predict this, a STRING search was performed (**Figure 4.5A**).

Notable proteins are YraL, YraM, YraN and DiaA (YraO) whose gene locations are present immediately upstream of *yraP*. YfiO (BamD) is also highlighted as possible partner, which is also a σ^E regulated gene and an essential lipoprotein that is a member of the Bam complex, along with Skp, which a periplasmic chaperone that is also involved in OMP biogenesis. LptC is a lipoprotein which is involved in LPS transport to the OM. Also highlighted are hypothetical protein DcrB, ubiquinone biosynthesis enzyme UbiD (Zhang and Javor, 2003) and lipoprotein DapX, which belongs to the same family as BamC.

Overall, these predictions suggest that YraP is involved in some aspect of OM biogenesis, with OMP or LPS pathways being the most likely candidates. It should be noted that these predictions are not based on experimental data, but rather from neighbourhood and co-occurrence interactions.

The String search also found YraP and its potential interacting partners in other organisms (**Figure 4.5B**).The major organisms where all the partners were present were in proteobacteria, which confirms the accuracy of the phylogenetic tree. With the other organisms, some of the interacting partners are absent. This can suggest that distant homologues to YraP may have a different function in

other organisms or that they are not direct homologues to YraP and have been identified erroneously.

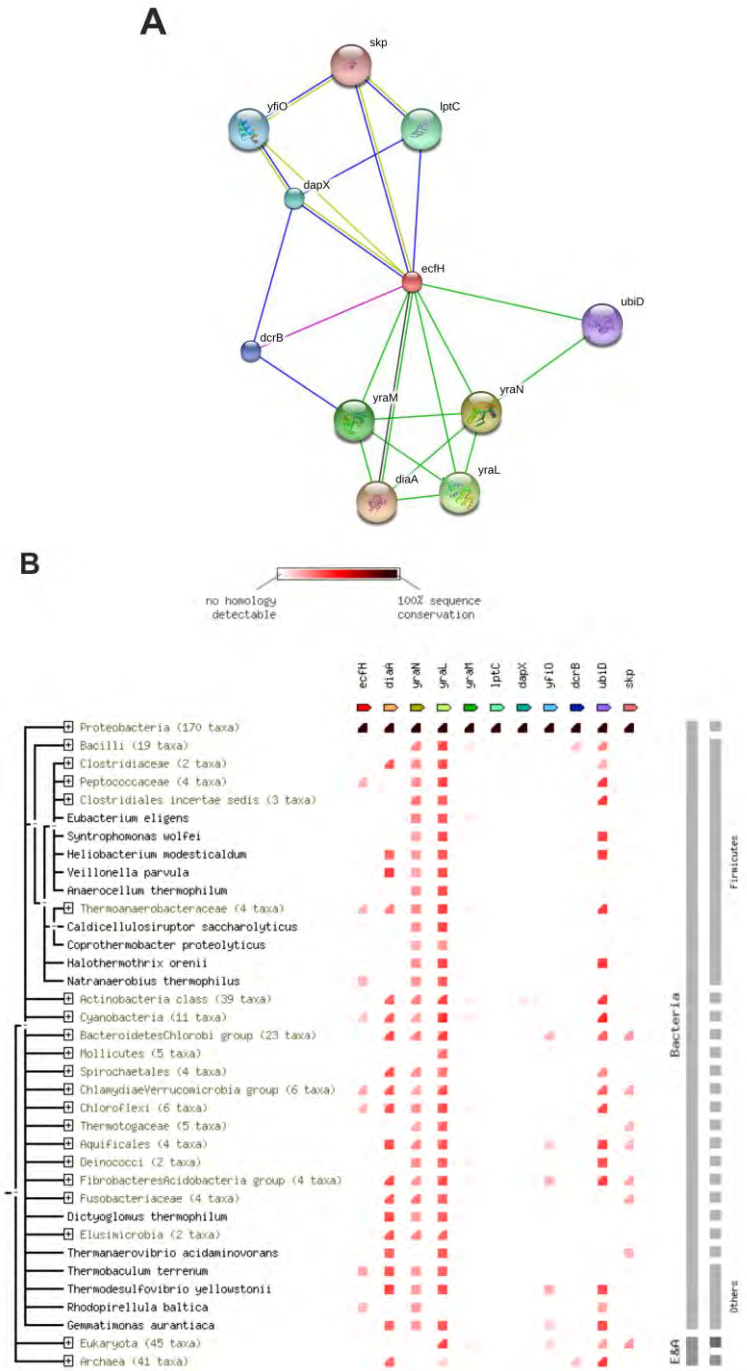


Figure 4.5A String analysis with YraP (labelled with the synonymous name EcfH). String highlights the possible interactions to other protein partners. **B** Co-occurrence of these possible partners in other organisms. Due to the absence of some of these partners, the homologues to YraP may not perform a similar function.

4.2.6 Secondary structure prediction of YraP

Psipred was performed to predict possible secondary structure elements of YraP. To aid the predictions, the signal sequence (residues 1-19) were omitted, as this region will be cleaved during lipoprotein maturation.

The Psipred prediction indicates that YraP contains a mixture of α-helical and β-strand content, with a total of 5 α-helices and 6 β-strands (**Figure 4.6)**. Two of the β-strands are very short, only consisting of two residues, which may indicate some inaccuracy in the prediction. There is also a large extended loop region between residues 90 and 109 (99 and 118 in YraP), which could serve as the linker between the 2 BON domains of YraP.

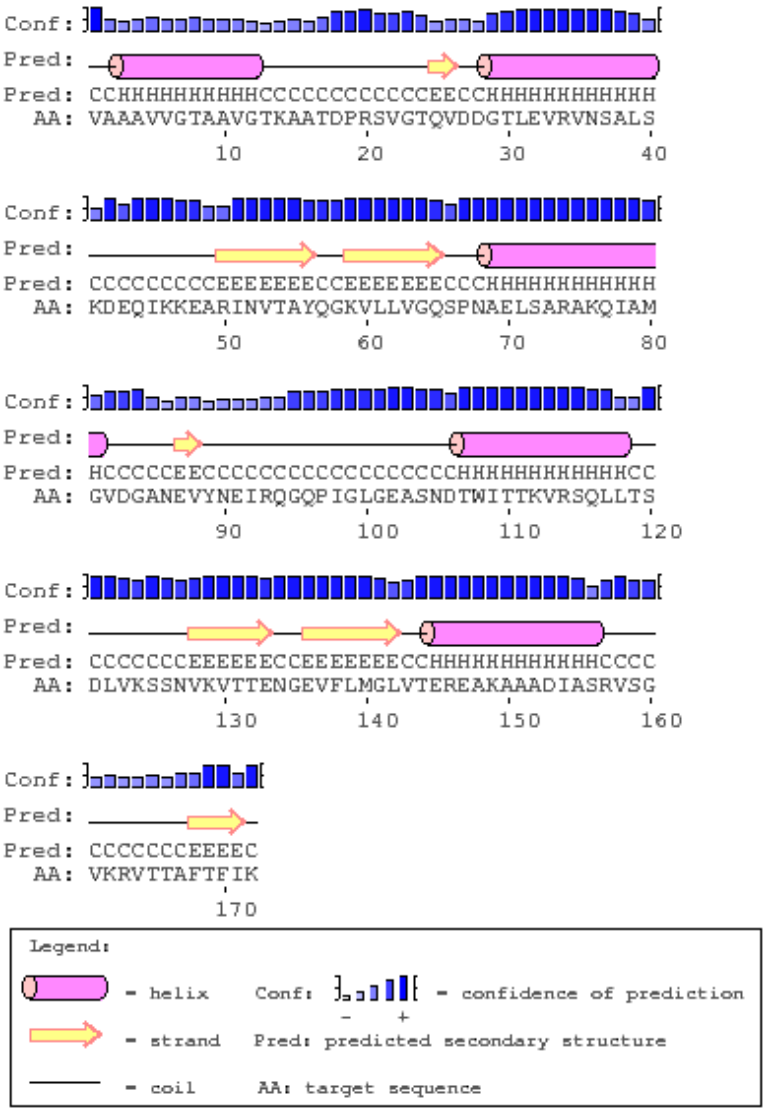


Figure 4.6 Psipred prediction of YraP lacking the signal sequence. A total of 5 α-helices and 6 β-strands have been predicted, with a possible linker region between residues 90-105 separating the two possible domains.

4.2.7 Tertiary structure predictions

Two different approaches were used to predict the structure of YraP from the protein sequence, Phyre2 and I-Tasser.

The Phyre2 prediction compares the input sequence to known homologues. If these homologues have known structures, Phyre2 uses these structures as a template to guide the fold of the protein. The output is a model of the protein, with a percentage sequence coverage to the homologues.

The model produced from the Phyre2 prediction has 97.5% confidence and uses 64% coverage (110 residues) to another protein, rv0899 in *Mycobacterium tuberculosis* (Teriete *et al*., 2010) (**Figure 4.7A**). This model shows an aligned structure, with α-helices on one side and β-strands on the other. 3 α-helices and 6 β-strands have been identified, although a 2 domain architecture is not immediately apparent.

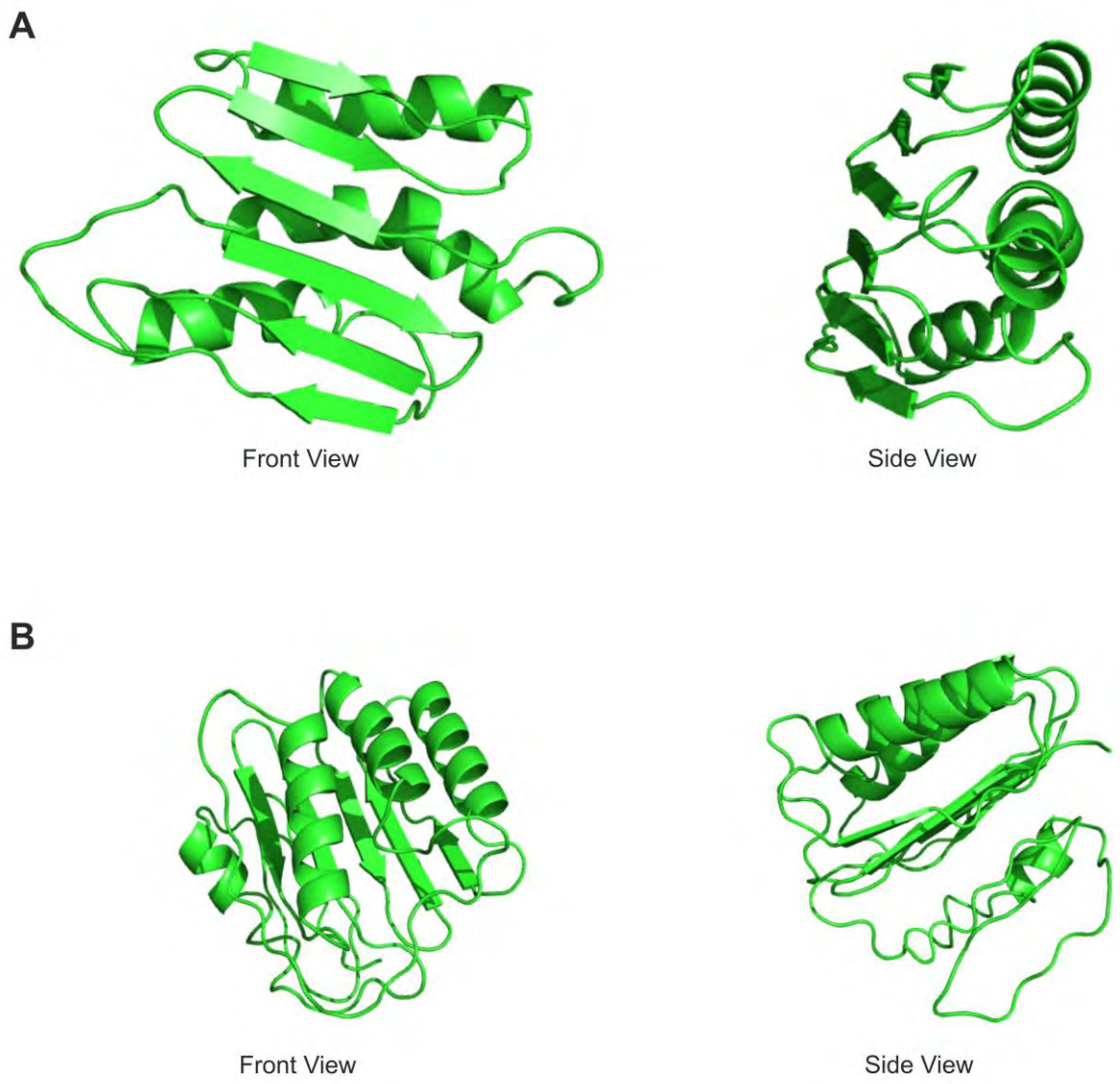


Figure 4.7A Prediction of the YraP structure using Phyre2. A flat, compact fold is produced, with a total of 3 α-helices and 6 β-strands organised as 2 'layers'.
B Top scoring model of YraP from the I-Tasser prediction. A total of 4 α-helices and 9 β-strands are identified, with the overall fold organised as 3 'layers'. The number of secondary structure elements from each of the predictions differ from the Psipred prediction and the 2 domain architecture is not apparent from either prediction.

I-Tasser (Roy *et al*., 2010), performs a similar prediction, in that it uses homology modelling for structure prediction. There are also further algorithms used, which refine the hydrogen bonding and remove potential steric clashes. Finally, another set of structure alignments are performed, which try to determine a potential function of the protein. A total of 5 models are produced from the prediction.

The top-scoring model from the I-Tasser predication is shown in **Figure 4.7B.** A mixed α-helix and β-strand protein is produced, which agrees with both the secondary structure and the Phyre2 predictions. However, 4 α-helices and 9 β-strands are produced, which are arranged in 3 'layers'. The 2 domain architecture is also not apparent on this model.

4.3 Discussion

The use of bioinformatic tools has allowed some predictions to be made on the possible function and structure of YraP. The use of BLAST and Pfam have identified that YraP contains a 2 BON domain architecture that may be able to bind to phospholipids. Using phylogenetic analysis, homologues of YraP are highly conserved in gamma proteobacteria, but Pfam has also identified other 2 BON domain proteins in alpha, beta and epsilon proteobacteria. Although the sequence is too different to conclusively say they are functional homologs, the architectural similarity strongly suggests a common function. Thus, the protein YraP is widespread in bacteria, indicating an important role in proteo-bacteria.

Both experimental and predictive data suggest that YraP has a role in OM biogenesis. Onufryk *et al*. 2005 have demonstrated that Δ*yraP* strains are sensitive to SDS, indicating a compromised OM barrier. STRING has predicted possible partners that are involved in LPS and OMP biogenesis, although there is a lack of experimental data behind these predictions. However, further proof that YraP may have a role in OM biogenesis, is from the phenotypic landscape screen, which shows that Δ*yraP* cells show sensitivity to various conditions such as vancomycin and SDS. Strains that harbour knockouts or depletions of genes that are involved in OM biogenesis also show sensitivity to similar conditions.

The structural predictions of YraP have produced some conflicting results. The three predictions used, Psipred, Phyre2 and I-Tasser have demonstrated that YraP contains both α-helices and β-strands. The actual number of α-helices and β-strands differ significantly and with the 3D models produced with Phyre2 and I-

Tasser, the 2 domain architecture is not observable as well as any similarity between the potential domains. The variations produced between the tertiary structures of the models means that these predictions are not accurate. To accurately determine the structure of YraP will have to be performed independently of these models.

(Nichols *et al*., 2011) performed a high throughput screen of various strains of *E. coli* harbouring gene knockouts and depletions and monitored their susceptibility to a variety of compounds. Using this approach, they were able generate a phenotypic landscape of mutants and were able to group genes that showed similar sensitivities to the various compounds as genes that have similar functions.

It has previously been reported that gene knockouts of members of the Bam complex show sensitivity to compounds such as vancomycin, rifampicin and SDS (Sklar *et al*., 2007a). We searched phenotypic landscape database to screen Δ*yraP* mutants and compared them to knockouts or depletions of genes involved in OM biogenesis. **Figure 4.8** shows the top scoring susceptibilities of *yraP*, *bamE* and *lptD* mutant strains. SDS, cholate and vancomycin are common between the three strains, suggesting that YraP may have a role in OM biogenesis.

YraP		BamE		LptD	
DEOXYCHOLATE-2.0%	-18.1322380	VANCOMYCIN-50	-26.1279640	MALTOSE	-2.5949150
SDS-0.5%	-11.1285990	VANCOMYCIN-20	-23.5454310	PEROXIDE-0.1	-1.8570700
SDS0.5%/EDTA0.1	-9.4105700	CHLOROPROMAZINE-24	-19.1909270	AMPICILLIN-1.0	-1.7837430
SDS-1.0%	-7.3577200	VANCOMYCIN-20	-13.9741440	CISPLATIN-100	-1.7439120
BILE-2.0%	-6.5991160	CEFSULODIN-24.0	-10.5114780	CLARYTHROMYCIN-1.0	-1.6778160
CHOLATE-2.0%	-6.2024070	SPIRAMYCIN-20	-10.1252570	PEROXIDE-0.5	-1.5433620
BILE-1.0%	-5.8270440	AZITHROMYCIN-1.0	-9.6137830	ACETATE	-1.5191030
SDS-2.0%	-5.4885640	TRITONX-0.2%	-9.0348350	POLYMYXINB-4.0	-1.5131930
DIBUCAINE-1.2	-5.3052370	TRITONX-0.03%	-8.4874650	N-ACETYLGLUCOSAMINE	-1.4413090
PH4	-5.2326330	VANCOMYCIN-10	-8.2660770	TAUROCHOLATE-0.5%	-1.3730000
SDS-3.0%	-5.2089240	BACITRACIN-200	-7.1966030	BACITRACIN-300	-1.3466060
VANCOMYCIN-50	-4.1573740	CEFSULODIN-18.0	-6.9537350	CHOLATE-0.1%	-1.3003980
CHOLATE-1.0%	-4.0965680	BLEOMYCIN-2.0	-6.7783680	VANCOMYCIN-10	-1.2988700
BILE-0.5%	-3.7074200	CLARYTHROMYCIN-10.0	-6.3354440	AMOXICILLIN-1.0	-1.2851370
SDS-4.0%	-3.1464870	BLEOMYCIN-0.5	-6.2427490	CEFACLOR-1.0	-1.2843850
CHIR090-0.075	-3.0511970	EGTA-1.0	-6.1933140	METHOTREXATE-25	-1.2825020
EGTA-2.0	-2.9549810	DIBUCAINE-1.2	-5.9679260	FUSIDICACID-20	-1.2424180
EGTA-1.0	-2.9036480	ERYTHROMYCIN-5.0	-5.9491700	CISPLATIN-50	-1.2306070
CECROPINB-0.1	-2.9020740	THEOPHYLLINE-100	-5.7893960	INDOLICIDIN-0.1	-1.1930530
HYDROXYUREA-5.0	-2.8938730	AZITHROMYCIN-1.0	-5.7228420	DOXORUBICIN-1.0	-1.1926840
CHIR090-0.05	-2.6617640	CLARYTHROMYCIN-10.0	-5.7189750	AMOXICILLIN-1.5	-1.1889830
NACL-300	-2.6156390	SPIRAMYCIN-20	-5.5357380	THIOLACTOMYCIN-1	-1.1515920
THIOLACTOMYCIN-50	-2.5332590	VANCOMYCIN-50	-5.4589390	GLUCOSAMINE	-1.0968310
ACRIFLAVINE-10	-2.5002230	BACITRACIN-300	-5.3039390	RADICICOL-1	-1.0716510
VANCOMYCIN-20	-2.3855400	ERYTHROMYCIN-10.0	-5.2049990	A22-2.0	-1.0604410
DIBUCAINE-0.8	-2.3624910	RIFAMPICIN-2.0	-5.1575940	HIGHCOBALT-0.1	-1.0376940
PROPIDIUMIODIDE-50	-2.2908400	SDS1.0%/EDTA0.5	-5.1496630	ACTINOMYCIND-10.0	-1.0032160
HIGHCOBALT-0.1	-2.1989080	ACTINOMYCIND-10.0	-5.1399540	45C	-0.9523370
NOREPINEPHRINE-100	-2.1604460	PH4	-5.1300460	CECROPINB-0.3	-0.9503570
PHLEOMYCIN-0.5	-2.1544520	CLARYTHROMYCIN-5.0	-4.8415950	TUNICAMYCIN-3.0	-0.9221310
DEOXYCHOLATE-0.5%	-2.1280330	BLEOMYCIN-0.1	-4.8229070	SDS-0.5%	-0.9012280

Figure 4.8 Top scoring sensitive conditions for YraP, BamE and LptD mutants.
Conditions are highlighted in grey with scores highlighted in blue.

Chapter 5

Structure determination of YraP by NMR

5.1 Introduction

One of the major aims of this project was to determine the structure of YraP. A solved structure could potentially aid in determining its function, by comparing its structural elements to other proteins with similar folds. Additionally, knowledge of the structure can allow other experiments to be considered, such as ligand or drug binding, based on key structural elements of the protein. Furthermore, as a 2 BON domain architecture protein has not yet been solved, a structure of YraP would contribute greatly in the study of this family of proteins.

We used NMR to solve the solution structure of YraP. We also hoped to use an X-ray crystallography approach, but we were unable to generate protein crystals of YraP. In order to appreciate the use of NMR as a structural biology tool, the relevant NMR theory will be discussed in the following sections.

5.1.1 NMR as a structural biology tool

Nuclear magnetic resonance (NMR) spectroscopy is one of the major techniques used in structural biology. In the case of protein biology NMR can be used to generate atomic resolution structures in solution. Although the resolution of protein structures generated by NMR can be ideal, depending of the data quality, generally cannot match the resolution of protein structures generated from X-ray crystallography. Furthermore, there is also size limitation with NMR, with either proteins or complexes of proteins that are generally above 30 kDa being unsuitable, due to the complexity of signals generated (Yee et al., 2005).

Although atomic resolution may be a factor in choosing an approach for structural studies, there are some notable advantages in the use of NMR over X-ray crystallography. X-ray crystallography requires the generation of protein crystals, which can be a rate limiting step. Furthermore, with crystallography, there is the presence of potential artefacts from crystallisation due to the conditions used, which may bias proteins to adopt conformations that may not be physiologically relevant. NMR is able to remedy this by working with proteins under more natural solution conditions. This has the additional benefit that dynamics, conformational changes and potential interactions can be studied with NMR that may not be amenable in X-ray crystallography. However, the cost of generating isotope labelled protein samples as well as the cost involved in using NMR spectrometers can be a hindrance in choosing NMR over X-ray crystallography. It should be noted that NMR is a non-destructive technique and complete recovery of the protein is possible after experimentation.

5.2 NMR Theory

5.2.1 Physical properties of the nucleus

NMR is able to occur because the nuclei of certain atoms possess a property known as spin. Spin is characterised by the nuclear spin quantum number *I*, which can take integer and half-integer values (1/2, 1, 3/2, 2 etc.) and is caused by the presence of uneven protons or neutrons in certain nuclei. Nuclei which have zero spin ($I = 0$), i.e. an even number of protons and neutrons, are not amenable to NMR (Carrington and Machlachlan, 1967).

The property of a nucleus having spin enables it to have angular momentum (P). As all nuclei have charge, the effect of spin on this charge means that these nuclei have a weak magnetic field associated with them, described as a magnetic moment (μ) (Farrar, 1987). The relationship between μ and P is described by the magnetogyric ratio (γ), which is calculated with the formula

$\gamma = \mu/P$ rad T^{-1} s^{-1}

The magnetogyric ratio is essentially a constant that indicates how 'magnetic' a particular nucleus is.

As nuclei with spin have their own magnetic moment, when they are placed in an external magnetic (B_0) field, they experience a torque that forces them into precession about the axis of the external magnetic field (**Figure 5.1**). This motion is referred to as Larmor precession and occurs at the Larmor frequency (v). v is directly related to the strength of B_0 as well as the γ of the nucleus. The Larmor frequency can be described with the formula

$v = \gamma B_0 / 2\pi$ Hz

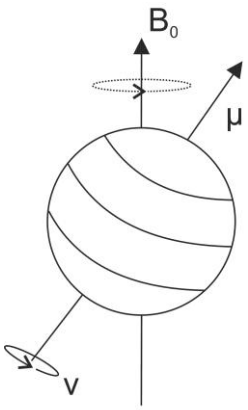


Figure 5.1 The effect of a nucleus in an external magentic field. The nucleus is a charge particle with magnetic momentum. When placed in a an external magnetic field (B_0), the nuclues experiences a torque, which causes it to precess around the axis of the magnetic filed at the Larmor frequency v.

In a magnetic field, nuclei with spin are able to take up 2*I* + 1 possible orientations with respect to the axis of the magnetic field. This is due to angular momentum associated with spin being quantized and can only take on a restricted range of values. In the case of nuclei with *I* = 1/2, such as ^{1}H, ^{13}C and ^{15}N, there are two possible orientations; parallel and anti-parallel to the field (**Figure 5.2**). There are energy levels associated with these orientations, with the lower energy (α) state being the parallel orientation and the higher energy (β) state being the anti-parallel orientation (Fukushima and Roeder, 1987).

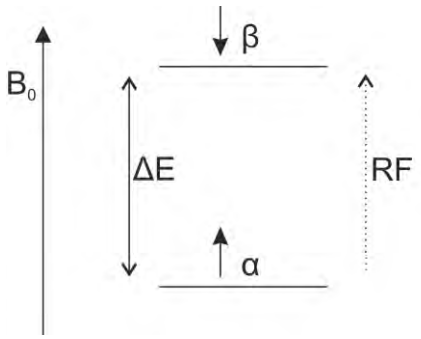


Figure 5.2 Allowed energy states of spin 1/2 nuclei in an external magnetic field. Spin 1/2 nuclei can only adopt one of two orientations in a magnetic field (B_0), a parallel lower energy state (α) and an anti-parallel higher energy sate (β) as indicated by the arrows. Nuclei are able to move between states by absorbance of radiofrequency (RF) radiation, or loss of energy, that corresponds to the ΔE between the states.

In the absence of B_0, there is no preference for either of the states as they are essentially energetically equivalent. The effect of B_0 causes these nuclei to adopt either an α or β state. As NMR is a spectroscopic technique, it requires both absorption and emission of electromagnetic radiation. Nuclei in the α state can become promoted by the absorbance of energy as well as nuclei in the β state can lose energy by relaxation and fall to the α state (Farrar, 1987). The frequencies that are required to cause these transitions directly correspond to Larmor frequencies of the nuclei and can be described with this formula:

$\Delta E = h\nu = h\gamma B_0 / 2\pi$

Where h = Planck's constant

NMR essentially records the population difference between the α and β states. At equilibrium under the effect of B_0, the populations can be described with the Boltzmann distribution equation

$$N\alpha / N\beta = e^{\Delta E/RT}$$

Where N = Number of nuclei in a particular spin state, ΔE = Energy of transition, R = Gas constant and T = Absolute temperature

The Boltzmann equation indicates that ΔE for the transition between the α and β states is small. It also indicates that there will be more nuclei in the lower energy α state than in the β state. Furthermore due to the small energy transition between the states it means n the population differences between the α and β states are also small only in the order of 1 to 1000. This makes NMR a relatively insensitive technique, therefore requiring high sample concentrations, often in the order of the mg range for proteins. Sensitivity can be improved with increasing B_0 field strength, although this does not remedy the sample cost.

5.2.2 Vector model of NMR

The previous description of NMR has described how a single nucleus behaves in a magnetic field. The vector model is used to describe how a population of nuclei behave in a magnetic field, as well as the effect of what happens when a pulse of EM radiation is added which ultimately causes the generation of spectra from NMR.

In the case of a set of I = 1/2 nuclei, the effect of an external magnetic field will cause them to align and precess around the axis, designated as the $\pm z$ axis.

According to the Boltzmann distribution, there will be an excess of nuclei in the α state. After cancelling out the nuclei in the β state from the α state, there will be a resultant magnetic vector that is from the excess of nuclei in the α state (Leyden, 1977). This vector is orientated parallel to the applied field and is designated as a bulk magnetisation vector *M*. As this magnetisation exists on the +*z* axis, it is referred to as longitudinal magnetisation (**Figure 5.3**).

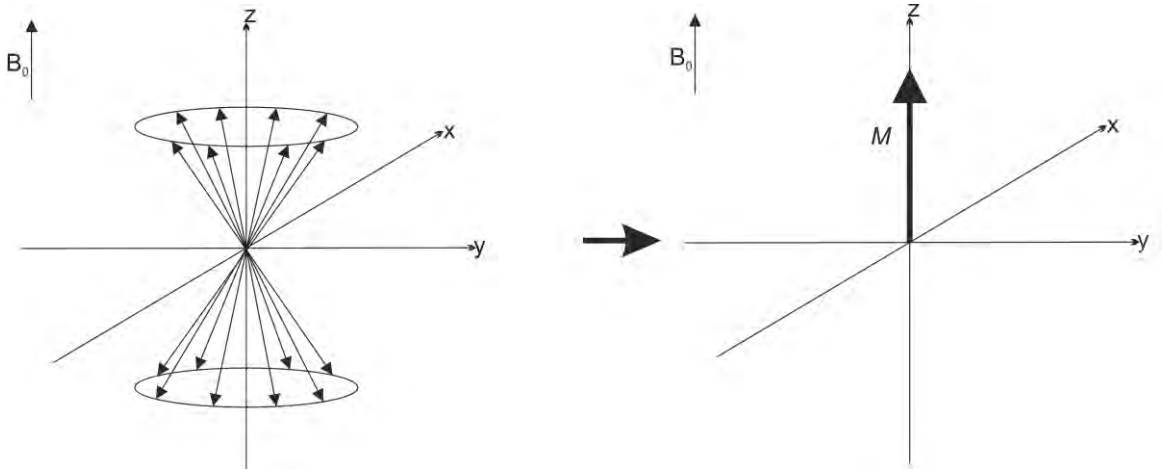


Figure 5.3 Nuclei behave as a population with a net magnetic vector. Nuclei orientate around B_0 (z-axis), with random orientations around the xy-plane. As the xy-magnetisation is random, this cancels out. The Boltzmann distributions states that there is an excess of spins present in the parallel α orientation than in the β orientation. After cancelling these out, there is a bulk magnetic vector *M*, which is parallel to B_0.

In NMR, a second magnetisation vector is applied as an RF pulse to perturb the equilibrium set up in the above case. Using this second field (B_1), which is perpendicular to B_0 and applied via a transmitter coil, causes M to move from the +*z* axis onto the ± x/y axis, causing transverse magnetisation (**Figure 5.4**). It should be noted that precession is still occurring at the Larmor frequency. In order to visualise this as static vectors on the vector model, a rotating frame of

reference is used, whereby the axes used are rotating at the same frequency as the nuclei.

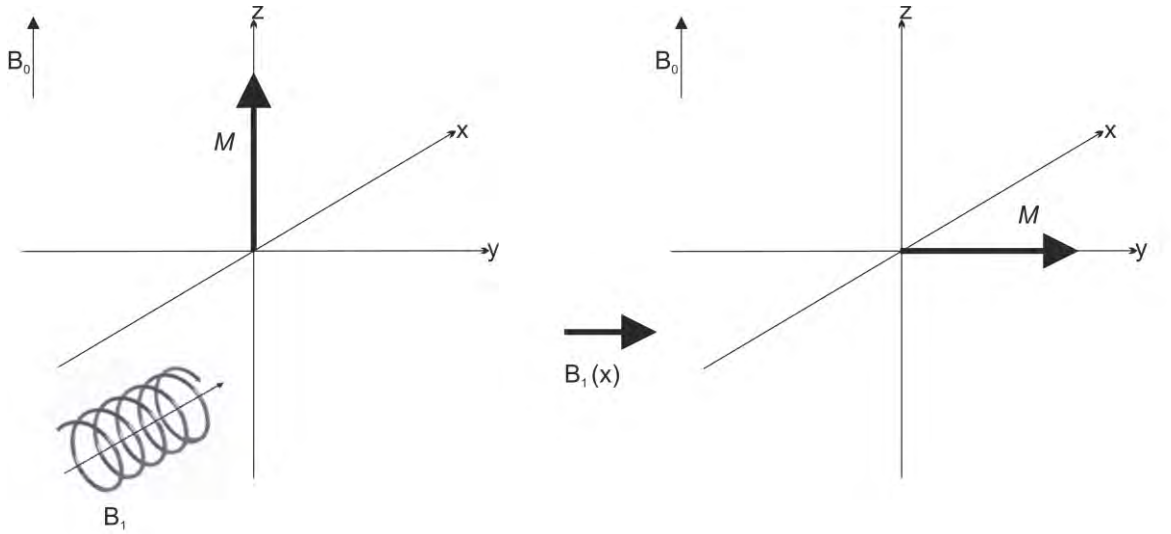


Figure 5.4 The effect of an orthogonal field on *M*. Applying a pulse in the x plane, causes *M* to move into the y-axis. This shifts longitudinal magnetisation to transverse magnetisation, which can be detected in NMR.

The effect of B_1 causes the transverse magnetisation vector to oscillate around the ± *x*/*y* axis, depending on the strength of field and pulse length of B_1. This oscillation causes the production of a current in a coil surrounding the sample and gives rise to the NMR signal (Hayes *et al*., 1988). Only transverse magnetisation is able to produce this current. Magnetisation will not stay in the ± *x*/*y* axis indefinitely, but will return back to the +*z* axis, by the loss of energy in a process known as relaxation (discussed below). As this oscillating current decays with time, it produces a free induction decay (FID) signal (**Figure 5.5**). The FID is composed of many different oscillating signals from different nuclei decaying over time (Stilbs, 1987). Fourier transformation of the FID yields the different precession frequencies of the nuclei which gives rise to the NMR spectrum.

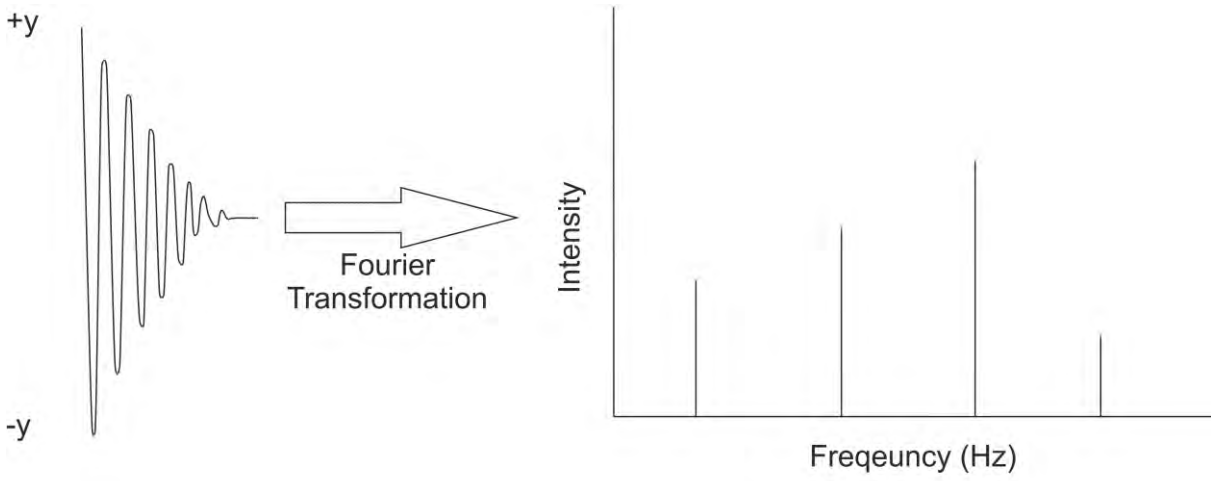


Figure 5.5 Transformation of FID. The generation of transverse magnetisation will give rise to decaying sinusoidal signal called a FID (Free Induction Decay), which is composed of many different oscillating nuclei. Using Fourier transformation to express a time domain into a frequency domain, = the individual nuclei can be expressed as separate frequencies with intensities.

5.2.3 Relaxation

There are two relaxation processes that occur in NMR, longitudinal relaxation (T1) and transverse relaxation (T2). Both of these processes contribute to the decay of the FID.

T1 relaxation, or spin-lattice relaxation, is the process which restores the magnetisation vector M to the ±*z* axis which is brought about by the restoration of equilibrium population differences between the α and β states. T1 relaxation can occur by a number of mechanisms, the most important is the dipole-dipole interaction, which is a direct through-space magnetic interaction between two

spins. As spin nuclei tumble in solution, each nucleus exhibits different rotational and vibrational motions. This can change the relative orientation of dipolar coupled-spins, which in turn changes the magnetic field experienced at one nucleus due to fluctuating fields from its neighbouring spin nucleus. These fluctuating fields can provide a suitable means of relaxation for the spin nuclei. The energy lost by the nuclei through this process is done so in the form of heat to the surroundings, known as the lattice. Ultimately this leads to the increase in magnetisation in the ±*z* axis (Gong and Hornack, 1992).

T2 relaxation describes the loss of bulk magnetisation in the ± *x*/*y* plane. This does not relate to the restoration of the equilibrium populations, making it distinct from T1 relaxation. The major cause of this is the loss of phase coherence in the ± *x*/*y* plane.

Following a 90^{o} pulse which shifts M into the ± *x*/*y* plane, the spin nuclei are precessing at the Larmor frequency on this plane around the *z* axis. The vector model is a macroscopic representation of many spins and assumes that they are aligned or phase coherent at this point. However individual spins in the ± *x*/*y* plane may experience slightly different local magnetic fields which can be brought about by small variations in the local magnetic environment or from a non-homogenous static magnetic field. This effect causes the spins in the ± *x*/*y* plane to precess at different frequencies around the *z* axis. Ultimately, the spins are evenly distributed about the ± *x*/*y* plane and there is no resultant M. The spread of Larmor frequencies correspond to a spread of frequencies in the final

NMR spectrum, which display the individual resonances with defined linewidths rather than narrow lines. The linewidth of a peak can be described as

$\Delta v_{1/2} = 1 / \pi T_2$

T2 relaxation does not require the loss of energy to the lattice, but energy is retained by the transfer between spin nuclei that have identical v. These nuclei are able to undergo spin flipping whereby spin = 1/2 nuclei interchange between α and β states. Although there are no net changes in the population differences between the states, the average life time of a spin in the active state is reduced. This effect can cause line broadening and occurs ultimately until there is no magnetisation left in the ± *x*/*y* plane (Li and Hornack, 1994).

5.2.4 Chemical Shifts

When a spin active nucleus is placed in an external magnetic field such as B_0, it precesses about the axis of the field at its Larmor frequency. The nucleus is also surrounded by electrons, which generate magnetic fields of their own that act in opposition to the B_0 field. This in turn influences the resultant magnetic field that the nucleus experiences. This effect from the electrons is known as shielding and influences the precession rate of the nucleus and the frequency required to stimulate it into resonance. Differences in the chemical environment can modify the electron and distribution about the nuclei. In the case of proteins, there are many instances of different chemical environments for nuclei and hence a variety of different resonances can be seen for the nuclei in the protein. These

differences are what constitute the chemical shift (δ) differences that are observed and relates the dependence of energy transitions in nuclei to the local chemical environment. As chemical shifts are essentially the frequency differences between nuclei in different chemical environments, δ is measured using a relative scale. δ for any resonance can be defined as:

$\delta = v - v_{Ref} / v_{Ref}$, where v is the frequency resonance of the nuclei and v_{Ref} is the frequency resonance of a reference compound.

The frequency differences between nuclei are very small, in the order of 10^{-6} and the units of δ are dimensionless. The values of δ are multiplied by 10^6 to give rise to the ppm (parts per million) scale. In addition to making the numbers easier to work with, an added advantage to using ppm is that the chemical shift of any resonance in ppm is independent of the strength of the applied magnetic field. This allows δ measured from one spectrometer to be directly compared to those of another spectrometer (Freeman, 1988).

5.2.5 Chemical shift perturbations

As the chemical shift essentially gauges the local chemical environments of nuclei, this can be exploited to examine possible further changes in the local environments. Changes such as ligand binding to a protein can cause drastic alterations in the local chemical environments of the residues involved in ligand binding. This in turn affects the chemical shift of the the nuclei which is known as chemical shift perturbations (CSP). With HSQC experiments (discussed below), which measure ^{1}H and ^{15}N chemical shifts, CSPs can be observed for both nuclei. As CSPs associated with ^{1}H nuclei aare generally much greater than

those associated with ^{15}N nuclei, weighting is required to address this imbalance and normalise any observed CSPs. The standard formula used for this purpose is:

$$\Delta\delta_{Overall} = [(\Delta\delta_H \times 500)^2 + (\Delta\delta_N \times 50.7)^2]^{0.5}$$

5.2.6 HSQC

The Heteronuclear Single Quantum Coherence (HSQC) is a 2D NMR experiment. ^{15}N-HSQC spectra display the chemical shifts of all of the amide (NH) groups present in the protein via the ^{1}H and ^{15}N dimensions (Bodenhausen and Ruben, 1980).

In a ^{15}N-HSQC, magnetisation is transferred from the proton to the nitrogen atom in an NH group. Through chemical shift evolution of the nitrogen atom, the magnetisation is transferred back to the H atom which is then detected (**Figure 5.6**).

In theory, each amino acid would produce one peak, from their backbone amide group, with the exception of prolines, which lack an NH group due to the imino side chain. Other amino acids such as Glutamine, Aspargine, Tryptophan giving additional resonances from their side chains containing amide groups. However, due to uneven magnetisation, or overlapping peaks, not all of the predicted resonances may be seen.

The HSQC is a crucial experiment for structure determination using NMR, as well as being an important tool in gauging various properties of the protein, such as the folding state or binding to ligands. The dispersion of the HSQC can also give other details of the protein, such as the presence of secondary structure groups, such as α-helices and β-sheets.

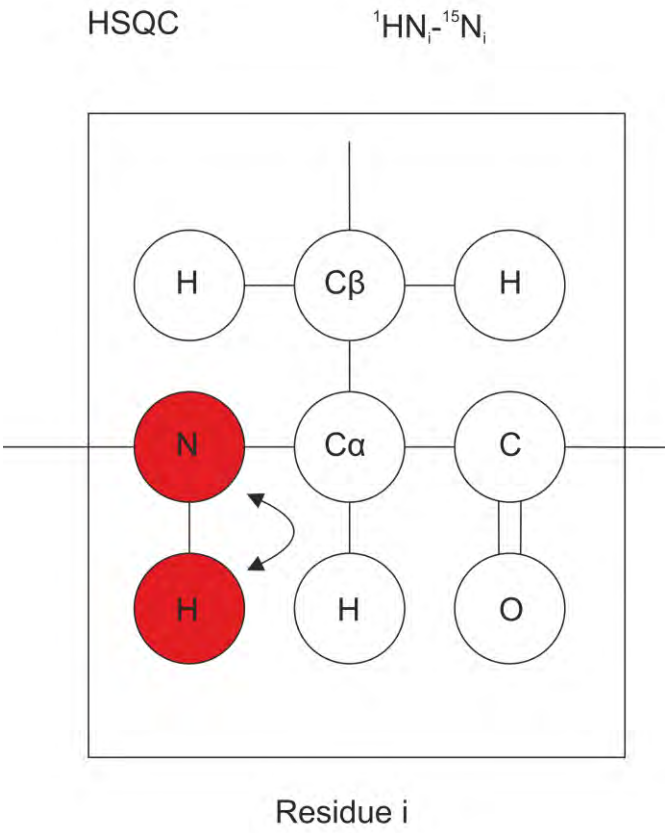


Figure 5.6 Magnetisation pathway of the HSQC. Magnetisation is transferred between the ^{15}N and ^{1}H group of the i residue, which gives rise to a 2D spectrum.

5.2.7 Sequential backbone assignment

The HSQC only gives a 2D representation of the protein, by displaying amide (NH) resonances via the Proton and Nitrogen dimension. In order to determine which peaks belong to which residue in the protein, a full assignment must be

carried out. To do this, further NMR experiments are performed alongside the HSQC. These are 3D experiments that utilise the same Proton and Nitrogen dimension, with a third dimension for Carbon resonances. By correlating the common Proton and Nitrogen dimensions and examining along the Carbon dimension, it is possible to find resonances that belong to the CA (α carbon), CB (β carbon) and CO (carbonyl) of a residue. To further the assignment process, the 3D experiments are performed in a pair wise manner, with magnetisation being transferred from an NH group to the different backbone carbon atoms. Doing this allows one experiment to display the resonances of the current residue (termed i) and the other experiment with the resonances of the current residue (i) and the preceding residue (i-1). Provided that the primary sequence of the protein is known and by following this connectivity between the experiments, and relaying the new information to the Proton and Nitrogen resonances, it is possible to assign the protein using this method. Further help is found from certain amino acids displaying characteristic resonances for their atoms. This piece of information can be vital in ensuring that assignments are correct for a particular region of the protein.

5.2.8 HN(CA)CO and HNCO

This pair of experiments allows backbone assignment of CO groups in the protein. The HNCO experiment transfers magnetisation through the HN_i, via a process termed 'J-coupling' to the preceding CO_{i-1} residue. Therefore, this experiment should only produce a single resonance corresponding to the CO group of the i-1 residue at a defined Proton and Nitrogen ppm. In terms of signal to noise ratio, the HNCO is the most sensitive experiment, which can be used to

confirm correct backbone assignments, especially in regions of strong overlap on the HSQC.

The HNCO is run alongside the HN(CA)CO. The HN(CA)CO transfers magnetisation through the CA_i and CA_{i-1} to both the CO_i and CO_{i-1} (**Figure 5.7**). However, due to stronger coupling between the N_i - CA_i, than the N_i - CA_{i-1}, the CO_i resonance is generally more intense than the CO_{i-1} resonance.

On overlaying the spectra of these two experiments, the CO resonance that is common to both spectra corresponds to the CO_{i-1} and the individual resonance in the HN(CA)CO is the CO_i.

5.2.9 HNCA and HN(CO)CA

This pair of experiments allows sequential backbone CA assignments. The HNCA allows magnetisation transfer to both the CA_i and CA_{i-1}, subsequently producing two CA resonances. As the coupling between the N_i - CA_i is greater than the N_i - CA_{i-1}, the CA_i generally appears to be more intense out of the two resonances.

Any CA resonance information in the HNCA needs to confirmed by the HN(CO)CA. This experiment transfers magnetisation through the CO_{i-1}, only producing a single resonance belonging to CA_{i-1}. The resonance produced from this experiment correlates to the backbone NH_i, which allows the correct assignments of CA groups. On overlaying the spectra of these two experiments,

the CA resonance that is common to both spectra corresponds to the CA_{i-1} and the individual resonance in the HNCA is the CA_i.

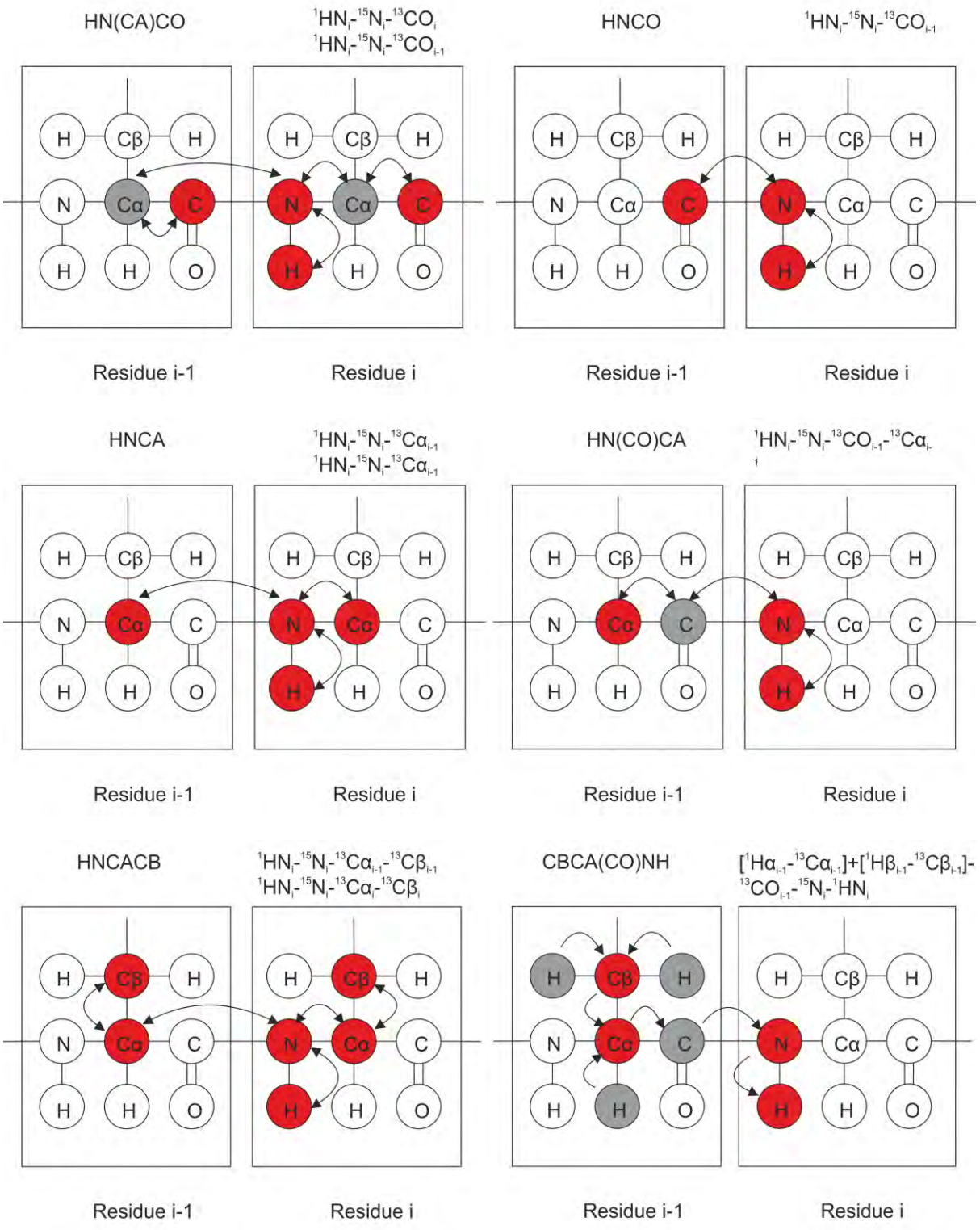


Figure 5.7 Magnetisation pathways of the HN(CA)CO and HNCO experiments. Atoms coloured in red are detected in the spectra and atoms coloured in grey indicate that magnetisation is transferred through them. Experiments are performed in a pairwise manner, to display the atoms in the i and i-1 residues.

5.2.10 HNCACB and CBCA(CO)NH

Another pair of experiments that are essential for full backbone assignments are the HNCACB and CBCA(CO)NH. These experiments display the resonances of both CA CB groups. These experiments are very useful in that they act as a backup for the CA assignments from the HNCA and HN(CO)CA experiments as well as providing the side chain CB assignments. Rather than only performing the HNCACB and CBCA(CO)NH experiments the two sets (CA and CB experiments) are normally performed in conjunction as the resolution of the CB experiments is far lower as a larger ppm range is required to be sampled. As CA resonances tend to be over a smaller ppm range than the CB resonances, it makes accurate CA assignment impossible.

With the CBCA(CO)NH, magnetisation is transferred from the HB_{i-1} and the HA_{i-1} to the CB_{i-1} and CA_{i-1} respectively. From here, it is then transferred to the CO_{i-1} finally onto the NH_i group. This experiment should ideally produce two peaks, that correspond to the CA and CB atoms of the i-1 residue.

In the HNCACB experiment, magnetisation is transferred from the NH_i, to both the CA and CB of i and i-1 (**Figure 5.7**). Potentially four resonances should be observed from this experiment, with the CA_i and CB_i having greater intensities

than the CA_{i-1} and CB_{i-1} due to uneven coupling between the groups. Another notable feature about this experiment is that CA peaks are positive whereas CB peaks are negative, due to the phasing used in this experiment. This can aid immensely in differentiating the CA and CB resonances for sequential backbone assignment.

Comparing the spectrum from the HNCACB to the CBCA(CO)NH, the identification of i and i-1 CA and CB groups can be determined. It should be noted that not all of the resonances may be observed, especially in the case of the HNCACB experiment, where four peaks are normally expected. This can be due to uneven magnetisation transfer, dynamics or due to the presence of glycine residues, which lack CB groups.

5.2.11 H(CCO)NH

This 3D experiment allows the identification of the side chain protons on an i-1 residue, using the NH group of the i residue for signal detection. Magnetisation is transferred from the side chain nuclei to their directly attached carbon atoms on the i-1 residue. Magnetisation is then transferred between the carbon atoms, onto the carbonyl atoms and finally to the proton on the NH group of the i residue (**Figure 5.8**). This produces a 3D spectrum containing one nitrogen and two proton dimensions.

5.2.12 (H)C(CO)NH

The (H)C(CO)NH allows the identification of the side chain carbon atoms on an i-1 residue in a similar manner to the H(CCO)NH. The magnetisation is transferred in the same way, except that carbon atoms are detected (**Figure 5.8**).This experiment produces a spectrum with carbon, nitrogen and proton dimensions.

5.2.13 HCCH TOCSY

This 3D experiment is used to check if the side chain carbon and proton assignments are correct. Magnetisation is transferred from the side chain protons to the neighbouring side chain carbon atoms. This is then followed by further carbon magnetisation which is then transferred to other side chain protons (**Figure 5.8**). In theory, every side chain carbon and side chain protons, including CA and HA, should give a resonance. No inter-residue magnetisation occurs in this experiment. All of the resonances observed are for i residue only.

The spectrum from this experiment produces a diagonal, with a Carbon dimension, a direct Proton dimension and an indirect Proton dimension. To analyse this spectrum, one selects the Carbon and Proton ppm of a known proton directly attached carbon resonance as determined from the H(CCO)NH and (H)C(CO)NH experiments. The proton peak from this selection would lie on the diagonal, with the same ppm for both proton axes. By examining along the indirect proton dimension, it is possible to find proton resonances attached to other side chain carbon groups in the same residue. The assignments produced from this experiment are self checking, in that one is able to repeat the process by selecting a directly attached proton to a different side chain carbon and

examine the indirect proton frequency, which would be a different diagonal peak. If the side chain carbon and protons have been assigned correctly, then it should be possible to see the other side chain protons at their correct ppm in the indirect Proton dimension. In theory, this can be repeated for the other side chain protons, with a list of assignments at different carbon ppm, all having the same ppm values along the indirect Proton dimension.

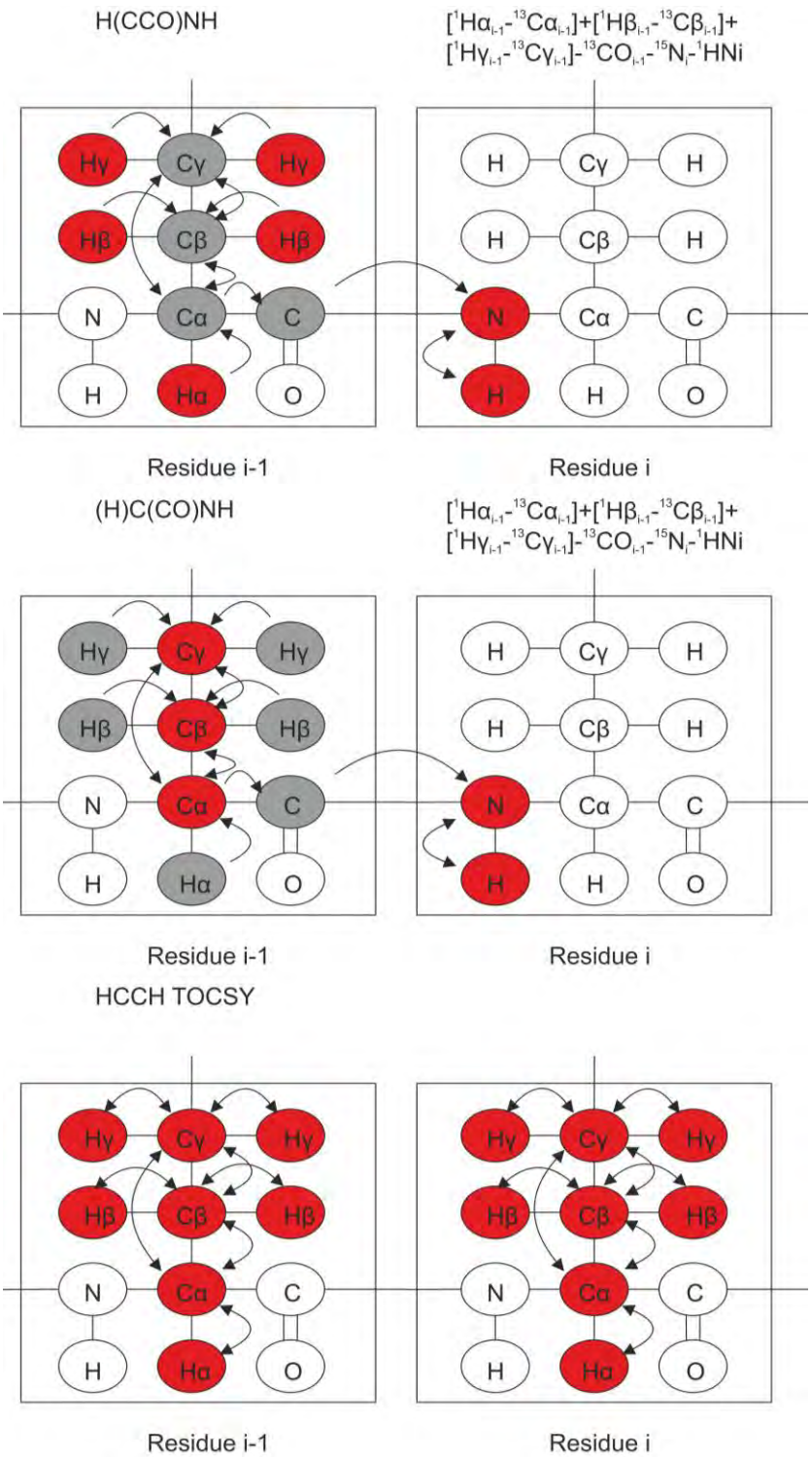


Figure 5.8 Magnetisation pathways of the side chain carbon and protons in the HCCONH and CC(CO)NH experiments. Atoms coloured in red are detected in the spectra and atoms coloured in grey indicate that magnetisation is transferred through them.

5.2.14 NOESY experiments

NOESY (Nuclear Overhauser Effect Spectroscopy) is a powerful tool in generating accurate 3D protein structures from NMR data. NOESY experiments differ from the other NMR experiments in that magnetisation is transferred through space rather than through bonds. If a proton is in close proximity to another proton (within 5 Å), it will give rise to a resonance signal. However, it is also prone to some error as protons that are bonded to the same residue are also in close proximity and can give rise to signals as well (**Figure 5.9**). As a general rule, protons that are 3 bonds away from the proton examined, will display signals. One has to be careful in assigning NOESY spectra, ensuring that potential resonances are not from closely bonded protons. However, previously determined assignments formed from the other spectra can remedy this problem.

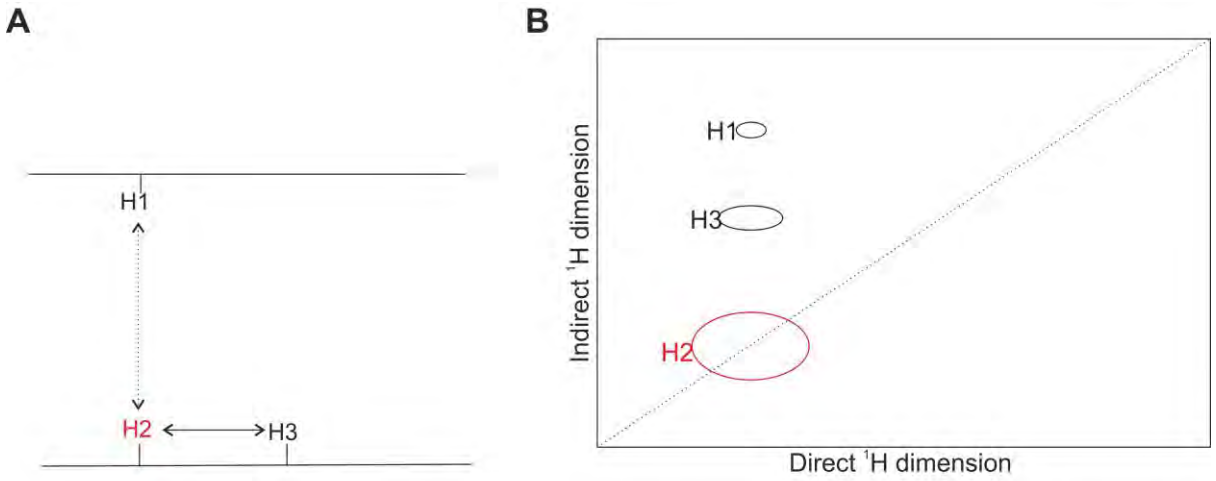


Figure 5.9 Schematic of NOESY interactions. A Probing from proton H2, the NOESY spectrum of this proton will display the neighbouring proton H3 (solid arrow) as well as displaying the resonance of H1 (dashed arrow) as a through space NOE interaction. **B** Schematic spectrum of the NOESY interactions in **A**. At the proton and carbon/nitrogen frequencies of H2, H2 appears as a large peak on the diagonal, indicated in red. The NOE of H3, which is a close through bond interaction, will produce a peak, with a different ppm in the indirect ^{1}H dimension. H1 will also produce a peak with a different ppm in the indirect ^{1}H dimension, but as this is a through space interaction that is further away than H3, its intensity will be much less than H3.

The strength of this resonance is proportional to the distance, with protons in close proximity producing stronger NOE resonances. Generally, protons that are closely bonded will generate strong resonances. By collecting a large number of NOEs, 3D constraints can be generated from this data, which is essential in solving the structure of the protein.

Over the course of this study, three major NOESY experiments were carried out that were used to solve the structure of YraP. A ^{15}N edited NOESY, ^{13}C edited NOESY, and a ^{13}C Aromatic NOESY.

The ^{15}N edited NOESY displays NOE resonances of protons within 5 Å of an amide proton. These are powerful signals for determining the secondary structure of the protein, in that potential residues that are involved in hydrogen bonding (H-bonding) can be identified from the amide resonances given from the backbone amide groups. For cases such as determining a potential α-helix, the n+4 periodicity can be observed in the assignments. In this study, a deuterated ^{15}N edited NOESY was also performed. A notable feature of this experiment is that only NOE resonances between amide protons are observed. This yields a simpler spectrum with fewer peaks, which can be used to determine or confirm secondary structure elements such as β-sheets.

The spectrum produced from this experiment has three dimensions, a Nitrogen dimension, direct Proton dimension and an indirect Proton dimension. This spectrum contains a diagonal, which corresponds to the directly attached proton. Potential NOESYs can be found by examining the indirect Proton dimension.

The ^{13}C edited NOESY is a major experiment for structure determination. This experiment displays NOEs from protons attached to the carbon residue. This is powerful in that potential side chain interactions can be determined, which influences the tertiary structure restraints of the protein therefore determining the overall fold. Like the ^{15}N edited NOESY, the spectrum contains three axes, a Carbon dimension, a direct Proton dimension and an indirect Proton dimension, where potential NOEs are found.

Finally, a ^{13}C aromatic NOESY is performed to identify possible through space interactions from protons attached to carbons in aromatic rings. Aromatic residues can have a function in maintaining the protein core by hydrophobic interactions with other residues, or can be surface exposed and have a role in the function of the protein. The use of aromatic NOEs can determine what the case may be and the data produced from this experiment can yield very helpful restraints for structure determination.

5.2.15 Hydrogen/Deuterium (H/D) exchange

H/D exchanged HSQCs can be a powerful tool in determining residues involved in H-bonding, which can be further used as restraints for structure calculations. H/D exchange works on the principle that if a protonated protein is placed into dueterated conditions, NH groups will exchange into ND groups. However, NH groups that are either involved in H-bonding or are buried inside the protein will take longer to exchange. By choosing an appropriate time for this exchange and performing HSQCs on this exchanged protein, exchanged peaks in the HSQC will not give a signal. Any peaks that give a signal can be presumed to be involved in H-bonding or are not surface exposed. This list of resonances can be crucial in determining the H-bonding network of a protein.

5.2.16 Structure calculations with CYANA

All of the chemical shift values, NOESY peaks and dihedral angle restraints from TALOS can be used for structure determination. In this study, the automated NOESY assignment software CYANA was used for this purpose, which enables quick structure calculations without introducing user bias.

With NOESY data, there is the presence of ambiguous assignments, either due to signal overlap or erroneous identification. As the NOESY data is in essence a list of the distance restraints between atoms and strongly influences the tertiary structure, errors in this data can prove to be problematic for structure calculations and can generate incorrect models of the structure. CYANA is able to circumvent these problems by introducing a weighting function for ambiguous assignments.

A typical CYANA run consists of 7 iterative cycles of ambiguous NOE assignments, with structure calculations using torsion angle restraints. Between the cycles, information of the NOE assignments is transferred between the intermediary 3D structures. During early cycles, CYANA ranks the ambiguous NOE assignments and generates a score list. These low scoring NOEs are initially removed and initial structures are generated only with the higher scoring NOEs i.e. unambiguous NOEs. The structures that are generated at a given cycle are used to guide the NOE assignments in the following cycles. With later cycles, CYANA is able to make better estimates of the ambiguous NOEs and factors in the low scoring NOE assignments as well as removing spurious and erroneous assignments. The final cycle produces the best possible structure with the data that has the least number of NOE violations.

CYANA uses simulated annealing to calculate structures. In brief, heating steps are applied to completely unfold the protein and start from a random state, where torsion angles are treated independently and H atoms are not factored due to the large potential steric clashes. The protein is then cooled, with torsion angle

dynamics, H atoms and van der Waals interactions factored in with weighting criteria, which allows the protein to fold. At the final cycle, the structure that is produced contains the minimum energy and least violations, which is an accurate representation of the fold.

5.3 Results

5.3.1 Expression and purification of YraP for NMR studies

One of the major aims of this project was to determine the protein structure of YraP, which would in turn help in deducing a possible function of the *yraP* gene. As YraP is predicted to be an OM lipoprotein, expression of this lipoprotein could prove to be difficult, with the chance of achieving much lower yields per growth, compared to the expression of soluble proteins, as well as the difficulty of solubilising the protein and using it for structural studies.

In order to circumvent these possible problems, we used a modified pET26b-*yraP* vector. A notable feature of this vector is that its native signal sequence (residues 1-19) of the protein has been removed. Cysteine 19, which is the acylation site for lipoprotein formation, has been substituted to an Isoleucine, which prevents any possibility of acylation. This protein is also C-terminally 6xHis-tagged, which aids purification as well as improving solubility. Finally, to ensure a natural folding environment, the vector contained a PelB signal sequence, which would allow targeting to the periplasm.

Initial expression trials with this vector showed the presence of a thick band at roughly 20 kDa following induction with 1 mM IPTG, corresponding to the estimated size of YraP (**Figure 5.10**). This result suggested that this vector was suitable for protein expression.

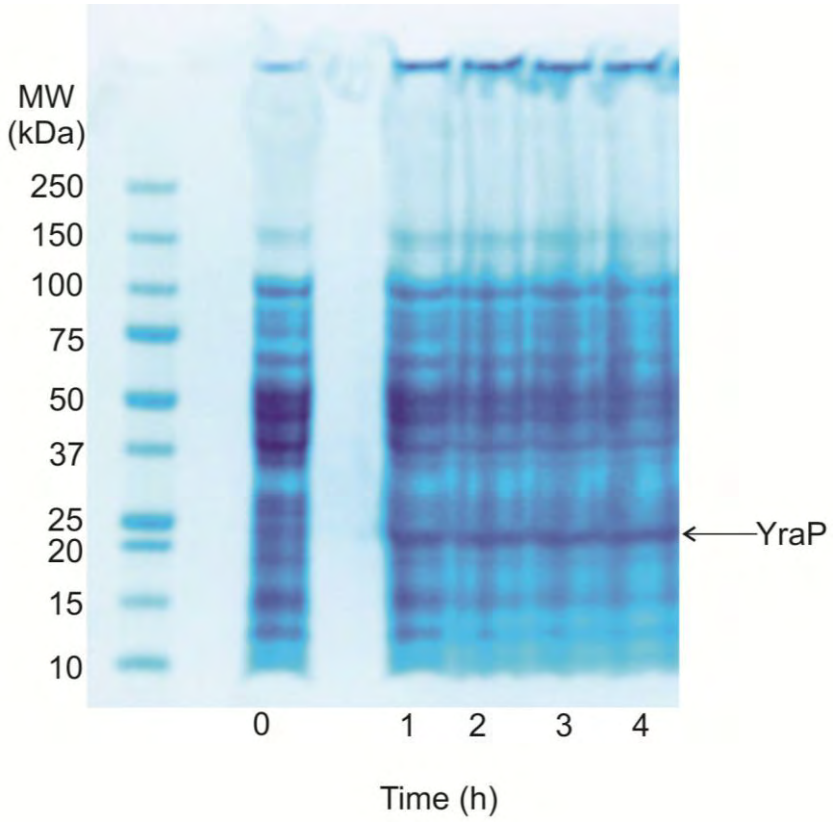


Figure 5.10 Expression trial of the pET26b-*yraP* vector. SDS-PAGE performed with 20 µl of cell lysates of BL21 (DE3) cells containing pET26b-*yraP* from 0-4 hours after induction with 1 mM IPTG. Proteins were detected by Coomasie staining.

Using 2 litre LB growths, with induction and overnight growth at 18^{o}C, YraP was overexpressed. Following lysis and centrifugation, the protein was present in the supernatant fraction, indicating that it was soluble (**Figure 5.11A**). Nickel affinity purification was performed on the soluble fraction, resulting in the recovery of overexpressed YraP in the eluate, removing most of the other contaminating proteins. Further purification using size exclusion chromatography, produced an almost pure sample of YraP (**Figure 5.11B**), although there was another major protein species present. Confirmation of the purity of YraP was determined by performing SDS-PAGE on the collected YraP fraction, which were essentially free from other contaminating proteins (**Figure 5.11C**). After concentration of the

purified protein, concentrations of 1.2 mM at volumes of 2 - 2.5 ml easily obtained, making it ideal for NMR studies.

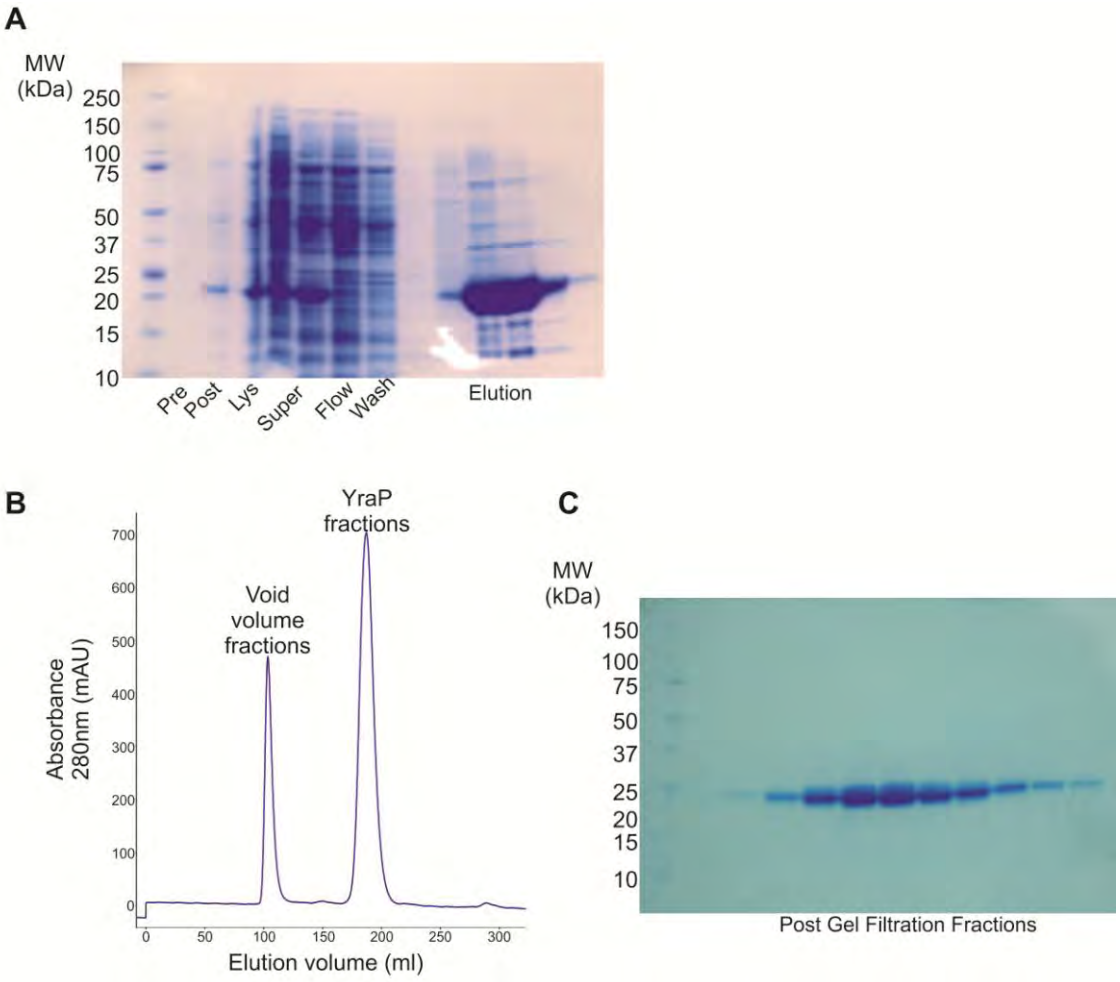


Figure 5.11 A high yield of pure YraP can be expressed and purified. **A** Coomasie stained SDS-PAGE of 20 µl of BL21(DE3) cells with pET26b-*yraP* samples from 2 litres of growth and purification using Nickel affinity chromatography with a 5 ml His Trap HP Column (GE Healthcare). The lanes are Pre (pre-induction), Post (post-induction), Lys (lysate), Super (supernatant after centrifuging out the membranes), Flow (flow-through), Wash (wash-through) and elution from the His-trap purification, with 1 ml fractions collected. **B** Size exclusion chromatography using a Superdex S75 328 ml column (GE Healthcare) with 5 ml of concentrated post Nickel affinity purification. **C** Coomasie stained SDS-PAGE of 20 µl fractions containing YraP after size exclusion chromatography.

To ensure that the protein was folded, 1D proton NMR was performed on the sample (**Figure 5.12**). The spectra showed sharp defined signals as well as signals below 0 ppm, which generally correspond to folded proteins.

In order to perform further NMR studies and to solve the structure, YraP was labelled with NMR active nuclei ^{13}C and ^{15}N. Growths were repeated in M9 minimal media, containing ^{13}C-glucose and ^{15}N-ammonium sulphate to label the protein with these nuclei. Generally these growths take longer or the protein yield can decrease. Over the course of this study, none of these effects have been observed, highlighting the reproducibility of this expression and purification procedure.

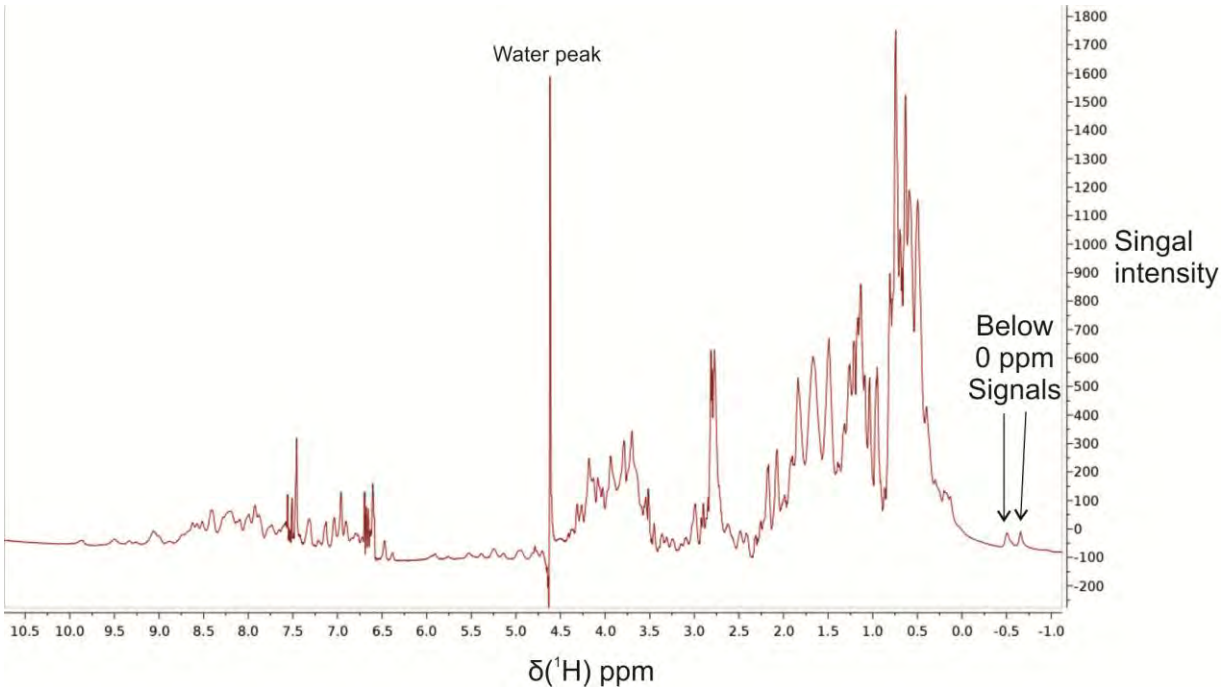


Figure 5.12 1D NMR of YraP indicates that it is folded. 1D NMR performed with 800 µM of YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25°C.

5.3.2 HSQC analysis of ^{13}C, ^{15}N double labelled YraP

The ^{1}H-^{15}N HSQC spectrum of YraP shows well dispersed resonances (e.g. peaks spread over 6-10 ppm in the ^{1}H dimension (**Figure 5.13**).There are a few areas of overlap, seen in the central region between 7.8 - 8.8 ^{1}H ppm and 118 - 125 ^{15}N ppm, which could prove to be problematic for assignment. Also certain resonances indicative of characteristic amino acids can be seen, such as glycines near the top part of the spectrum 105 - 110 ^{15}N ppm, possible typtophan residues at lower left side of the spectrum 9.5 - 10.5 ^{1}H ppm and 125 - 130 ^{15}N ppm. There are also NH_2 aspargine and glutamine side chains at 6 - 6.5 ^{1}H ppm (as highlighted by the lines between the doublet peaks).

The dispersion of the HSQC indicates the presence of secondary structure. The central peaks in the area 118 - 121 ^{15}N ppm and 7.5 - 8.5 ^{1}H ppm indicate the presence of α-helical content and peaks between 6.8 - 7.2 ^{1}H ppm and 9.2 - 10.2 ^{1}H ppm show potential β-sheet content. This is consistent with the predictions that YraP contains these motifs.

From this spectrum, 225 peaks were identified. The estimated number of peaks in the HSQC of YraP can be calculated by taking the total number of residues and subtract the number of proline residues present, to give a value of 173 peaks. A possible explanation why there are more peaks picked than estimated could be due to the presence of loop regions that are able to exist in multiple conformers and produce another set of resonances, the picking of peaks that are weak and are most likely noise and the presence NH_2 side chains.

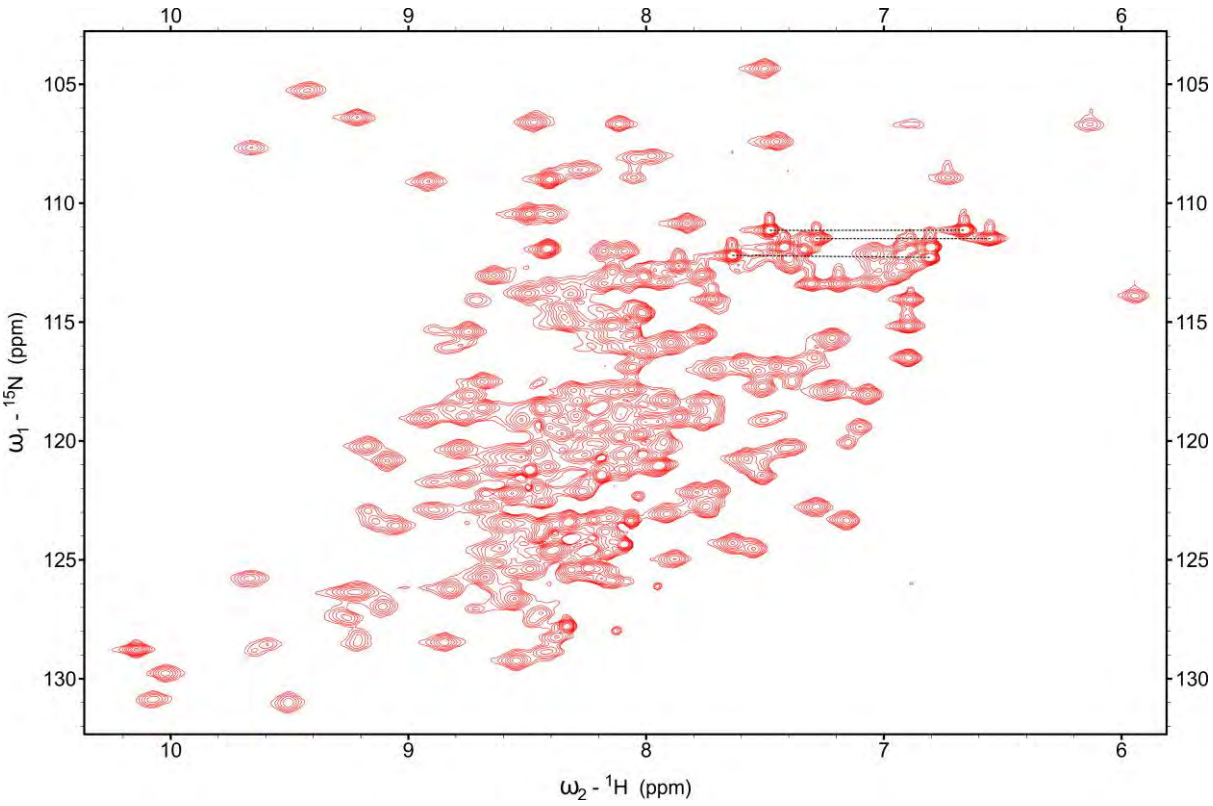


Figure 5.13 ^{1}H-^{15}N HSQC of YraP. HSQC performed using a Varian Inova 800-MHz spectrometer, with 400 μM ^{15}N-labelled YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25°C. Dashed lines indicate the NH_2 side chains of aspargine and glutamine doublets.

This spectrum shows that YraP is dispersed and has characteristic α-helical and β-sheet resonances. Due to the little overlap, this is amenable for backbone assignment.

5.3.3 Spectra and assignment of the backbone experiments

In order to assign the protein, experiments displaying the resonances of the carbon and nitrogen atoms of the backbone of the protein need to be performed. In this study, six backbone experiments were performed. HNCO, HNCACO experiments to determine the resonances of i and i-1 carbonyl (CO) groups, HNCA, HNCOCA for i and i-1 CA atoms, HNCACB and CBCACONH for i and i-1 CA and CB groups.

Each of these experiments produced ideal spectra for sequential assignment purposes. An example of the assignment process for i and i-1 CO, CA and CB are shown in **Figures 5.14, 5.15 and 5.16**. The assignment of all CA, CB and CO groups in YraP was achievable with these spectra. Confidence of these assignments are from the fact that characteristic chemical shifts were observed for certain residues, such as the CB resonances of alanine residues resonating at approximately 20 ppm on the ^{13}C axis, as well as the unique resonances produced from other residues (e.g. serine, threonine, glycine). To further confirm that the assignments were correct, the assignment prediction match on Sparky was used. In brief, this prediction analyses the sequential assignments and tries to find the most likely place in the sequence where they fit. Performing this prediction for all the assignments fitted this to the sequence of YraP with a score 0.781. Scores less than 1 generally indicate a correct assignment position.

Although all of the CA, CB and CO groups have been assigned, there are some notable exceptions. A peak corresponding to N153 was unable to be found in ^{1}H-^{15}N HSQC. However the CA, CB and CO groups of N153 were found from the other backbone spectra. Possible reasons for why N153 is not observed in the HSQC could be due to signal overlap or that N153 is in a dynamic region of the sampling numerous conformational states.

There are three proline residues present in YraP, P38, P86 and P116. As prolines have an imino group rather than the more common amino group, they do not display a resonance on the HSQC spectrum, as they lack the necessary

hydrogen bound to a nitrogen atom. This can make sequential assignment of the backbone problematic as it is not possible to assign i-1 resonances from a proline residue. However the backbone assignments from the other experiments have circumvented these possible problems and the prolines with the preceding residues have been correctly assigned.

There are also some instances on the HSQC spectrum where two residues will resonate at a similar chemical shift. These residues are A35 + L50, D37 + V149, V47 + K78, R69 + T150, Q115+ T126 and N124 + E155. However, from unambiguous backbone assignment, these residues were correctly assigned and happen to display signal overlap on the HSQC.

Despite all of these potential caveats, almost every CA, CB, CO and NH group has a correct assignment, except for the NH group of residue N153 where a peak was not found on the HSQC spectrum (**Figure 5.17**). From this, 176 of the total 176 residues in YraP have been assigned (Appendix II).

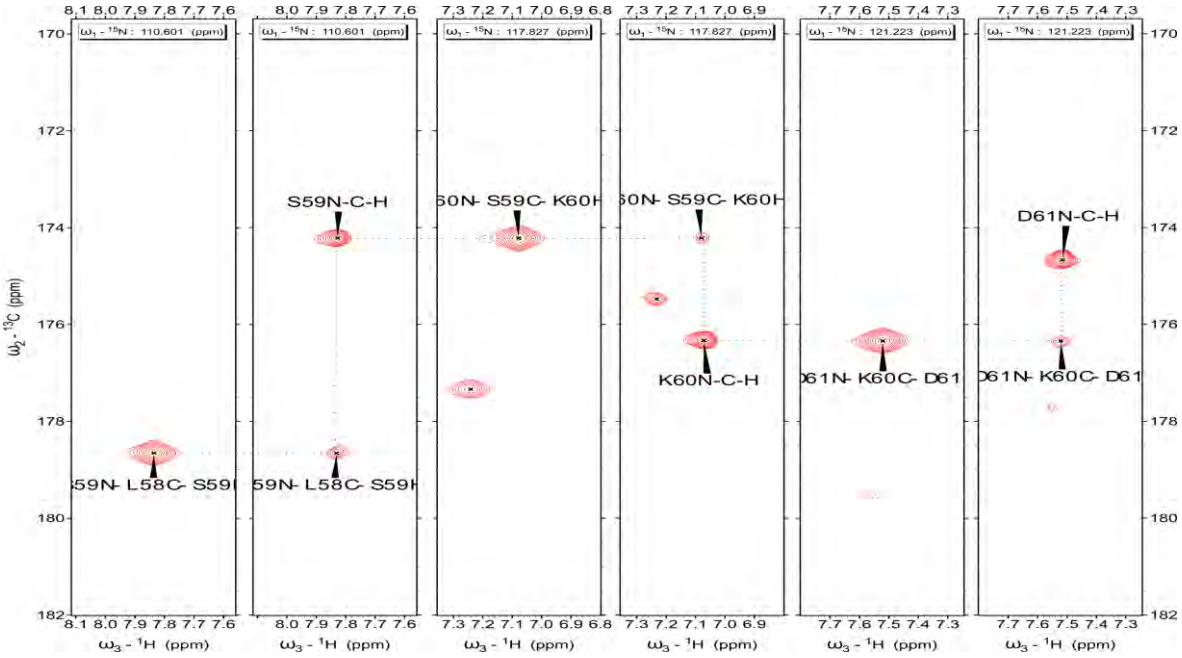


Figure 5.14 Sequential backbone assignments for the CO groups of residues L58-D61. Dashed lines indicate the sequential links determined. HNCO and HN(CA)CO experiments were performed with 1 mM ^{13}C-^{15}N YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25^{o}C on a 600 MHz Inova Varian spectrometer with cryoprobe.

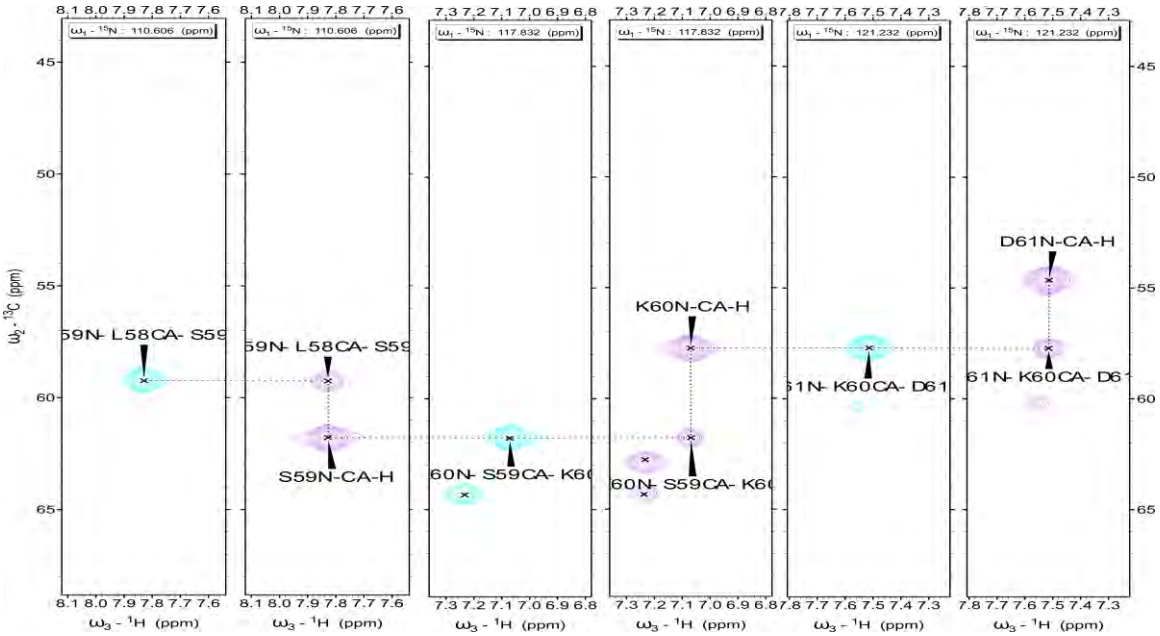


Figure 5.15 Sequential backbone assignments for the CA groups of residues L58-D61. Dashed lines indicate the sequential links determined. HNCA and HN(CO)CA experiments were performed with 1 mM ^{13}C-^{15}N YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25^{o}C on a 600 MHz Inova Varian spectrometer with cryoprobe.

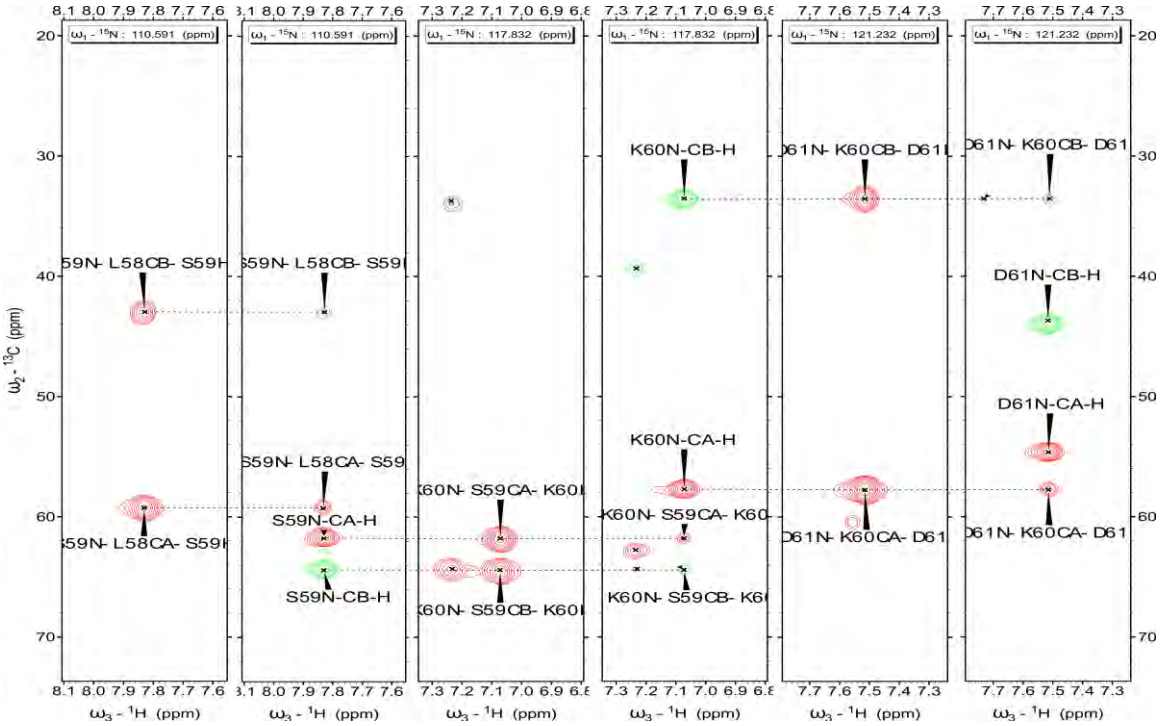


Figure 5.16 Sequential backbone assignments for the CA and CB groups of residues L58-D61. Dashed lines indicate the sequential links determined. HNCACB and CBCA(CO)NH experiments were performed with 1 mM ^{13}C-^{15}N YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25^{o}C on a 600 MHz Inova Varian spectrometer with cryoprobe.

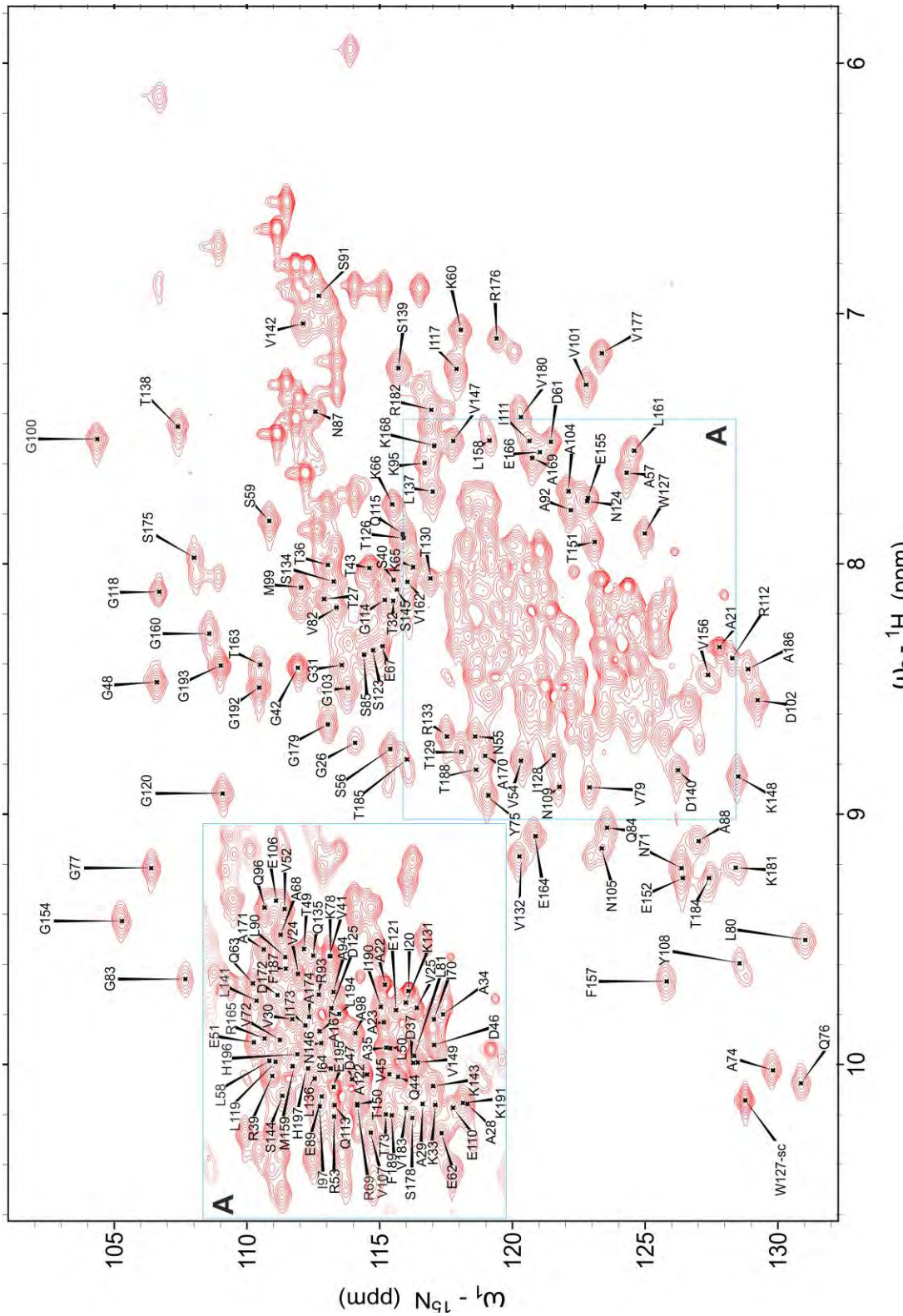


Figure 5.17 The backbone of YraP is almost completely assigned. ^{1}H-^{15}N HSQC of ^{13}C-^{15}N labelled YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25°C using a 600 MHz Inova Varian spectrometer with cryoprobe. Almost every residue is assigned on the HSQC apart from N153. There is a major overlapped area in the centre of the spectrum, which has been zoomed in (A).

5.3.4 Side chain assignments

Although assignment of the backbone can give potential secondary structure information of a protein, side chain assignments are crucial for determining its tertiary structure. To get these assignments, (H)CCONH and (H)C(CO)NH NMR experiments were performed to determine the side chain protons and carbons. Using information from these assignments, a HCCH TOCSY was also performed as a further proof to confirm the assignments.

The HC(CO)NH allows assignment of carbon attached protons of the i-1 residue from the transfer of magnetisation from i NH group. One notable aspect about this spectrum is that the i-1 HA group (proton attached to CA) has a characteristic chemical shift at approximately 4.1 ppm in the Proton axis, which was observed for most of the residues. Most of the side chain protons were assigned except for two particular regions, V107 - I111 and T151 - E155. These regions are thought to be flexible loops in YraP, which would explain the lack of assignments for those residues.

The (H)C(CO)NH displays the resonances of the side chain carbon atoms. Due to the limitation of this experiment, certain carbon atoms do not resonate, such as the CD (C-delta) of glutamine or CE (C-epsilon) of aspargines, as well as certain carbon atoms in aromatic rings. For the most part, many of the carbon atoms in residues resonated at their characteristic chemical shifts, with very few missed assignments.

The HCCH TOCSY was used to refine the side chain carbon and proton assignments. The resolution of this experiment is far higher than the previous experiments, but is far more complicated to interpret. The magnetisation pathway is different to the HCCONH and (H)C(CO)NH in that it does not transfer from a backbone NH group, but rather from the carbons and attached protons in the residue. Thus all resonances displayed are from proton and carbon frequencies only, making it an ideal tool for refinement of side chain assignments.

With the use of these three experiments, approximately 80% of the possible side chains of YraP have been assigned (Appendix II).

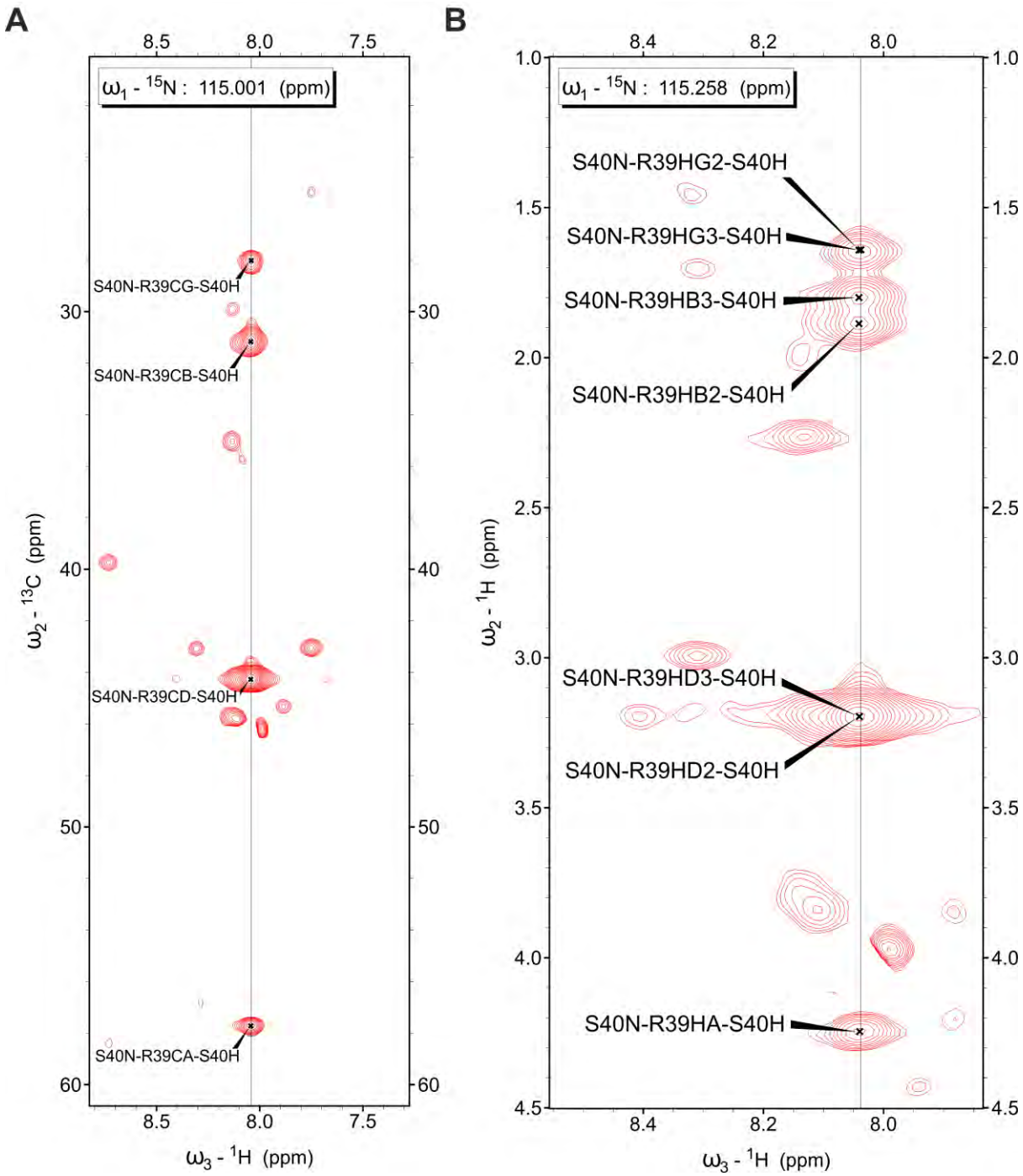


Figure 5.18 Side chain assignments of YraP. R39 CA resonance from CC(CO)NH experiment (**A**) and R39 side chain proton resonances from HCCONH experiment. Both side chain carbons and protons can be assigned using these spectra. These experiments were performed with 1.2 mM ^{13}C-^{15}N labelled YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 using a 800 MHz Inova Varian spectrometer with cryoprobe. Assignments were checked with the HCCH TOCSY.

5.3.5 Hydrogen - Deuterium exchanged HSQC

To aid in structure determination a Hydrogen/Deuterium (H/D) exchange HSQC was performed. In brief, ^{15}N-lablled YraP was lyophilised and resuspended in deuterated buffer (50 mM NaP, 50 mM NaCl, pH 6.0). This would allow surface exposed NH groups to exchange with the deuterons (ND) in the buffer. Any NH groups that are either buried in the core of the protein and not accessible to the surface, or those that are involved in H-bonding in the protein, will not be exchanged. As this is a ^{15}N HSQC, any NH that has been exchanged to a ND group will not give a resonance signal. By overlaying the H/D exchanged spectrum with the HSQC of YraP, a selection of residues were identified as possible residues involved in H-bonding or buried in the core (**Appendix IV**).

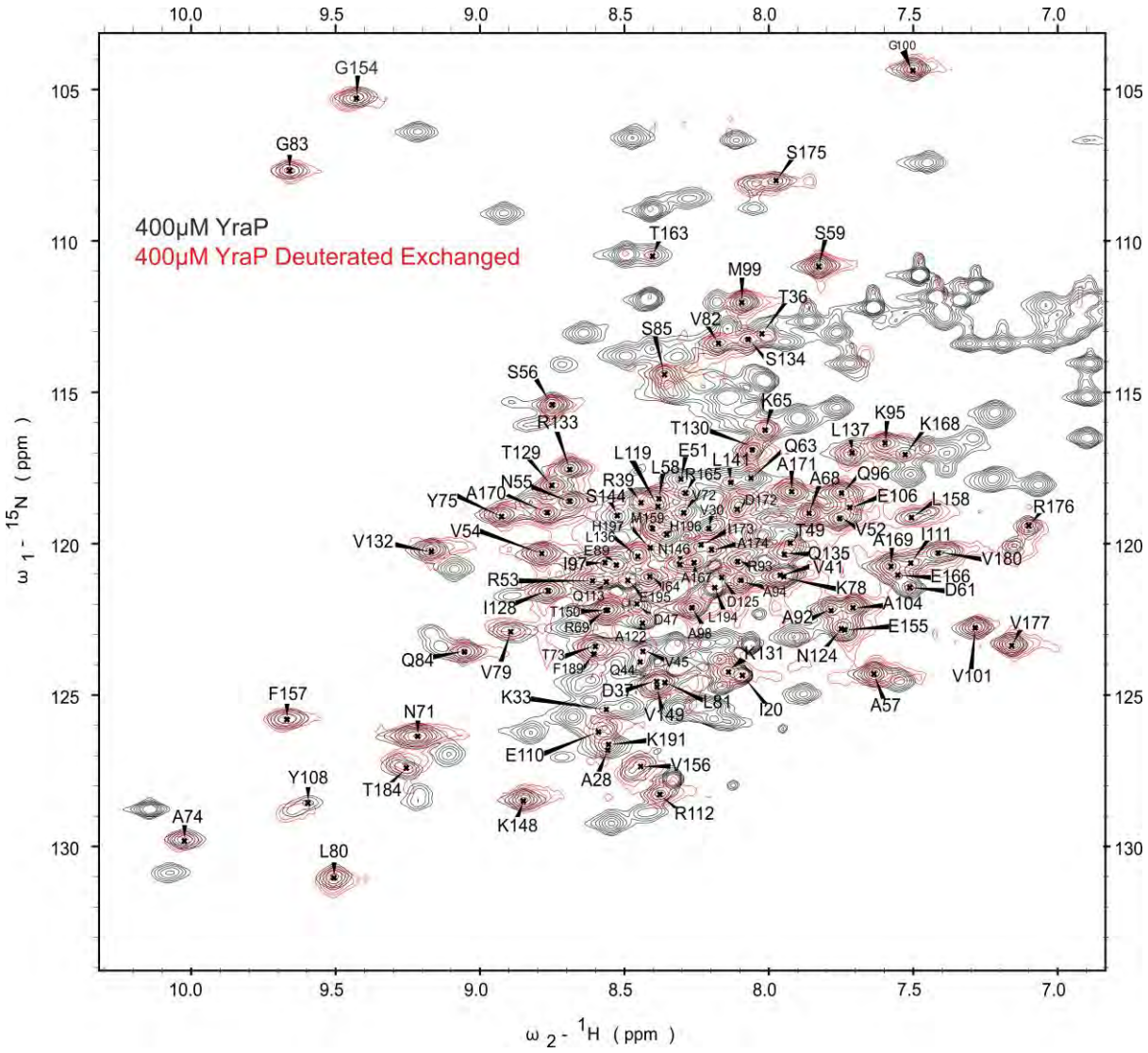


Figure 5.19 H/D exchanged HSQC of YraP. Overlaid spectra of 400 μM ^{15}N labelled YraP (black) in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 and 400 μM ^{15}N labelled YraP that was exchanged in deuterium for 2 hours, performed on 600 MHz Inova Varian spectrometer with cryoprobe at 25^{o}C. Several assignments disappear from the exchanged spectrum. Assignments that have remained are labelled on the spectrum, which are residues most likely to be H-bonded or not surface exposed.

5.3.6 Nuclear Overhauser Effect Experiments

The overall fold determines the 3D structure of a protein. In order to solve the structure of YraP using NMR, NOESY experiments were performed to gain information on the tertiary structure. NOESY experiments display possible through space interactions between protons, which allow the generation of distance restraints for structure calculations. We performed three major types of NOESY experiments, ^{15}N edited NOESY, ^{13}C edited NOESY and an Aromatic NOESY.

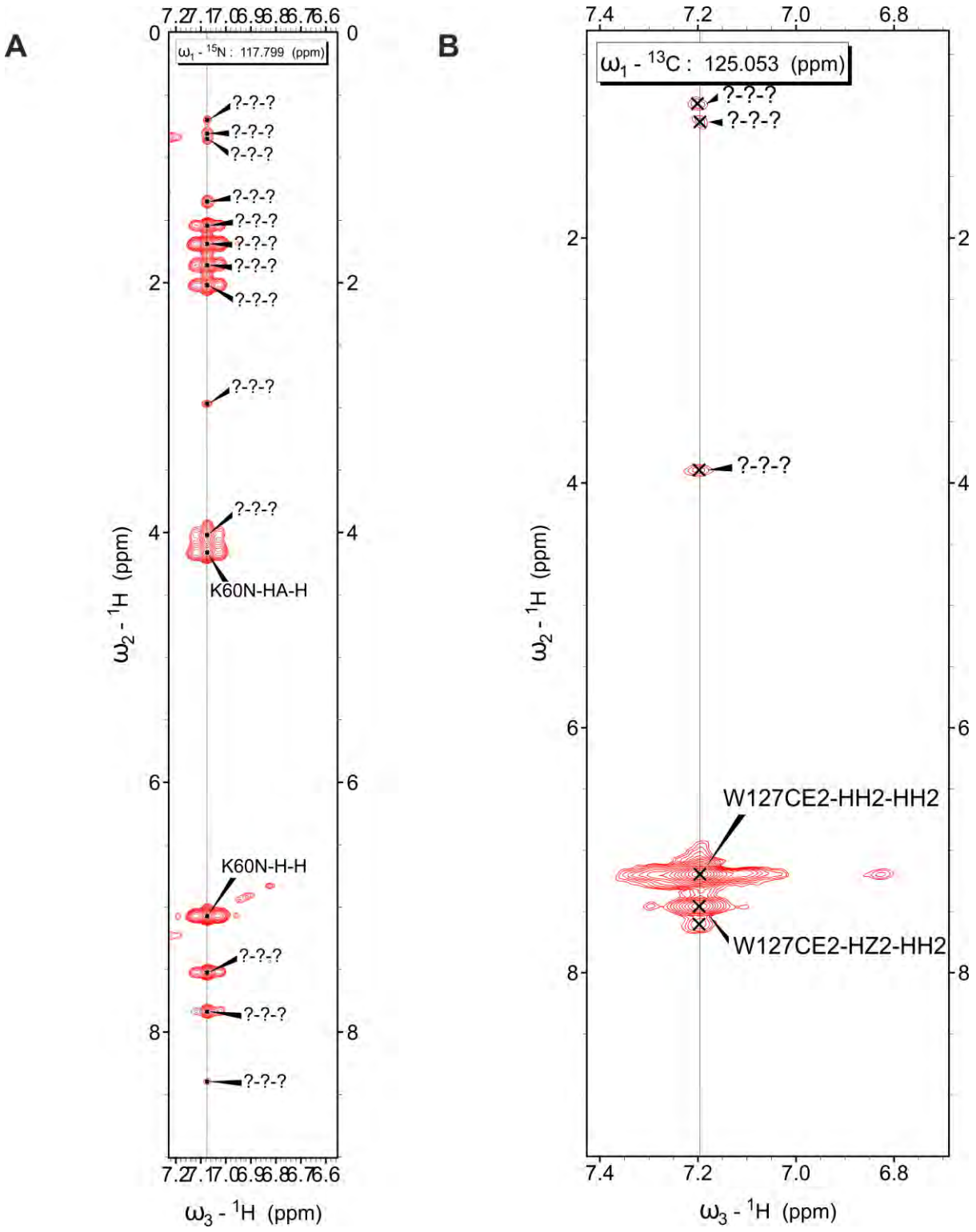
The ^{15}N NOESY displays protons within 5 Å of an amide residue. The spectrum produced from this experiment gave approximately 4500 peaks and was fairly clean, with little baseline noise. We were able to assign most of the diagonal peaks as well as identify possible residues in close proximity of the cross-peaks. To aid the structure calculations, a deuterated ^{15}N NOESY was performed, which yields a much simpler spectrum. The only resonances present are from the NH groups, which have been exchanged from deuterons back to protons. This can be a valuable tool in mapping secondary structure elements. As both experiments involve magnetisation transfer to the NH group, the same regions of overlap that were observed in the HSQC, were observed in both ^{15}N NOESY spectra. However, due to the quality of the spectra, we were able to make unambiguous assignments.

The ^{13}C NOESY-HSQC displays all protons that are within 5 Å of all carbon-linked protons present in the protein. These are the most common groups in

proteins and hence this spectrum is the most complicated but also contains the most information. Approximately 5500 peaks were picked for this spectrum. In comparison to the ^{15}N-NOESY, there were only approximately 1000 additional peaks, this relative small increase is due to the large increase in complexity and signal overlap, making unambiguous peak assignments difficult. As an example, the majority of the glycine assignments were found to be on top of each other, due to these residues possessing similar chemical shifts to each other.

The aromatic NOESY allows the identification of close proximity protons from carbon-linked protons on aromatic rings. These are mainly from phenylalanine, tyrosine and tryptophan residues. In YraP there are only 5 aromatic residues. Because of this there were only approximately 80 peaks produced in the aromatic NOESY spectrum. From this data it allowed the assignment of 4 out of the 5 residues, Y75, Y108, W127 and F189. We were unable to assign F187. It should be noted that Y75 seemed to possibly exist in another state, as in the ^{13}C dimension where it was assigned, there was another diagonal peak approximately 0.15 ^{1}H ppm upfield that displayed very similar NOE interactions to that of the assigned Y75 peak.

The list of peaks used can be found in Appendix III.


A
ω1 - 15N : 117.799 (ppm)
?-?-?
K60N-HA-H
K60N-H-H
ω2 - 1H (ppm)
ω3 - 1H (ppm)
B
ω1 - 13C : 125.053 (ppm)
W127CE2-HH2-HH2
W127CE2-HZ2-HH2

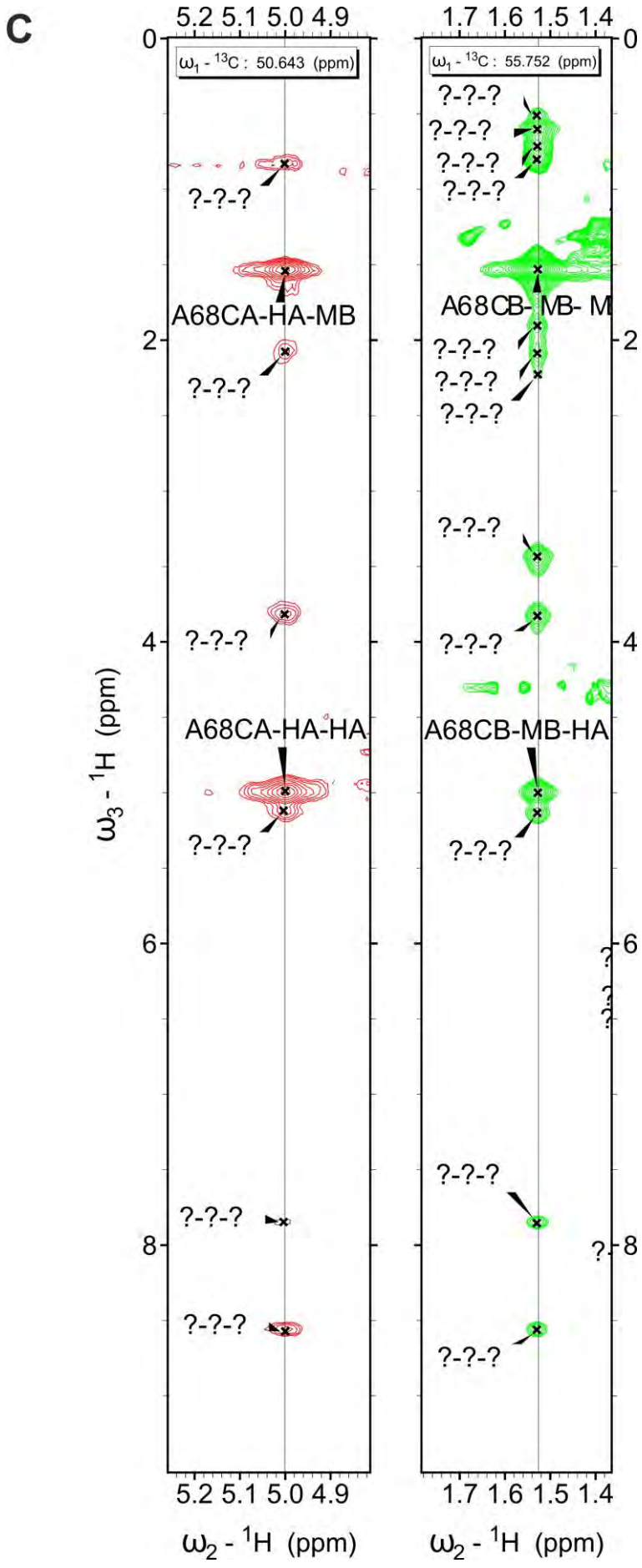

C
5.2 5.1 5.0 4.9
1.7 1.6 1.5 1.4
ω1 - 13C : 50.643 (ppm)
ω1 - 13C : 55.752 (ppm)
?-?-?
A68CA-HA-MB
A68CB-MB-M
A68CA-HA-HA
A68CB-MB-HA
0
2
4
6
8
ω3 - 1H (ppm)
ω2 - 1H (ppm)

Figure 5.20 NOESY assignments of YraP. Experiments were performed with 1.2 mM ^{13}C-^{15}N YraP in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25^{o}C using a 800 MHz Inova Varian spectrometer with cryoprobe. **A** ^{15}N edited NOESY of K60. The NH and HA protons have been assigned. The other unassigned peaks correspond to potential through space interactions with other protons. **B** Aromatic NOESY of W127. This spectrum shows potential NOE interactions from the aromatic rings that would be unobtainable with other experiments. **C** ^{13}C edited NOESY of A68. There are 2 assignable resonances for alanines, the HA and the methyl side group protons (MB). In this spectrum, observing at the CA frequency, both HA and MB protons are assigned. Then by observing the same HA and MB resonances at the CB frequency confirms that these assignments are correct.

5.3.7 Secondary structure predictions using backbone assignment data

With almost complete assignment of the backbone, secondary structure predictions were made using TALOS (Cornilescu *et al*., 1999), CSI (Wishart and Sykes, 1994) and ΔCα - ΔCβ method (Metzler *et al*., 1993). These secondary structure predictions can be used to ensure that backbone assignments are correct and can provide torsion angle restraints for structure calculations.

TALOS is a program which compares experimentally derived backbone chemical shift data with those of a database compared of 200 proteins. It then compares the φ and ψ angles produced from the chemical shift data and assigns a likely secondary structure element based on comparisons with the database. By plotting each of the angles for each residue, regions where the φ and ψ values are distant correspond to α-helical content whilst regions with the φ and ψ values close together correspond to β-strands.

CSI (chemical shift index) is another prediction program which compares chemical shift data to its database. Comparisons are based on the chemical shift differences against some predefined 'random coil' values, using chemical shifts derived from HA, CA, CB and CO assignments. The output is given as scores of 1, 0 and -1, which correspond to β-strand, flexible region and α-helix respectively.

The ΔCα - ΔCβ method is another prediction that compares the Cα and Cβ chemical shifts of residues to a standard set of values that would be observed in unfolded proteins. Comparisons between folded and random coil chemical shifts

show that Cα resonances tend to shift downfield in α helices and upfield in β-strands. The opposite is observed with Cβ resonances, with upfield shifts in α helices and downfield shifts in β-strands. A subtraction between the Cα and Cβ differences enhances the correlation between these chemical shifts and secondary structure elements. Positive values correspond to α helices whereas negative values correspond to β-strands.

Common secondary structural trends are observed for YraP with all three secondary structure prediction methods (**Figure 5.21**). The predictions estimate 4 α helices between residues D46-I64, A88-D102, G120-L136 and E164-V180. TALOS and ΔCα - ΔCβ predictions suggest that there is a break in the α-helix of D46-I64 at residues 60-61. 6 β-strands have been predicted between residues E67-A74, V79-G83, N105-N109, S145-T151, E156-L161 and V183-T188. Two repeats of αββαβ secondary structural elements are seen, which likely corresponds to the 2 predicted BON domains present in the protein. Due to the similar predictions between all three methods, it seems highly likely that the secondary structure elements are present at those residues.

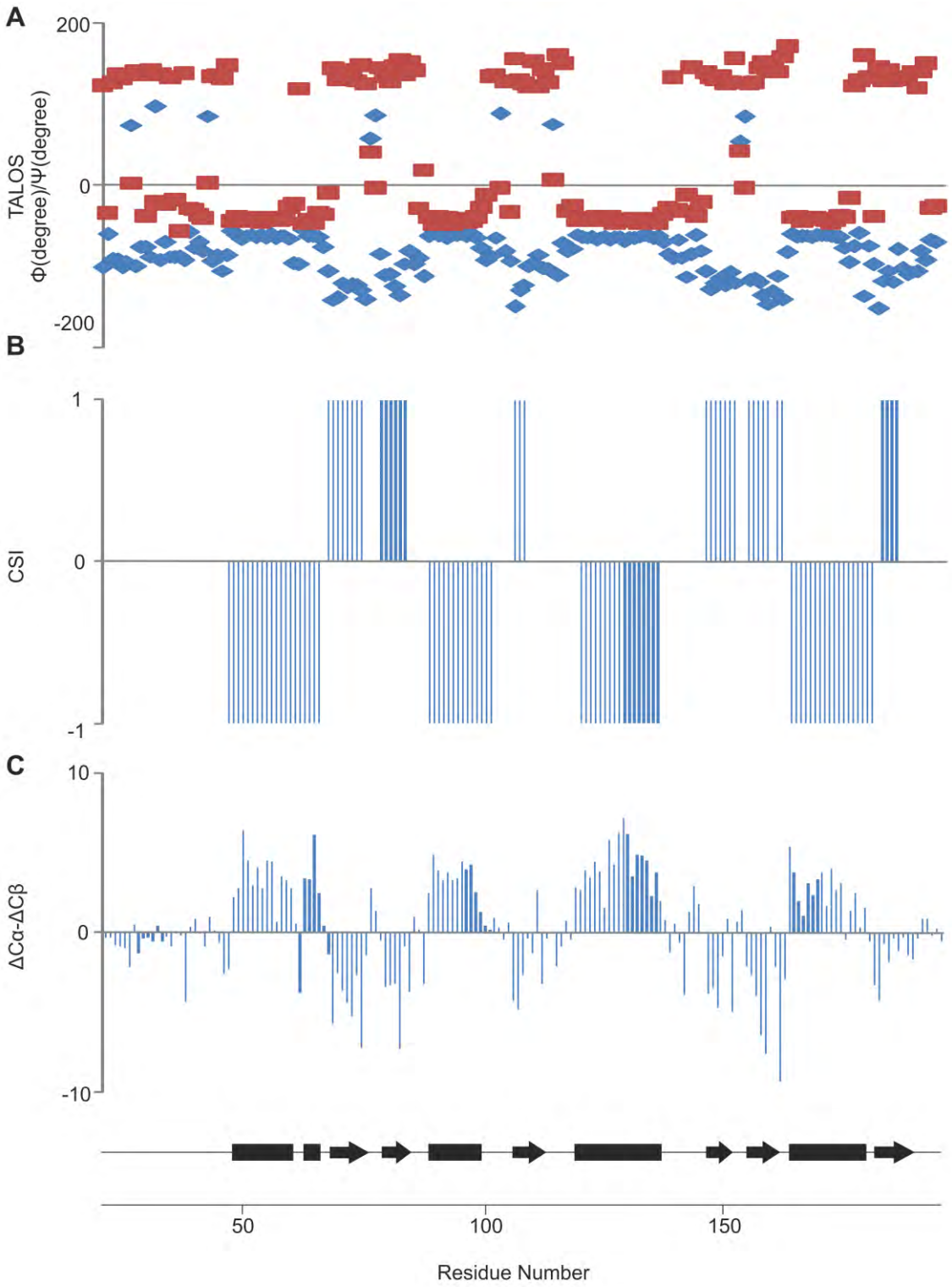


Figure 5.21 Secondary structure predictions of YraP using chemical shift data. Graphs of the data generated from **A** TALOS, **B** CSI and **C** CΔα - CΔβ predictions are shown. With TALOS, φ angles are plotted in red and ψ in blue. Predicted secondary structure elements are shown for α-helices (rectangles) and β-strands (arrows) as well as their approximate residue numbers in the protein.

5.3.8 Determining Hydrogen Bonds

5.3.8.1 Hydrogen - Deuterium exchanged HSQC

In order to generate an accurate structure, the H-bonding network needs to be determined. H-bonds are essential in keeping the secondary structure of the protein. In structure calculations, H-bonds can serve as strongly weighted restraints which can aid the simulated annealing and ultimately yield to a more refined and accurate model of the protein.

The H/D exchange HSQC experiments highlight which amides are protected from exchange and hence involved in hydrogen bonding. However this information provides only one partner of the hydrogen bond. No information is given regarding its partnering carbonyl. In order to elucidate this, initial structures were generated using NOE data solely. These early structures, though not well resolved showed key secondary structural elements, for example the α helices of YraP were resolved in the first structure calculations and were consistently present in subsequent calculations.

The H-bonds in α helices have an n+4 periodicity, with the CO group of a residue H-bonding with the NH group of a residue that is 4 residues further in the sequence. By observing this on the calculated structures, we could map out the H-bonding network for the α-helices. To confirm that these residues are able to form H-bonds, the peaks corresponding to these residues on the H/D exchanged ^{15}N HSQC were inspected. If the peaks were present (i.e. not exchanged to deuterons), then the H-bond was confirmed and used as a restraint in subsequent structure calculations. As more and more of the structure was

resolved more hydrogen bond restraints could be incorporated. The positions of the β-sheets were also determined and the H-bonds were mapped to these regions using the same method. The full list of H-bonds that were identified is present in Appendix V, which were used to generate the structure of YraP.

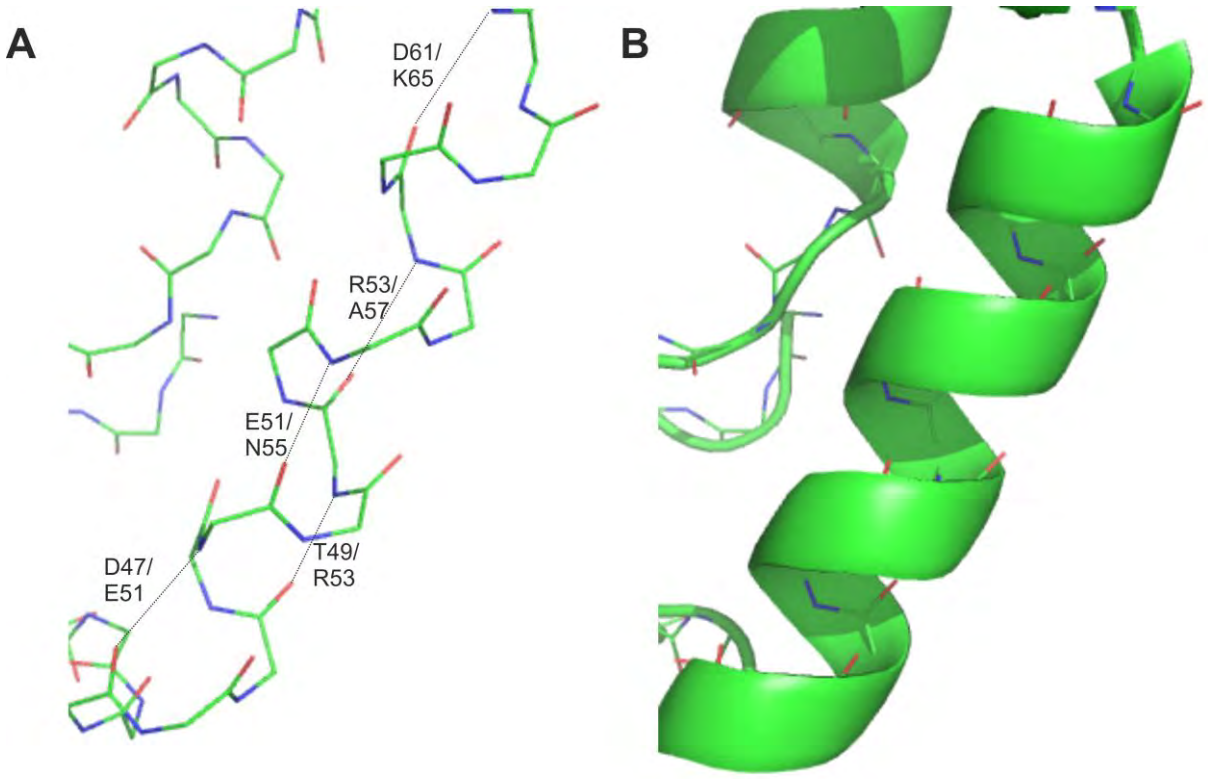


Figure 5.22 H-bonds identified in an α-helix of YraP. From intermediate structure calculations, H-bonds with an n + 4 periodicity could be observed. By mapping these, H-bond restraints could be used in further structure calculations. **A** An example of some of the H-bonds found in the first α-helix of YraP. Residues involved are labelled and H-bonds are indicated with the grey line which goes from the O-atom of the first residue (red) to the N atom of the second residue (blue). **B** The α-helical structure that is formed from using these H-bond restraints.

5.3.8.2 Identification of β-strands using deuterated ^{15}N-edited NOESY data

With the later structure calculations, β-strand regions identified. Finding interactions between β-strands, such as the H-bonds between them, is more difficult to resolve in structure calculations and was a problem for YraP. This can be because β-strands lack the regular n+4 periodicity present in α-helices for H-bonding and that they can exist as a mixture of both parallel and anti-parallel orientations. Because of this, in order to get the correct set of H-bonds between the β-sheets, the correct orientation needs to be defined in the simulated annealing calculations. If the wrong H-bond network is used to determine the β-sheets, this can result in either strong violations or the production of an erroneous structure.

To aid finding the correct H-bonds, a combination of intermediate structure calculations and the deuterated ^{15}N edited NOESY were used together with H/D exchange analysis. These structure calculations, although not completely resolved for the β-strands, allowed identification of the regions that contained them and would give an idea of which residues are in close proximity to each other and correctly orientated for H-bonding. A deuterated ^{15}N-edited NOESY was used as it provided a far simpler spectrum than a standard ^{15}N-NOESY as it only provided amide-amide NOESYs. These are the predominant NOE interactions observed between β-strands and hence provided a clear fingerprint of interactions between residues, thus enabling clear matching of β-strands together. Finally H/D exchange was used to identify those amides involved in H-bonding. Utilising all three techniques allowed identification of the corresponding

carbonyl binding partner and thus the incorporation of a hydrogen bond in subsequent structure calculations. This procedure was pivotal in resolving both the parallel and anti-parallel β-strands in YraP and refining the structure.

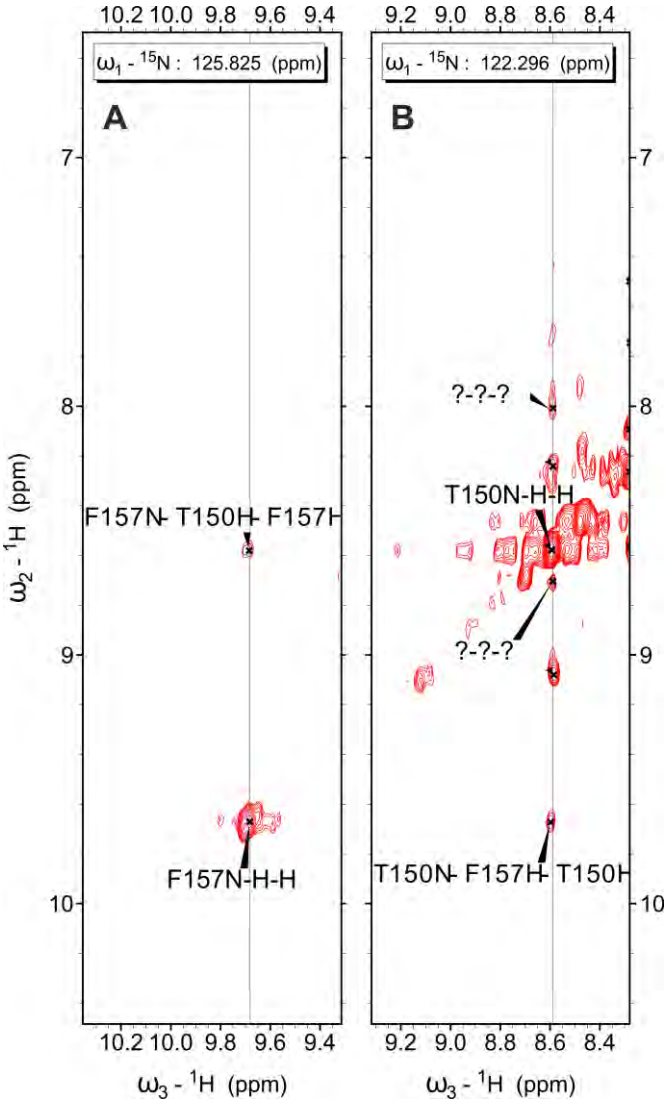


Figure 5.23 NOESY assignments in the deuterated ^{15}N-edited NOESY experiment. By identifying assignments between residues, H-bonds can be determined in β-strands. In this example, T150 and F157 are predicted to be in separate β-strands that are in close proximity to each other. (A) From F157NH, an assignment has been made to T150H at 9.6 ppm. (B) From T150NH, a similar assignment is made to F157 at 9.65 ppm. These assignments together with the list of amides involved in hydrogen bonding from the H/D exchanged HSQC spectrum, has led to the identification of a H-bond between these two residues.

5.3.9 Solving the structure of YraP using NMR data

The combination of the previous NMR experiments has yielded data that can act as various parameters for correct structure calculations. From the backbone assignment data, combined with the predictions from TALOS, can give an accurate estimate of the Φ and Ψ angles of the protein. This in turn allows predictions of possible secondary structure elements for certain regions of the protein. This can be aided further by the addition of H-bonds, which are strong restraints in refining the secondary structure. Obtaining assignments for the side chain protons as well as their respective NOESY interactions can yield 3D restraints. This is beneficial for structure calculations in that the overall fold is determined from this data.

The generation of the correct structure from NMR data is an iterative process. A structure is calculated from a given set of data. This structure is analysed for potential errors, usually given from the list of violations from CYANA. The data is adjusted accordingly and another calculation is performed, with its set of violations. This is repeated until a structure is produced that has very few violations and the overall fold is accurate.

Over the course of this study, numerous structure calculations were performed to generate the final structure of YraP. This section details some of the structures from these calculations, the problems associated with them and the actions that were performed to obtain the structure with the least number of violations.

5.3.9.1 Initial structure

Initial structure calculations were performed using assignments from backbone and sidechain data, together with TALOS data only. All picked NOESY peaks were used, with 4500 and 5500 peaks from ^{15}N and ^{13}C NOESY data respectively. This dataset included very weak peaks very close to the noise level, as well as the very strong peaks associated with overlapped assignments. No aromatic NOESY data was used in this calculation.

The structures produced from this calculation display a compacted ball shaped structure. No individual domains can be identified, which is not expected from the predictions of YraP. However 4 α-helices are resolved in this calculation at residues 46-66 (α1), 89-98 (α2), 119-139 (α3) and 166-175 (α4). This corresponds to the secondary structure predictions by TALOS, CSI and ΔCα-ΔCβ .

There are some notable features about these helices. With α1, there is break of the helix into a loop region at D61, which is a bent region that forms a small α-helix between E62 and K66. These two α-helices will now be referred to as α1 (between residues 46-66) and α1' (between residues 62-66).

α3 possesses produced a helix that has an almost 'U'-shaped kink in it. This helix looks inaccurate and is probably a result of either erroneous assignments or NOESY peaks that have caused this conformation of the helix.

This structure calculation did not resolve any β-sheet content. The secondary structure predictions and HSQC of YraP suggest that YraP contains 6 β-sheets. Generally, β-sheets are harder to resolve than α-helices and accurate NOESY assignments are needed to resolve them.

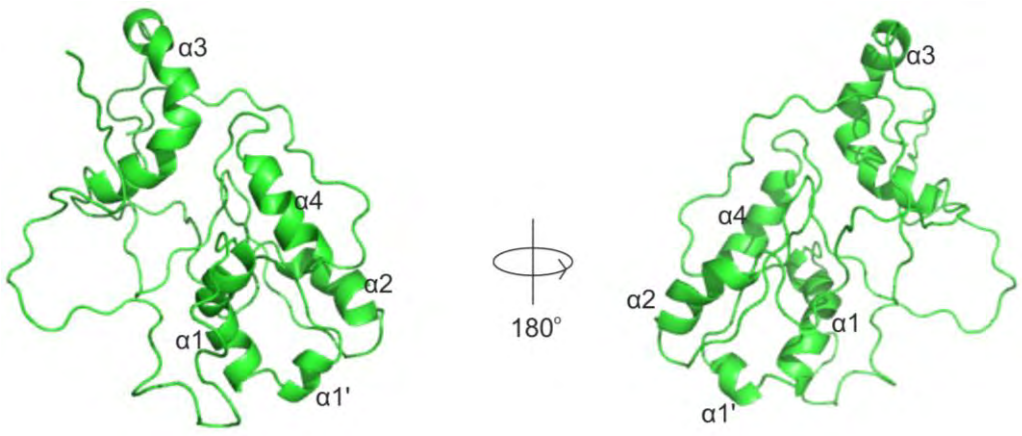


Figure 5.24 Initial structure calculations using NOESY, TALOS, backbone and sidechain assignments with CYANA. No aromatic NOESY data was used. This data has allowed the identification of 4 α-helices. α3 is quite bent, suggesting that there are erroneous NOEs in this data.

5.3.9.2 Intermediate structures - Resolving anti-parallel β-strands

A large number of peaks were identified in the NOESY data. There was a chance that some of these were erroneous and were introducing inaccuracies in the structure calculation runs. Likely candidates include overlapped peaks, where the peak intensity would be higher than were it an individual resolved peak. Other candidates include very weak peaks which could be noise. To solve these problems, the ^{15}N and ^{13}C NOESY data were filtered by intensity. The removal of both very strong and weak peaks happened during these filtering steps. Also, H-bonds for the known α-helices were introduced.

The structures produced from these runs start to show the formation of the β-strand content. Two pairs of anti-parallel β-strands were observed in the protein. The two domain architecture of YraP is beginning to show and each domain contains 2 α-helices and 2 anti-parallel β-strands. This structure contains extended loop regions between residues 100-115 and 177-189. These loops are not present in the predictions but are thought to produce β-strands. Indeed in the structure these regions can be seen to adopt extended structures due to the TALOS restraints used. However, due to NOE conflict, resolution of where these β-strands reside cannot be resolved.

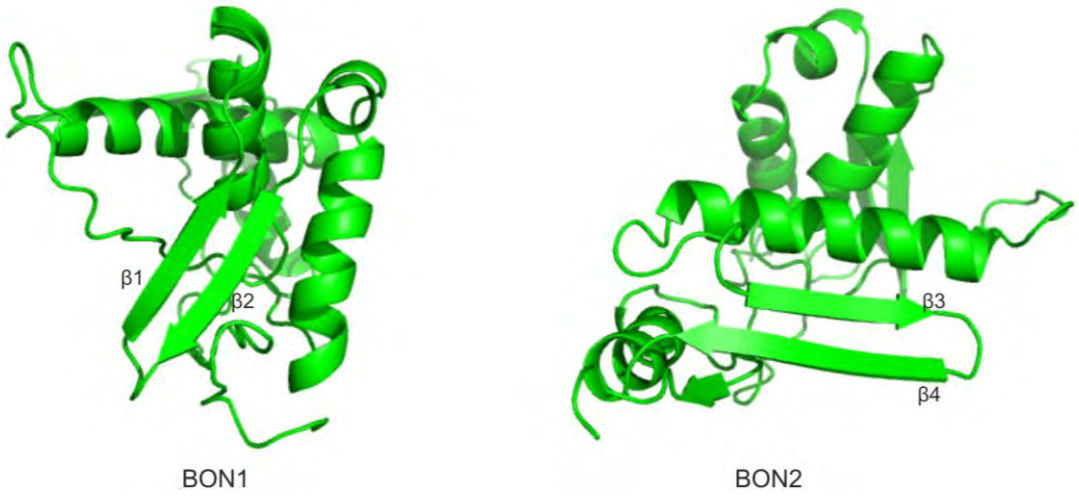


Figure 5.25 Representative structure of YraP using limited NOESY restraints.
These calculations have allowed the identification of anti-parallel β-strands in each of the BON domains. The orientations of each of the domains may be incorrect.

5.3.9.3 Intermediate structure - Resolving parallel β-strands

Following the pervious structure, several modifications were made. All of the NOESY spectra were analysed extensively for any errors. For example, by

analysing and correcting any assignments in remaining unresolved regions and both strong peaks that were near the diagonal and weak peaks that may just be noise were removed. The data used in this calculation included ^{15}N,^{13}C and ^{15}N deuterated amide-amide data, along with an updated H-bond list containing the H-bonds between the already resolved α-helices and β-sheets. Diagonal peak filtering was used as well to remove any overly strong assignments, also included was a longer cooling time to further allow resolution of any problematic NOEs.

The structure from this calculation finally shows the presence of the previously unresolved β-strands. The structure shows they are paired in a parallel fashion with the already resolved anti-parallel β-strands to form a three stranded β-sheet in each domain. All of the predicted secondary structure elements are resolved. The orientations of these elements have changed from what has been observed in previous calculations. Each domain has an almost parallel orientation of its secondary structure elements. However, the orientations between the domains are almost perpendicular. Also α3 seems to be in close proximity to BON1, due to the tight loop turn between residues 112-116. This seems to be an unfavoured conformation and is most likely from strong NOEs on either α3 or β1, forcing BON2 into this orientation.

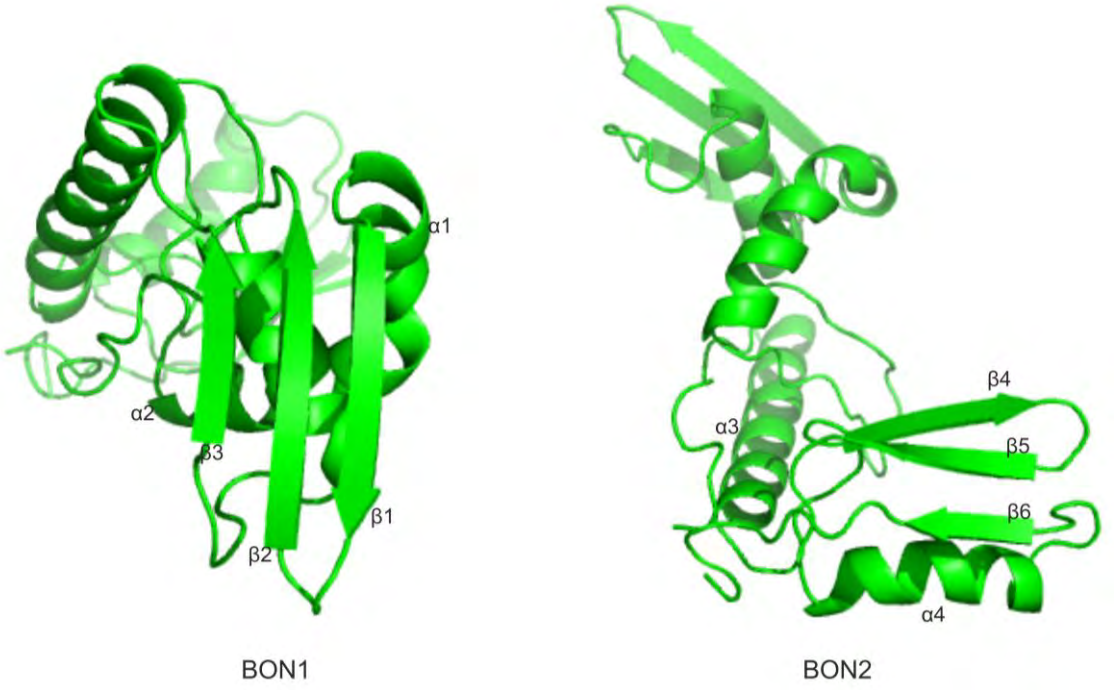


Figure 5.26 Representative structure calculation with all of the secondary structure elements resolved.

The addition of assigned NOESYs in the ^{15}N deuterated NOESY data and H-bonds in between the α-helices and β-strands have allowed a third parallel β-strand to be solved in each of the BON domains.

5.3.10 Final calculations

The final set of calculations mainly involved using the previous data that was further cleaned. As the data was reasonably accurate, shorter structure calculations could be performed, speeding up the structure determination process. These set of calculations were performed in an iterative process, where structures were calculated and a set of violations were produced. By addressing these violations (i.e. addressing errors, analysing NOE patterns and removing erroneous data), local changes in erroneous regions in the protein could be made, which would alter the final fold at the end of the simulated annealing run. Once most of the violations had been corrected, the final structure of YraP determined. This work was performed by Dr. T. J. Knowles.

5.3.11 The final structure of YraP

Using NMR analysis, a high resolution structure of YraP in solution has been determined with an RMSD of 1.83 Å. This structure is the lowest free-energy conformer produced from the structure calculations with this NOESY data and the top 20 structures produced from these calculations align very well in the core structured elements (**Figure 5.27, Table 5.1**).

The solved structure contains an unstructured N-terminus (**Figure 5.28**) and two BON domains (BON1 and BON2) that are separated by a small flexible region (residues 112-118). Both of the BON domains contain a mixture of α-helical and β-sheet and are organised in a similar manner with a repeated αβαββ pattern. All of the predicted secondary structure elements have been identified and are labelled as α1 (47-59), α1' (62-66), β1 (70-74), β2 (79-83) α2 (88-99) and β3 (106-111) which comprise BON1 and α3 (119-138), β4 (148-152), β5 (155-159) α4 (164-176) and β6 (182-185) which make up BON2, with small flexible regions connecting each of these secondary structure elements (**Figure 5.29**).

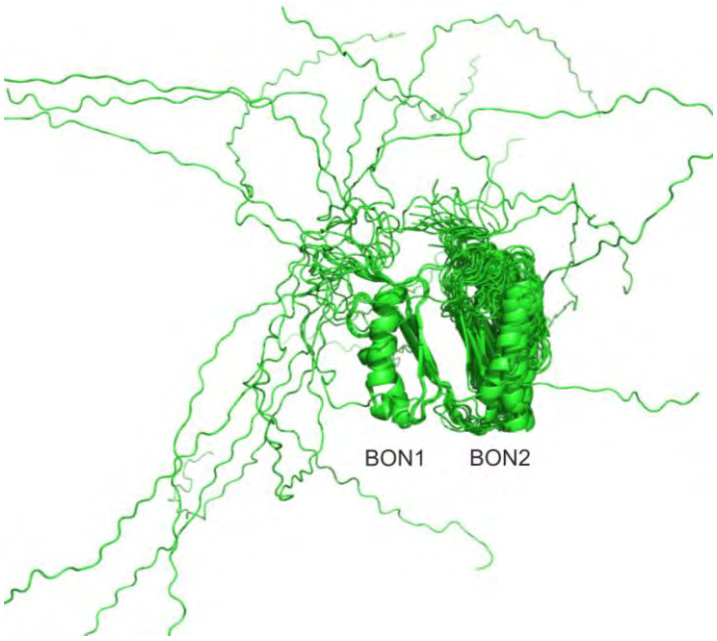


Figure 5.27 The top 20 structures of YraP calculated using CYANA. Strong alignment is seen in the key structural elements that contain BON1 and BON2 whereas the flexible N-terminus is not aligned in all 20 structures.

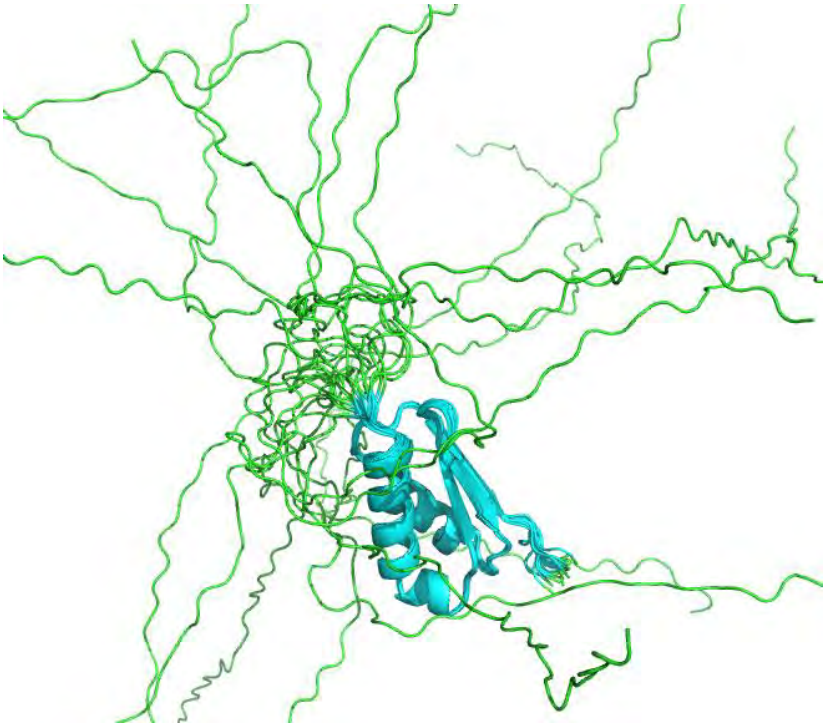


Figure 5.28 The N-terminal region of YraP is flexible. Alignment of the ensemble of the 20 structures of YraP. The N-terminal region (residues 20-46) is coloured green and BON1 is coloured cyan (residues 47-111). The structures converge at BON1, where alignment is observed. With the N-terminus, no alignment is observed in the 20 structures suggesting that it is flexible.

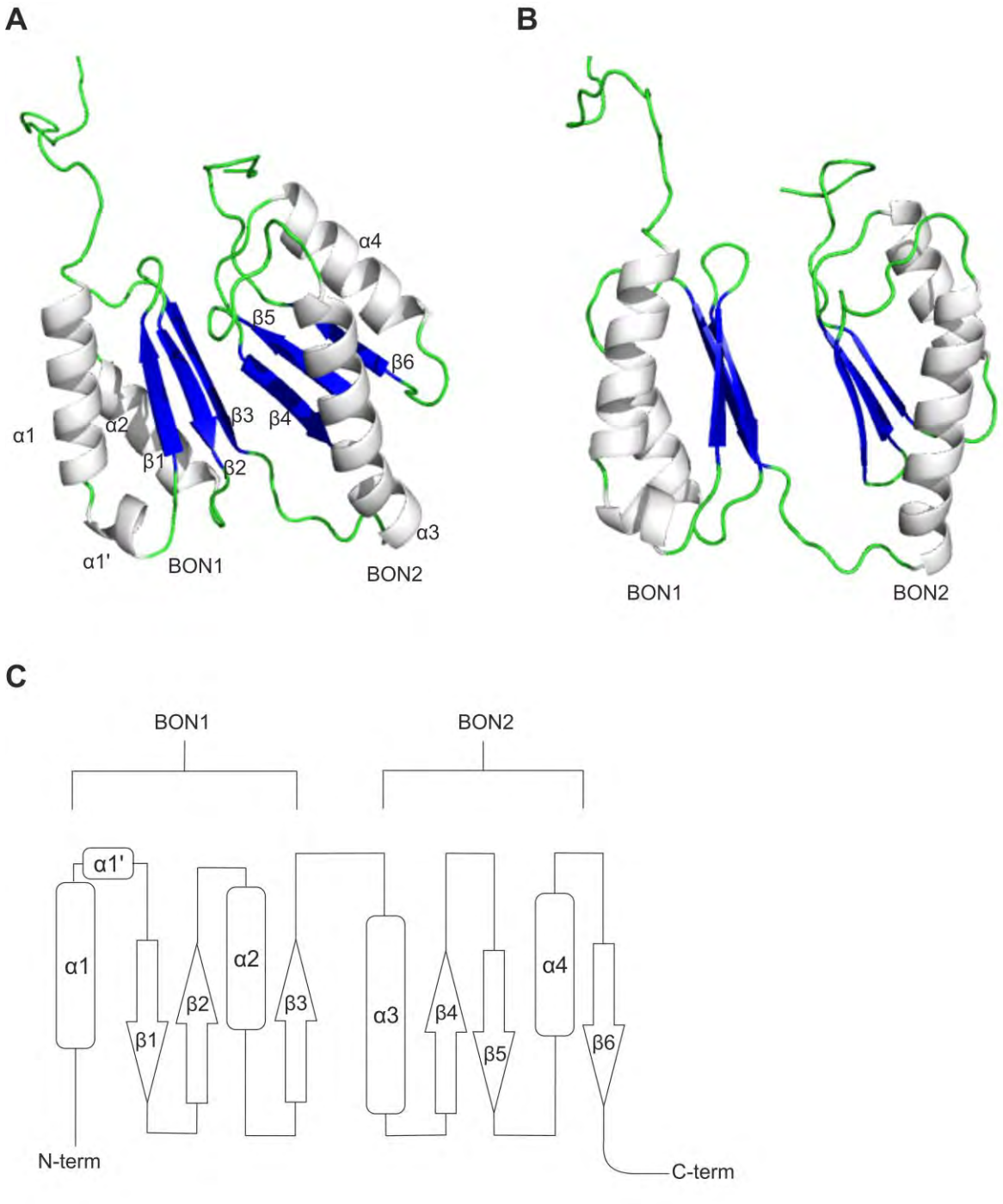


Figure 5.29 The solution structure of YraP. This represents the lowest free-energy structure of YraP. **A** Cartoon representation of YraP showing all of the secondary structure elements and the overall αββαα fold of each BON domain. α-helices are coloured white and β-strands blue. **B** Side view of YraP, showing that each of the BON domains consists of 2 α-helices and 3 β-sheets and is arranged in an almost box like manner. **C** Schematic representation of YraP, showing the order of the secondary structure elements.

Table 5.1 Statistics of the final structure of YraP

Number of restraints	
NOE	
Total	1805
Intraresidue (i=j)	346
Sequential (|i-j|)=1	509
Medium range (1 < |i-j|≤4)	496
Long range (|i-j| > 4)	454
Hydrogen bond restraints	120
Dihedral-angle restraints	244
RMSD from distance restraints (Å)	0.0276 ± 0.002
RMSD from covalent geometry	
Bond lengths (Å)	0.004 ± 0.000
Angles (degrees)	0.468 ± 0.015
Impropers (degrees)	1.368 ± 0.080
Ramachandran plot (%)	
Most favoured regions	89.0
Additionally allowed regions	9.3
Generously allowed regions	1.3
Disallowed regions	0.4
RMSD (Å)	
Global (46-190)	
Backbone	1.21 ± 0.44
Heavy atoms	1.83 ± 0.4
Domain 1 (46-112)	
Backbone	0.60 ± 0.12
Heavy atoms	1.36 ± 0.18
Domain 2 (118-190)	
Backbone	0.78 ± 0.13
Heavy atoms	1.54 ± 0.19

There is one notable difference between the two BON domains in YraP. The α1 helix of BON1 contains a small loop region followed by a 90^{o} turn back into a helix. This is not observed in BON2 where α3 contains no intervening loop region, but does contain a small kink. This particular helix is designated α1' and is a normal α-helix which retains its n+4 periodicity and not a 3_{10} helix.

In each BON domain, the pair of α-helices are orientated in the same direction, almost parallel to each other. Also the 3 β-sheets are paired, with β1 and β2 as an anti-parallel pair and β2 and β3 in a parallel pair. The α2, β1, β2, and β3 secondary structures are almost the same length in each of the domains. All of the secondary structure elements are orientated in a parallel manner, with secondary structure facing the same direction, producing a flat, box-like structure for each domain. The BON domains have aligned very well (**Figure 5.30**). With the overlay of the top 20 structures calculated, BON1 displays very little flexibility. BON2, although an aligned core region of YraP, is less well aligned than BON1. These BON domains in YraP also exhibit a similar secondary structure to the known BON domain structure of Rv0899 of *Mycobacterium tuberculosis* (Teriete *et al*., 2011) (**Figure 5.31**).

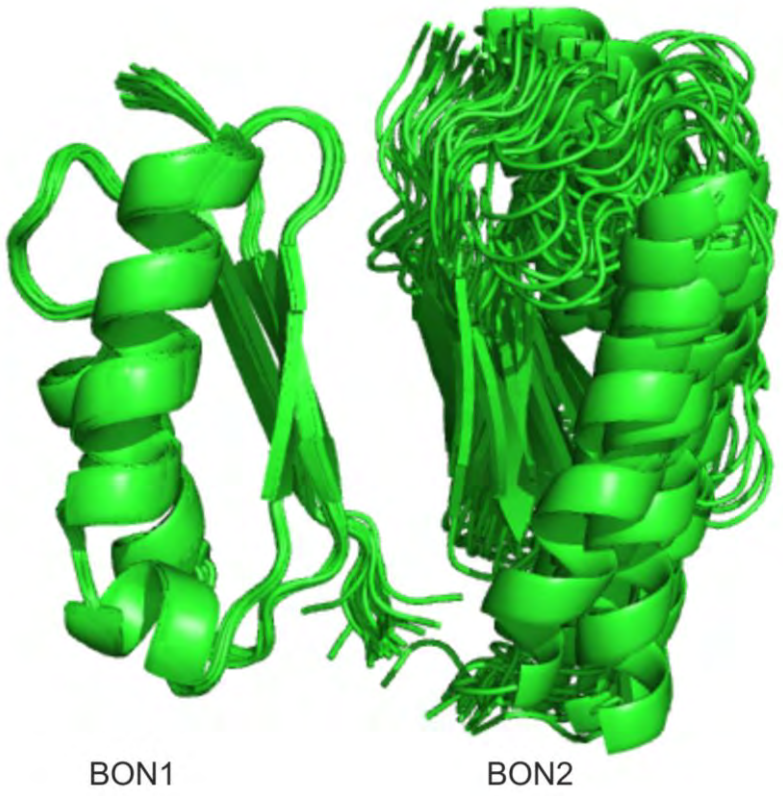


Figure 5.30 Top 20 structure calculation alignments of the individual BON domains. Although both BON domains have been resolved, the alignment of BON1 is much greater than the alignment of BON2, which is well less aligned, especially in its flexible loop regions.

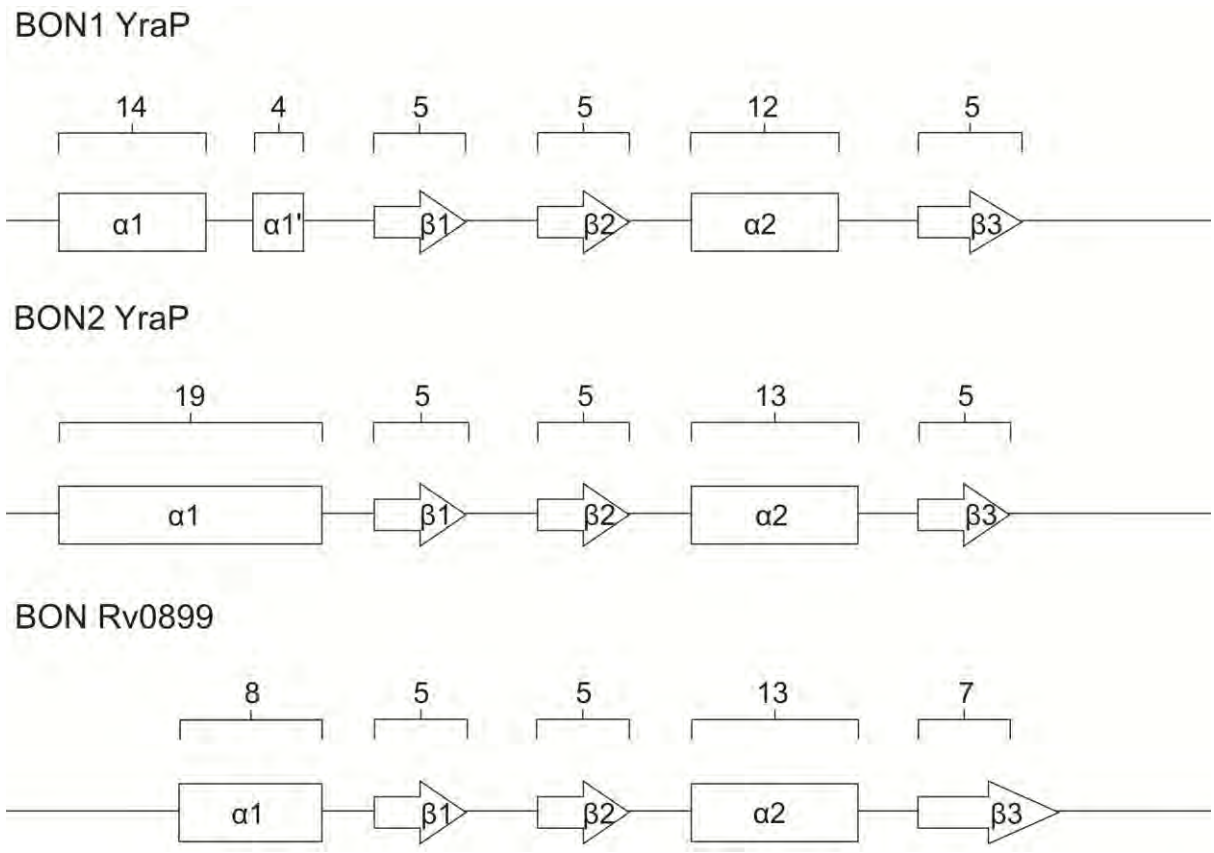


Figure 5.31 Schematic comparing the secondary structure of known BON domain structures. Secondary structure schematics are shown for BON domains 1 and 2 of YraP from *E.* coli and the BON domain of Rv0899 from *Mycobacterium tuberculosis*. α-helices are represented by rectangles and β-strands by arrows. The number of residues of each secondary structure element is displayed above them.

Figure shows a secondary structure comparison of the three known BON domain structures, two from YraP from *E. coli* (BON1 and BON2) and one from Rv0899 from *M. tuberculosis*. All of the BON domains share a αββαβ topology. A key difference between the domains is the α1 helix. In BON1 of YraP, the α1 helix is separated into 2 helices by a flexible region. This does not happen in BON2, which has a longer uninterrupted α-helix of 19 residues. α1 is much shorter in the BON domain of Rv0899, but it should be noted that the overall length of its BON domain is much shorter, approximately 46 residues compared to 61 residues for the BON domains in YraP. The other structural elements are roughly the same length.

The two BON domains are organised in an almost symmetrical manner with the β sheet content of each domain facing each other and the α helices on the outside. The BON domains themselves are orientated in an anti-parallel fashion, as evidenced from the different directions that the relative β-strands face (e.g. β1 faces an opposite direction to β4).

Interactions were observed between the BON domains that were mediated by Y75. NOE resonances were identified between Y75 and the residues, G160, L161, V162 and T188 (**Figure 5.32**). From the structure, G160, L161, V162 and T188 seem to be in close enough proximity to mediate interactions with Y75.

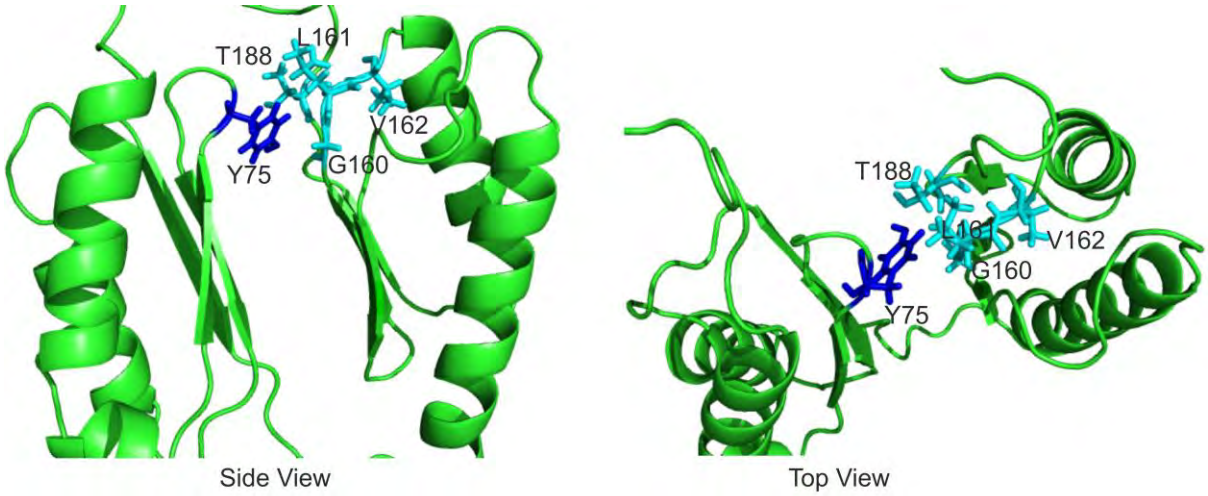


Figure 5.32 Possible residues that interact from BON2 to Y75 of BON1. Y75 (coloured blue) and other potential interacting residues, G160, L161, V162 and T188 (coloured cyan).

The hydrophobic surface of YraP was examined (**Figure 5.33**). There are some potential hydrophobic pockets present in each of the of the BON domains, which may be involved in ligand binding. The two that are present in BON1 are formed of residues L90, A92, A94 and I97 and the second site consisting of residues V101, A104 and V106. In BON2, there are also two potential hydrophobic binding sites consisting of residues F156, L157, M158, G159, L160 and V161 and the second site with F187, F189 and I190.

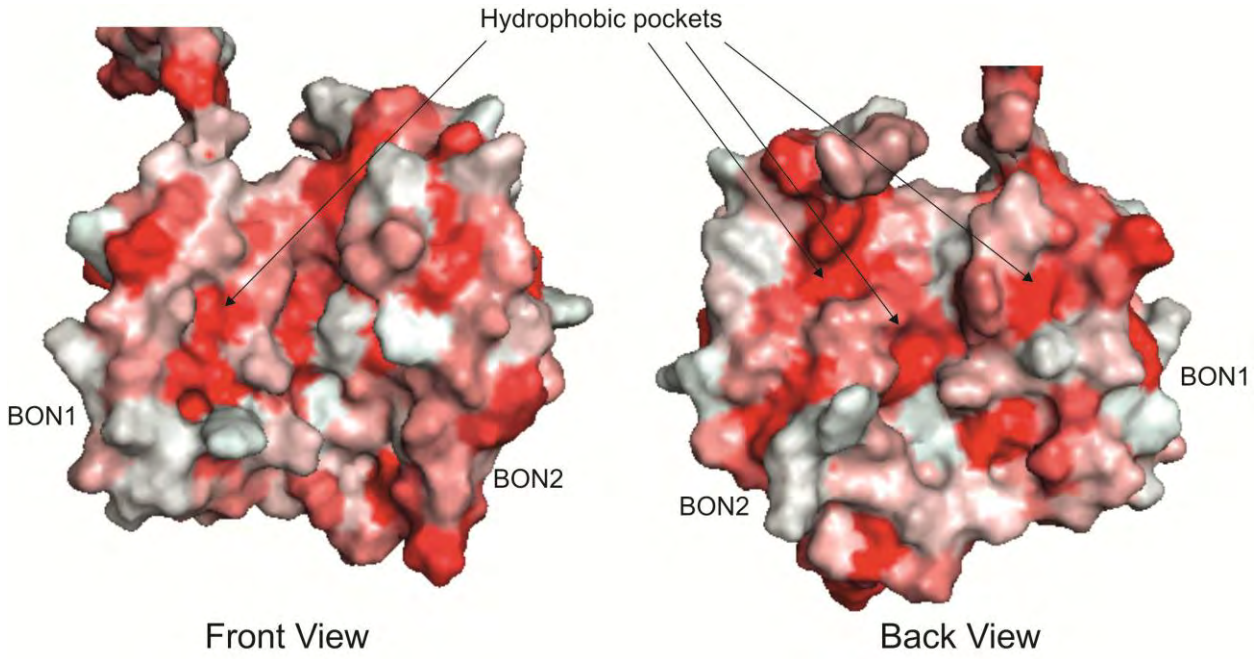


Figure 5.33 Hydrophobic surface map of YraP. The hydrophobicity of the residues are coloured according to normalised scale according to Eisenberg *et al*., 1984, with strongly hydrophobic residues coloured red and strongly hydrophilic residues coloured white. Possible hydrophobic pockets are highlighted.

5.4 DISCUSSION

One of the major aims of this project was to solve the structure of YraP. From initially starting with an expression able to produce soluble YraP, then over-expressing in labelled media, purifying and performing backbone, side chain and NOESY NMR experiments and finally structure calculations, the solution structure of YraP has been solved.

The use of pET26b-YraP has been essential in the expression of YraP. This vector has some modifications to the native YraP protein which has made it ideal to work with. The removal of the N-terminal lipobox signal sequence has prevented acylation of C19 and lipoprotein maturation. This combined with the pelB leader sequence has allowed periplasmic transportation of the protein and hence correct folding. The overall effect of these modifications has converted a lipoprotein to a fully soluble protein, which is folded in its native environment of the periplasm. This strategy has led to high level expression whilst preventing issues such as inclusion body formation and aggregation.

The combination of Ni-affinity and size exclusion chromatography has yielded a pure and soluble product. The buffers used in these purifications have not caused any issues with the protein. From a 2 litre LB growth, approximately 1.2 ml of 1.4 mM YraP (29.7 mg) was able to be produced. There was no noticeable a decrease in yield when grown in M9 media supplemented with ^{13}C glucose and ^{15}N ammonium sulphate. The protein produced from this purification is folded, which was confirmed by 1D NMR. Furthermore, YraP is very stable at room

temperature (25^{o}C) under the buffer conditions used (50 mM sodium phosphate, 50 mM NaCl, pH 6.0). Over the course of this study, YraP has not degraded under these conditions, which is especially important during certain NMR experiments, where the collection of data can take up to a week (backbone, NOESY experiments). We did not observe any decrease in NMR spectra quality performing these experiments.

The HSQC of YraP is well dispersed, displaying resonances for most of the NH groups. Both α-helical and β-sheet content could be observed, which matched with later secondary structure predictions and the solved structure of YraP. Using the HSQC and the backbone experiments [HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HNCO and HN(CA)CO], almost complete assignment has been achieved. Only the NH group of N153 could not be found in the HSQC, but its other atoms (CA, CB and CO) had assignments. From the structure N153 is present in a flexible loop region. A likely explanation for why N153 couldn't be found is because this residue might exhibit dynamics which would cause the intensity of its peak to weaken beyond detection. There is also the chance that N153 might have overlapping signals with another peak in the HSQC and hence is not easily detected.

Secondary structure predictions of the chemical shift data using TALOS, CSI and ΔCα-ΔCβ have identified 4 α-helices and 6 β-strands, with the possibility of a break in the first α-helix. This is somewhat different from the secondary structure predictions using the protein sequence (**Chapter 4**). However, the predictions using TALOS, CSI and ΔCα-ΔCβ are based on experimental chemical shift data

and are much more accurate. These predictions have ensured that the backbone assignments of YraP are correct, as well as the TALOS data being used as a strong dihedral angle restraint in the structure calculations. The secondary structure predicted from these methods are seen in the final structure of YraP, further demonstrating the accuracy of dihedral restraints used from the chemical shift data.

The side chain data from the C(CO)NH and H(CCO)NH experiments has been useful, with approximately 80% sequence coverage of YraP. Analysis of these experiments was hindered by the lack of assignments in regions V107-I111 and T151-E155. The lack of assignments would suggest these regions are in flexible loop regions. The final structure shows that T151-E155 is in a flexible loop region between β4 and β5, whereas V107-I111 are part of β3. Why V107-I111 was unable to be assigned in these spectra could possibly be due to uneven transfer of magnetisation using these experiments. However, assignments of these two problematic regions were resolved using the ^{13}C edited NOESY spectrum, which was also used to confirm that the other side chain assignments were correct alongside with the HCCH-TOCSY spectrum.

The NOESY data has allowed powerful 3D constraints for structure calculations from amide, aliphatic and aromatic protons using the ^{15}N edited, ^{13}C edited and aromatic NOESY experiments respectively. With the ^{15}N edited NOESY, approximately 4,500 peaks were picked, the ^{13}C edited NOESY had approximately 5,500 peaks and there were approximately 80 peaks from the

aromatic NOESY. Assignments were performed on each of these spectra. The majority of assignments were found on the ^{15}N edited NOESY, with problems with the overlapped regions that were present on the HSQC. However, the deuterated ^{15}N edited NOESY produced a much simpler spectrum of approximately 450 peaks, which allowed unambiguous assignment of the HA groups. In the ^{13}C edited NOESY, although it was assigned, there were regions of overlap which may have contributed to ambiguous assignments. As an example, the majority of the glycine residues had overlapping chemical shifts. Furthermore, the water signal was present prominently in this spectrum, making assignments close to this region difficult. Finally, the aromatic NOESY allowed the assignment of 4 of the 5 aromatic residues in YraP (Y75, Y108, W127 and F189), although we were unable to assign F187. The structure indicates that F187 is present in the final loop region and show dynamics which led to signal weakening of this residue.

Structure calculations of YraP have been an iterative process, in which erroneous structures were improved upon by increasing the quality of the data. Initial structure calculations were performed with all of the NOESY data, which yielded ball shaped structures that only had the α-helices resolved. Later structures were improved by the addition of assigning unambiguous NOEs peaks. Further improvements were made by removing excessive peaks present in the ^{13}C and ^{15}N edited NOESY spectra. From the original peak count, there were many peaks picked that had very weak intensities, just above the noise the baseline. Removing these improved the calculated structures greatly, as there no longer any spurious restraints that had to be accommodated for. Additionally, there were

several peaks that had large intensities. Many of these were through bond interactions of neighbouring protons, whereby assignment of them would help in generating the intermediate structures. However, many of these strong peaks were present very close to the diagonal and have either an unlikely intensity or were just noise from the diagonal. Signal overlap also contributed to this problem, where several peaks were on top of each other, which would give a combined intensity for the multiple peaks, when the real intensity of the individual peaks would be a fraction of this value. By removing these overly strong peaks, this led to more accurate structure calculations as these tight restrains would not be present to skew the structure into unfavourable conformations.

The use H-bonds as a restraint was invaluable in solving the structure. Initial calculations resolved the α-helices immediately. By examining the structures and residues in the H/D exchanged HSQC, H-bonds could be mapped for the α-helices. Using these restrains, anti-parallel pairs of β-strands could be resolved in each of the BON domains. Determining the last pair of parallel β-strands in each of the domains was problematic as these were parallel strands and required the correct orientation for them to be resolved. The use of the ^{15}N deuterated NOESY data allowed the identification of possible amide-amide interactions in those regions. Along with the H/D exchanged HSQC data and the intermediate structures of YraP, the residues potentially involved in H-bonding of these strands could be determined. By using these H-bond restraints and examining if there were any violations from adding in these H-bond restraints, the parallel strands β3 and β6 of BON1 and BON2 respectively, could be resolved.

The removal of the N-terminal unstructured region (residues 20-40) also aided the structure calculations for the intermediate structures. The N-terminus unstructured region had very few NOE peaks, but the presence of these interfered with resolving of the core structure elements. Once the core structure of YraP had been resolved, the unstructured N-terminus was factored back into the structure calculations and it did not affect the resolved core elements.

The NMR structure of YraP represents the first tandem BON domain structure of any protein. The solved structure contains an unstructured N-terminal region and two distinct domains. The individual domains represent a compact box-like structure consisting of 2 α-helices and 3 β-sheets each, with parallel orientations of these secondary structure elements. They are also organised in a manner that the α-helices face one side and the β-sheets are on another side. Both of the BON domains themselves look very similar, which is expected due to the high level of sequence homology between them.

The small 90^{o} loop that is present in α1 of BON1 that breaks the helix and produces another smaller helix, α1'. As α1' retains its n+4 periodicity, it is not a 3_{10} helix. The major residue involved in this break is D61. The NOESY data in this region is good with clear, unambiguous assignments made for this residue. There are very few ambiguous assignments as well as very little overlap in this region. Also in the majority of our calculations, this structural element has appeared, even in the early calculations. It is very likely that this structural feature is real and not an artefact from the structure calculations. This feature might be in place to aid the fold of BON1, which immediately has β1 after this region, which

is orientated in a parallel manner. A sharp 90^{o} turn from a flexible loop region followed by a small stretch of α-helix and then another 90^{o} sharp turn, is one solution to accommodate the parallel orientation of the β-sheet and α-helix. In BON2, this feature is not present and BON2 has a longer loop region after its α1 helix, which is able to fold around to orientate its β1 in a parallel manner. This is more similar to the already solved BON domain structure of rv0899 of *M. tuberculosis* (Teriete *et al*., 2010), which lacked a helix that was interspersed with a loop region. This could be a feature that is exclusive to YraP and may play a functional role. However further BON domain containing structures of other proteins would need to be solved to determine if this is the case.

The structure calculations have revealed that Y75 might be involved in mediating the interactions between BON1 and BON2. NOE peaks have been observed between Y75 and T150, G160, L161, V162 and T188. From the structure, all of these residues are facing each other. The possibility of interactions between Y75 and G160, L161, V162 and T188 seem to be feasible, as these residues are in close proximity and within the 5 Å limit for NOE signals and it seems very likely that Y75 could have a role in keeping the 2 BON domains together and orientated correctly.

The orientations of the BON domains themselves are orientated in a similar manner, with α-helices and β-sheets facing in the same directions, producing an almost symmetrical protein as the overall fold for YraP. This finding could be a common fold for proteins that contain multiple BON domains. As no other two BON domain architecture proteins have been solved, the structure in this study

cannot be compared to any other structures and thus structure checking programs such as Dali, cannot be used effectively. There may be multiple orientations in which the 2 BON domains can conform to. In order to confirm that the correct orientation between the 2 BON domains had been solved in our structure, we aimed to perform residual dipolar coupling (RDC) experiments, which would give the orientations of each of the domains. Unfortunately, we were unable to carry out these experiments as YraP was able to interact with the alignment media used (Pf1 phage and phospholipids).

Small angle X-ray scattering (SAXS) analysis of YraP was performed by Dr. T. J. Knowles. In brief, SAXS is a low resolution structural biology technique that can be used with proteins in solution. SAXS analysis generates a 3D molecular envelope, which can be used to 'fit' structures within its 3D set of coordinates. From the SAXS analysis, YraP can be fit into the molecular envelope, suggesting that the solved NMR structure of YraP is accurate **Figure 5.34**).

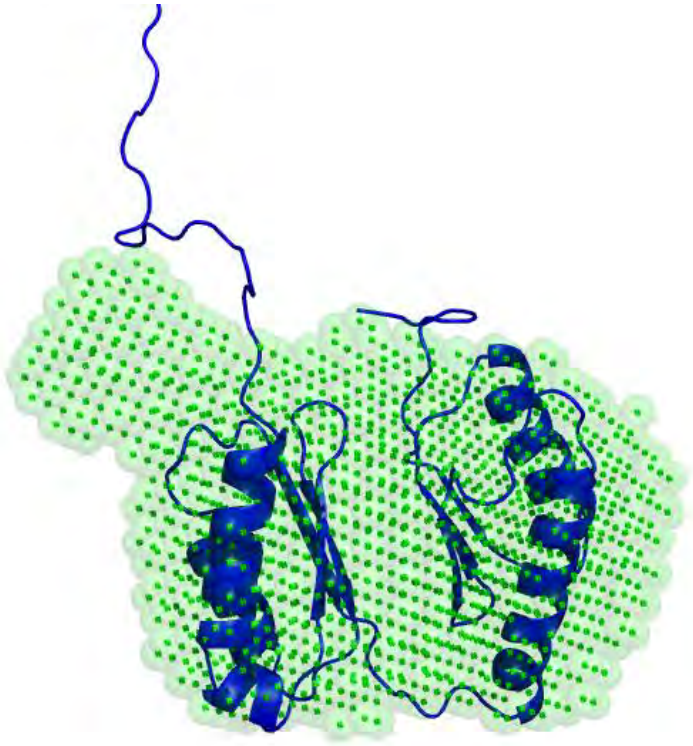


Figure 5.34 The solved NMR structure of YraP can be fitted into the SAXS molecular envelope. Damfiltered molecular envelope of YraP (green) superimposed with the NMR structure of YraP (blue). Damfilter is a modelling parameter that displays the regions of overlap from 10 independently determined SAXS models of YraP from the same dataset and removes regions of non-overlap from the model.

Chapter 6

Biochemical and functional analysis of YraP

6.1 Introduction

Although the structure of YraP has been solved, the function of *yraP* has still not yet been elucidated. The structure shows that YraP contains two BON domains, but due to a lack of similar known structures present, as well as very little information on known homologues, no apparent predictions can be made as to what this protein does.

The literature has hinted at possible roles that *yraP* may have, but with no real experimental data supporting these roles. Onufryk *et al.* 2005 identified that *yraP* is a member of the Sigma E regulon, suggesting a role in OMP or LPS assembly. Kuntumalla *et al.* 2011 assigned a putative function for *yraP*, as a factor involved in OMP biogenesis. However this was part of a proteomic study in *Shigella dysenteriae*, rather than an actual study of *yraP*, and *yraP* was just placed into a class of genes that may be involved in OMP biogenesis. To add weight to the claim that *yraP* may be involved in OMP biogenesis, Onufryk *et al.* 2005 also demonstrated that a strain harbouring a Δ*yraP* and Δ*surA* genotype, produced a synthetic lethal strain. SurA is a periplasmic chaperone that is involved in OMP assembly. The double knockout strain may suggest that *yraP* is part of this pathway and that the absence of *yraP* has severely disrupted it, producing the lethal phenotype, or it can also suggest that *yraP* and *surA* are part of different pathways, and the combination of disruption to both pathways has caused the observed lethality.

Yeats *et al*. 2003, using software prediction, have suggested that BON domains are able to bind phospholipids. However, to date, no actual experimental data

has been produced that definitively demonstrates a BON domain binding to phospholipids. Teriete *et al.* 2010 solved the first BON domain structure of the protein Rv0899 in *Mycobacterium tuberculosis*. This protein contains both a BON domain and another domain that is homologous to the OmpA-C-like superfamily of periplasmic peptidoglycan-binding proteins. Initially, they thought that this protein would form a membrane embedded β-barrel, but their solved structure indicated that this was not the case, but rather that it folded into a mixed α/β structure. Bioinformatic analysis on the BON domain from multiple sequence alignments revealed the presence of a conserved glycine residue at the end of the second β-strand. However, no experiments were performed on this finding, and it was only hypothesised that this glycine plays a crucial mechanical role rather than having any functional activity.

In order to elucidate a function for YraP, we undertook several genetic and biochemical experiments to probe its functional role. The results of these findings are present in this chapter.

6.2 Results

6.2.1 Confirmation of Δ*yraP* strain

The Δ*yraP* strain used in this study has been taken from the KEIO library (Baba *et al*., 2006), which is a collection of non-essential, in-frame, single gene knockouts. Essentially, the gene in question has been substituted for a kanamycin resistant cassette as described by (Datsenko and Wanner, 2000). Over the course of this study, inconsistencies have been observed working with the Keio library. To ensure that the *yraP* gene has been substituted for the kanamycin resistance gene cassette, *aph*, a set of PCR reactions were performed with primers designed upstream and downstream of the *yraP* gene and primers that are internal to the kanamycin cassette.

Using wildtype (WT) BW25113, and Δ*yraP* cells, PCR reactions were performed with *yraP* forward/kanamycin reverse (YF/KR) and *yraP* reverse/kanamycin (YR/KF) forward primer combinations (**Figure 6.1**) (primer sequences are present in **Table 2.5.4.3**). The *yraP* forward primer is present 175 bp upstream of the gene start point, and the reverse primer is present 153 bp downstream. The kanamycin cassette is 1452 bp long, and the kanamycin primers are present internally. Using these primer combinations, the YF:KR should give a product of 1027 bp and the YR:KF 852 bp. Products of these sizes indicate that the kanamycin cassette has correctly substituted the *yraP* gene. If lacking the kanamycin resistance cassette, no PCR product should be observed.

Gel analysis of the PCR reactions demonstrate that nothing is seen in the WT reactions, indicating that no kanamycin cassette is present, however in the Δ*yraP*

reactions, bands of the correct sizes are observed. This indicates that *yraP* has correctly been substituted for a kanamycin cassette in this Δ*yraP* strain.

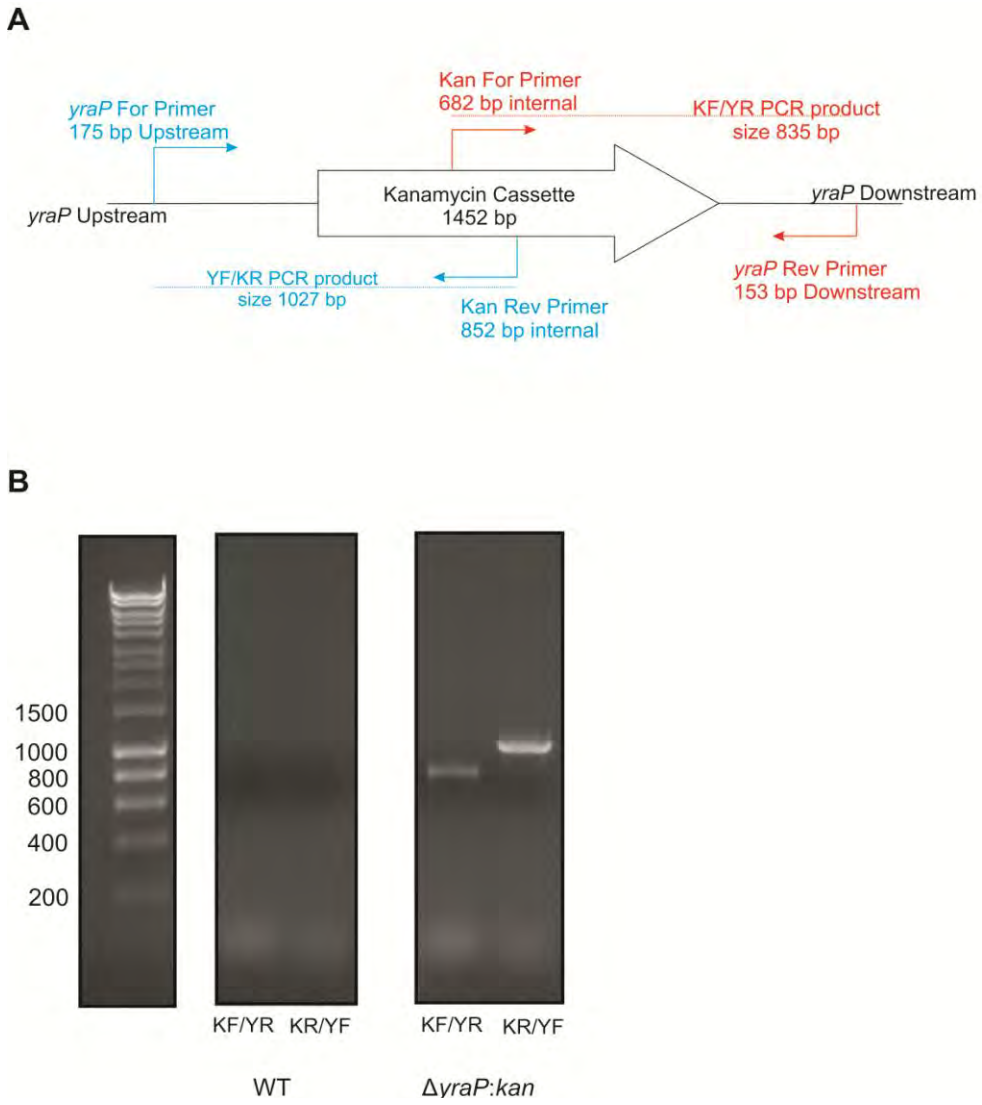


Figure 6.1 The *yraP* gene has correctly been replaced with a kanamycin resistance cassette in Δ*yraP* cells.

A Schematic of the expected kanamycin cassette substitution for *yraP* in Δ*yraP* cells. The schematic shows approximate positions where each of the primers anneal to and the estimated sizes for PCRs using KF/YR (red) and KR/YF (blue) primer pairs which are used to confirm a correct mutant. **B** 1% Agarose gel electrophoresis of the PCR products of KF/YR and KR/YF primer pairs in WT and Δ*yraP* cells. In WT cells, no products are observed for either of the reactions. With Δ*yraP* cells, PCR of KF/YR and KR/YF primer pairs have produced products of approximately 852 and 1027 bp, indicating that the *yraP* gene has been correctly substituted for a kanamycin primer in the correct orientation in Δ*yraP* cells.

6.2.2 Cloning of *yraP*

In order to study *yraP*, the full length gene was cloned into the expression vector pET20b. Primers were designed to amplify *yraP* from BW25113 genomic DNA by PCR, to produce full length wildtype and C-terminally His-tagged YraP respectively. The size of the *yraP* gene is 576 bp. From these PCR reactions, bands are seen at just below the 600 bp marker, indicating that the fragment has been amplified (**Figure 6.2**).

The fragments and pET20b vector were digested with appropriate restriction enzymes, ligated and transformed into XL1-Blue cells. Plasmids were extracted from the transformants and the correct constructs were determined by di-deoxy chain termination sequencing (Functional Genomics, Birmingham). Using this approach, constructs that can express full length wildtype and C-terminally His tagged YraP were generated.

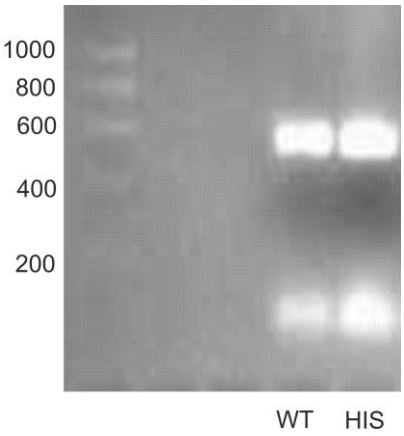


Figure 6.2 PCR for cloning WT and His-tagged *yraP* was successful
PCR of the amplified *yraP* products from genomic K-12 DNA. Both the WT and His-tagged primers have amplified a product just under 600 bp marker. The PCR reactions are pure, with no other products seen, except for free nucleotides at the bottom of the gel.

6.2.3 Δ*yraP* cells have an OM defect

As has been previously reported, Δ*bamE* cells display outer membrane phenotype that can be complemented by using a pET20b-*bamE* plasmid. We wanted to examine if this was the case with Δ*yraP* cells, in that an apparent OM defect could be observed under plating conditions and could be complemented by plasmid borne expression.

BW25113 Δ*yraP* cells were transformed with pET20b (a negative control) and pET20b-*yraP*. Individual transformant colonies were picked, and from a single pick, were streaked onto 4% SDS, 0.5% SDS with 0.5 mM EDTA and 100 µg/ml vancomycin LB-Agar plates, with a positive control 100 µg/ml ampicillin plate included. This was replicated an additional 3 times with other transformant colonies to ensure reproducibility.

BW25113 Δ*yraP* strains with pET20b were unable to grow on the test plates, but showed growth on the positive control ampicillin plate (**Figure 6.3A**). BW25113 Δ*yraP* cells transformed with pET20b-*yraP* were able to grow on all plates, demonstrating that complementation of the OM defect had been achieved.

In order to confirm that the complementation is a result of leaky expression from pET20b-*yraP*, western blot analysis was carried out on wildtype BW25113, Δ*yraP* pET20b and Δ*yraP* pET20b-*yraP* strains, with anti-*yraP* antibody. **Figure 6.3B** shows that WT BW25113 cells are able to produce YraP and that Δ*yraP* cells are unable to do so. With Δ*yraP* pET20b-*yraP* cells, YraP is produced is a slightly elevated level than WT. This indicates that leaky expression from pET20b-*yraP* is

greater than endogenous chromosomal expression in WT cells. The blot also shows that complementation is a result of expression from the pET20b-*yraP* vector.

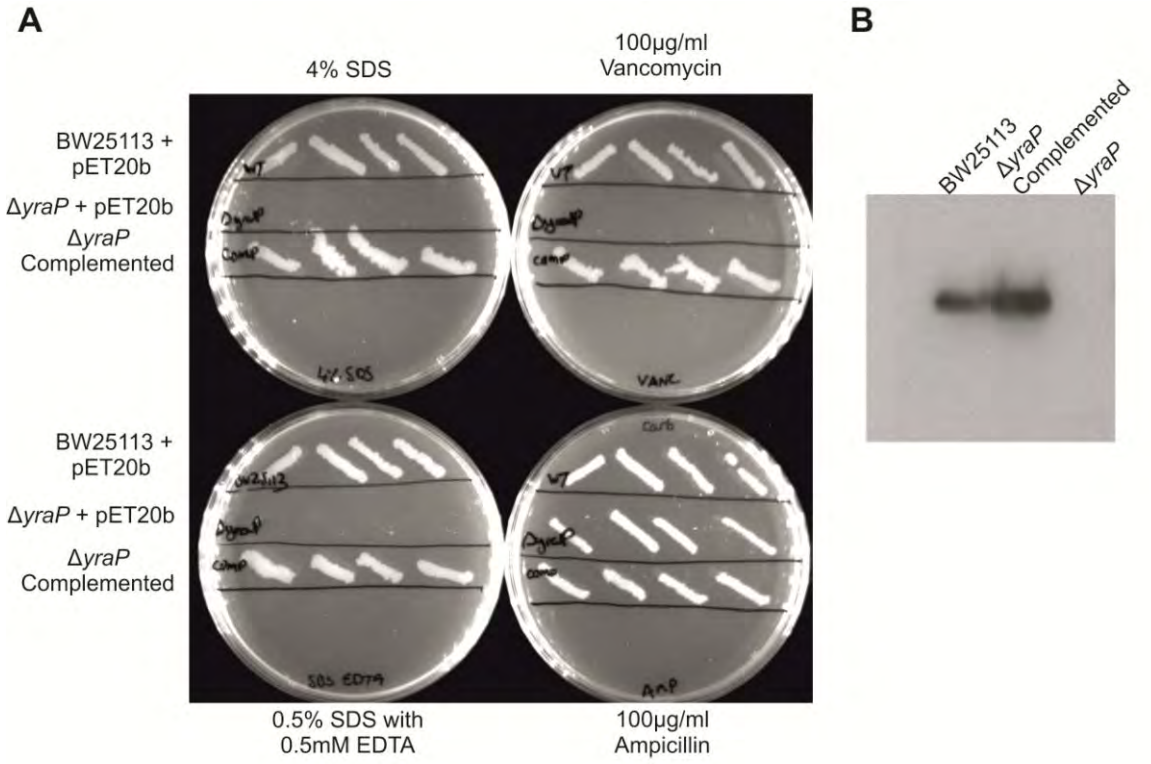


Figure 6.3 Plasmid borne expression of *yraP* from pET20b-*yraP* is able to complement Δ*yraP* defects. A Plating assay performed with streaks of BW25113 and ΔyraP cells transformed with pET20b (labelled BW25113 + pET20b and Δ*yraP* + pET20b), and with complemented Δ*yraP* cells transformed with pET20b-*yraP* (labelled Δ*yraP* Complemented). An individual transformant colony is picked and streaked onto test conditions 4% SDS, 0.5% SDS with 0.5 mM EDTA, 100 µg/ml vancomycin and 100 µg/ml ampicilln as a positive control to ensure that the cells had taken up the pET20b plasmid. Streaks are performed onto the same position on the different plates and are replicated an additional 3 times with different transformant colonies. BW25113 cells are able to grow on the test conditions, whereas Δ*yraP* cells fail to grow. The complemented Δ*yraP* cells transformed with pET20b-*yraP* (Δ*yraP* Complemented) behaves like WT cells, displaying growth on the test conditions. **B** Western blot of the above strains with 20 µl of normalised cell lysates. The blot indicates that pET20-*yraP* expresses in Δ*yraP* cells. The complemented strain is shown to express more YraP than WT, using uninduced leaky expression from pET20b-*yraP*.

6.2.4 YraP is OM localised

The N-terminal lipobox signal sequence of YraP suggests that it is localised to the outer membrane. From a functional perspective, this would make sense as Δ*yraP* strains exhibit outer membrane permeability defects. This has been observed with the accessory lipoproteins of the BAM complex, whereby individual knockouts of *bamB, bamC* and *bamE* also have permeability defects, and have been confirmed as outer membrane localised lipoproteins.

We wanted to confirm that YraP is an outer membrane localised protein. Sucrose density centrifugation was performed on BW25113 cells, which allows the separation of the outer membrane from the inner membrane by buoyancy. Due to the presence of LPS in the OM, which makes it more buoyant than the IM, the OM fraction is present near the bottom of the tube, with the less buoyant IM fraction nearer the top (**Figure 6.4A**). Fractions from the separation were analysed by boiling samples in SDS loading buffer and running SDS-PAGE. The presence of bands at 37 kDa is consistent with monomers of OMPs, most likely OmpA and OmpF, indicating that these are the OM fractions (**Figure 6.4B**). These bands are present in the earlier fractions, which is in accordance with the OM being separated first.

To confirm that YraP is an OM localised protein, western blotting analysis was performed on the fractions from the separation. Anti-BamC antibody (purified from rabbit) used as a positive control, which would highlight the OM fractions, with anti-YraP antibody (purified from rabbit), which would indicate which fractions YraP is present in. The Western blot shows that YraP is present in the

same fractions as BamC, with both proteins tailing off in intensity at the same fractions (**Figure 6.4C**). Although an IM marker is absent, this data suggests strongly that YraP is OM localised.

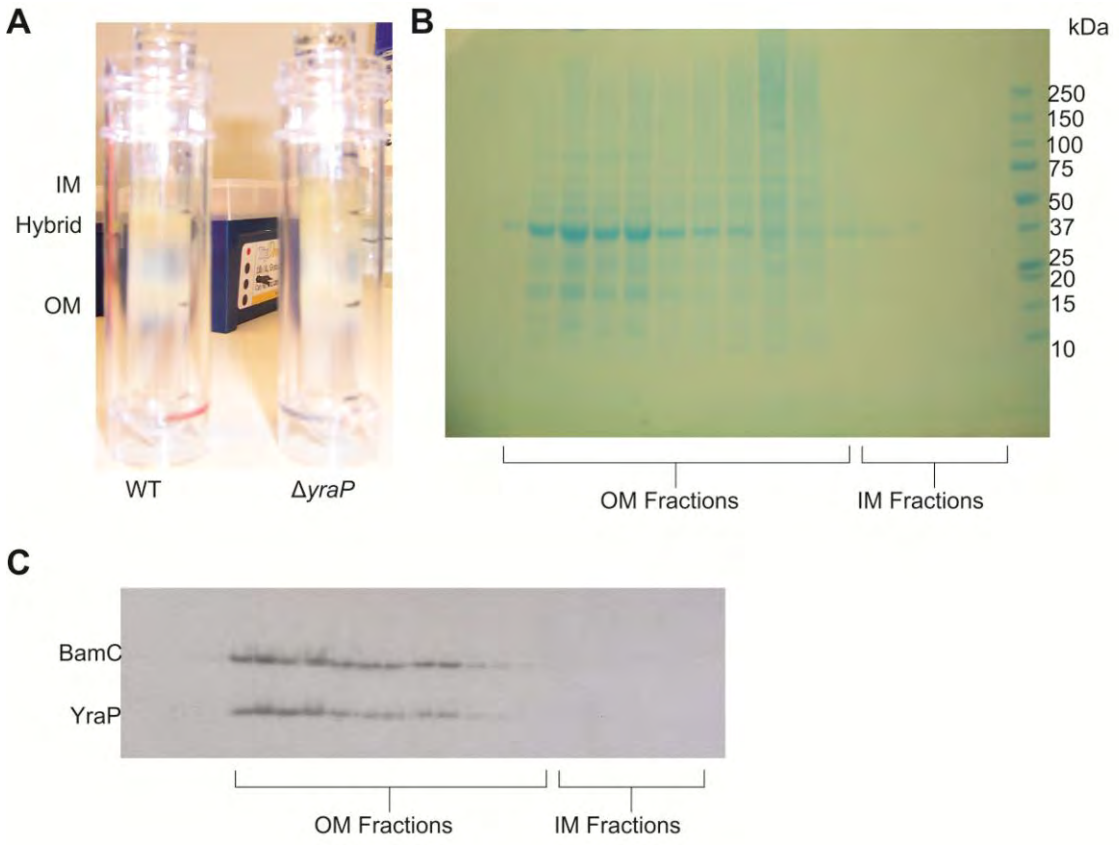


Figure 6.4 YraP is localised to the OM. A Sucrose density centrifugation performed on BW25113 cells. The OM is the fraction nearest to the bottom of the tube at the 42%/47% sucrose boundary. The IM is near the uppermost band and the IM/OM hybrid is seen in between the two bands. **B** Coomassie stained SDS-PAGE of samples taken from the centrifugation. 20 µl was taken from each 0.5 ml collected fraction, loaded in 2 x Laemmli buffer and boiled. The presence of OmpA and OmpF monomers are seen as bands at 37KDa, indicating that these are OM fractions. **C** Western blot of the sucrose density centrifugation fractions, probed with anti-YraP and anti-BamC antibodies. YraP is present in the same fractions as BamC, which is known to be an OM lipoprotein, strongly suggesting that YraP is localised to the OM.

6.2.5 YraP is not surface exposed

The sucrose density gradient result suggests that YraP is an OM localised protein. There is a possibility that YraP could be a surface exposed protein. Although far less frequent than periplasmic lipoproteins, surface exposed lipoproteins have been reported (Tokuda, 2009). Recently BamC has been shown to be surface exposed, possibly changing the current understanding of how the BAM complex may function (Webb *et al*., 2012).

In order to determine if YraP is surface exposed, BW25113 cells were incubated with anti-YraP antibody, which was then incubated with anti-rabbit-Alexa Fluor 488 conjugated antibody. This would in essence label any surface exposed YraP. As a positive control, membrane permeabilised cells were used, which would allow labelling of YraP inside the cells. Using fluorescence microscopy, YraP could be detected in the cells (**Figure 6.5**).

The fluorescence microscopy data shows that no fluorescence is observed for the non-permeabilised cells. However, with the permeabilised cells, fluorescent signal is observed, indicating that the labelling procedure has worked but that YraP is not a surface exposed protein.

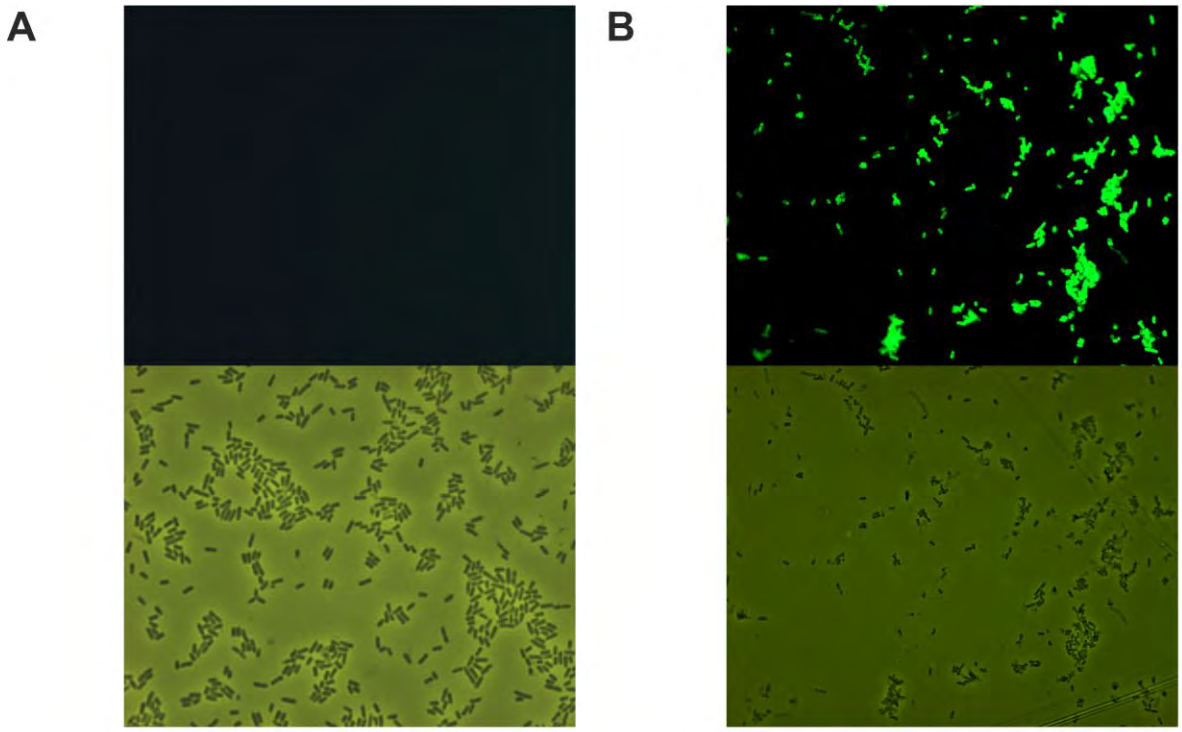


Figure 6.5 YraP is not a surface localised protein. Fluorescence microscopy performed with WT BW25113 cells using anti-YraP antibody, labelled with anti-rabbit-Alexa Fluor 488 conjuated antibody. **A** Unpermeabilised cells showing no fluorescence signal, indicating that YraP is not surface localised. **B** Permeabilised cells displaying fluorescence signal, indicating that YraP is internally localised. Images without fluorescent excitation are also shown.

6.2.6 YraP is monomeric in solution

In order to determine if YraP is able to self associate, analytical ultracentrifugation (AUC) was used to investigate the multimeric state of the protein. AUC sedimentation velocity experiments were ran with 80 μM His-tagged YraP (Abs_{280} = 0.7) in 50 mM sodium phosphate buffer and 300 mM NaCl.

Analytical ultracentrifugation allows various properties of proteins to be studied, such as molecular shape and size, as well as the distribution and interactions in

regards to the presence of different species, all within a solution environment. In the case of monitoring for multimeric interactions, the protein itself will give rise to peak of a certain Sverdburg (S) value. However, if oligomerisation occurs, peaks of larger S values will be observed. By observing the trace and integrating the peaks, it is possible to estimate if a protein is able to self-associate as well as determining the ratio between monomeric and oligomeric species.

Using SEDFIT to analyse the data, only one peak is observed at approximately 1.76S, which corresponds to a mass of approximately 23.3 kDa for a globular protein (**Figure 6.6A**). This result suggests that YraP is monomeric in solution. This data corroborates with elution profiles that are produced when performing gel filtration with purified YraP (**Figure 6.6B**). In the elution profiles, only a single peak is observed, and comparison to molecular weight markers further confirms that YraP is monomeric and is not able to self associate.

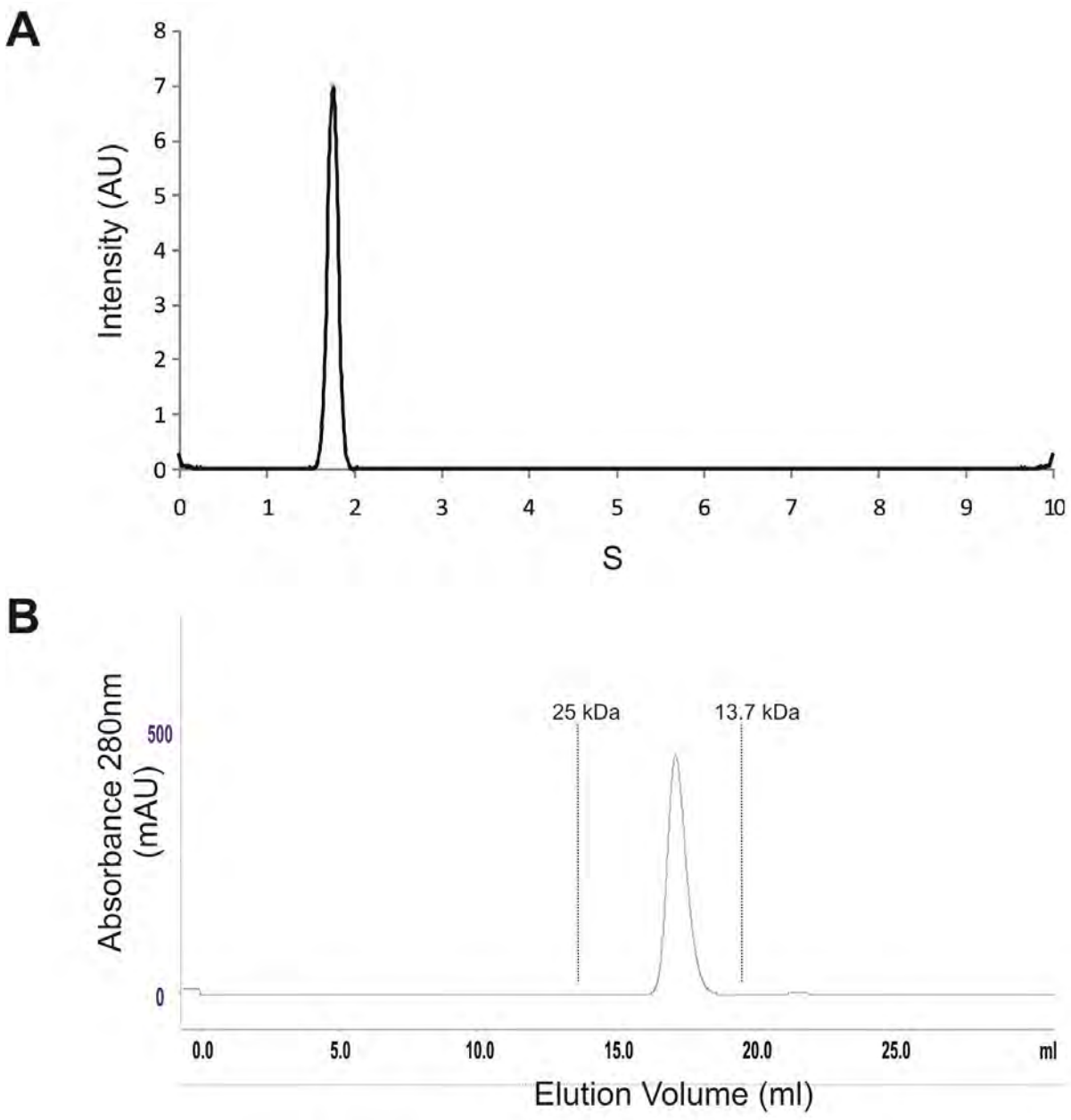


Figure 6.6 YraP is monomeric in solution A. SEDFIT analysed data of the sedimentation velocity run performed on 80 μM ΔN-YraP-C-His. The analysis shows only a single peak is observed at a 1.76 S (approximately 23.3 kDa), indicating that YraP is monomeric in solution. **B.** Elution profile of purified ΔN-YraP-C-His (350 μM), performed on a Superdex S200 column. YraP purifies as a single peak at an elution volume of 18 ml, corresponding to a mass of 21.4 kDa following column calibration, further demonstrating that it is monomeric. Molecular weight markers of 25 kDa and 13.7 kDa protein standards are shown.

6.2.7 YraP may not be a member of a protein complex

To gain a possible insight into the function of *yraP*, co-immunoprecipitation (Co-IP) experiments were performed to determine if there are any protein binding partners to YraP. This could potentially highlight YraP as a new factor in an already studied system or identify partners in a new previously unknown pathway. By using an anti-YraP antibody, which would bind to the bait protein, with DSP crosslinker to enhance protein-protein interactions, possible binding partners to YraP could be captured on Protein-A sepharose beads and observed by SDS-PAGE.

As YraP is predicted to be an outer membrane lipoprotein, Co-IPs were performed on purified outer membranes from WT and Δ*yraP* cells. As a positive control, anti-BamC antibody was used for one Co-IP, to see if the BAM complex could be immunoprecipitated using this method.

The gel from this Co-IP indicates that the procedure has worked, as members of the BAM complex, BamB and BamE, can be pulled down with anti-BamC antibody (**Figure 6.7A and B**). The Δ*yraP* fraction shows no bands, except for those belonging to the antibody light and heavy chains. With the WT fraction, only YraP is pulled down, but no other bands are seen. The same is observed with the cross-linked WT fraction, although less YraP is pulled down. This result suggests that YraP is not part of any OM complex.

As the previous Co-IP was performed with purified OM fractions, there is a possibility that YraP binds to other proteins in different subcellular localisations.

To test this possibility, Co-IPs were performed on WT and Δ*yraP* cells using whole cell lysates. The gel from this Co-IP (**Figure 6.7C**) showed the presence of bands that were present in the WT and WT cross-linked samples and that were not present in the Δ*yraP* sample. Bands were excised, digested with trypsin and identified by mass spectrometry. The major protein species that were present in WT samples and not in Δ*yraP* samples was YraP, which indicated that the identification procedure was accurate and another protein, YgaU.

A conserved domain search of YgaU predicts that it contains a BON domain and LysM, peptidoglycan binding domain (**Figure 6.8A**). YgaU has 14.3% sequence identity to YraP. As YraP contains two BON domains and is predicted to be in contact with the periplasm, where peptidoglycan is present, YgaU could be a potential binding partner to YraP. C-terminally his-tagged YgaU was cloned into the expression vector pET22b, expressed in BL21 (DE3) cells and purified. Binding studies were performed by NMR. The HSQC for ^{15}N-YraP in the presence and absence of YgaU, the results show that there were no chemical shift perturbations following the addition of 2:1 YgaU, indicating that these proteins were not interacting (**Figure 6.8B**).

As this Co-IP procedure can lead to the identification of false positives, the method was adjusted to prevent this reduce this problem. It seems highly likely that YraP would interact with proteins in the periplasm. To probe these potential interactions, pET26b-YraP-NMR was transformed into BL21 (DE3) cells, and soluble, periplasmic YraP expression was induced for 4 hours. The periplasm

was extracted by osmotic shock and Co-IPs were performed. BL21 (DE3) cells transformed with pET26b were used as a negative control.

The gel of this Co-IP (**Figure 6.7D**) shows that this procedure is less clean, with the presence of numerous bands in all of the samples. Both the YraP and empty BL21 (DE3) lanes look the same. However, in the cross linked YraP lane, there is the presence of one band just below the 50 kDa marker. Mass spec analysis identified this protein as the maltose binding protein, MalE. As this was an unlikely binding partner, no binding studies were performed with MalE.

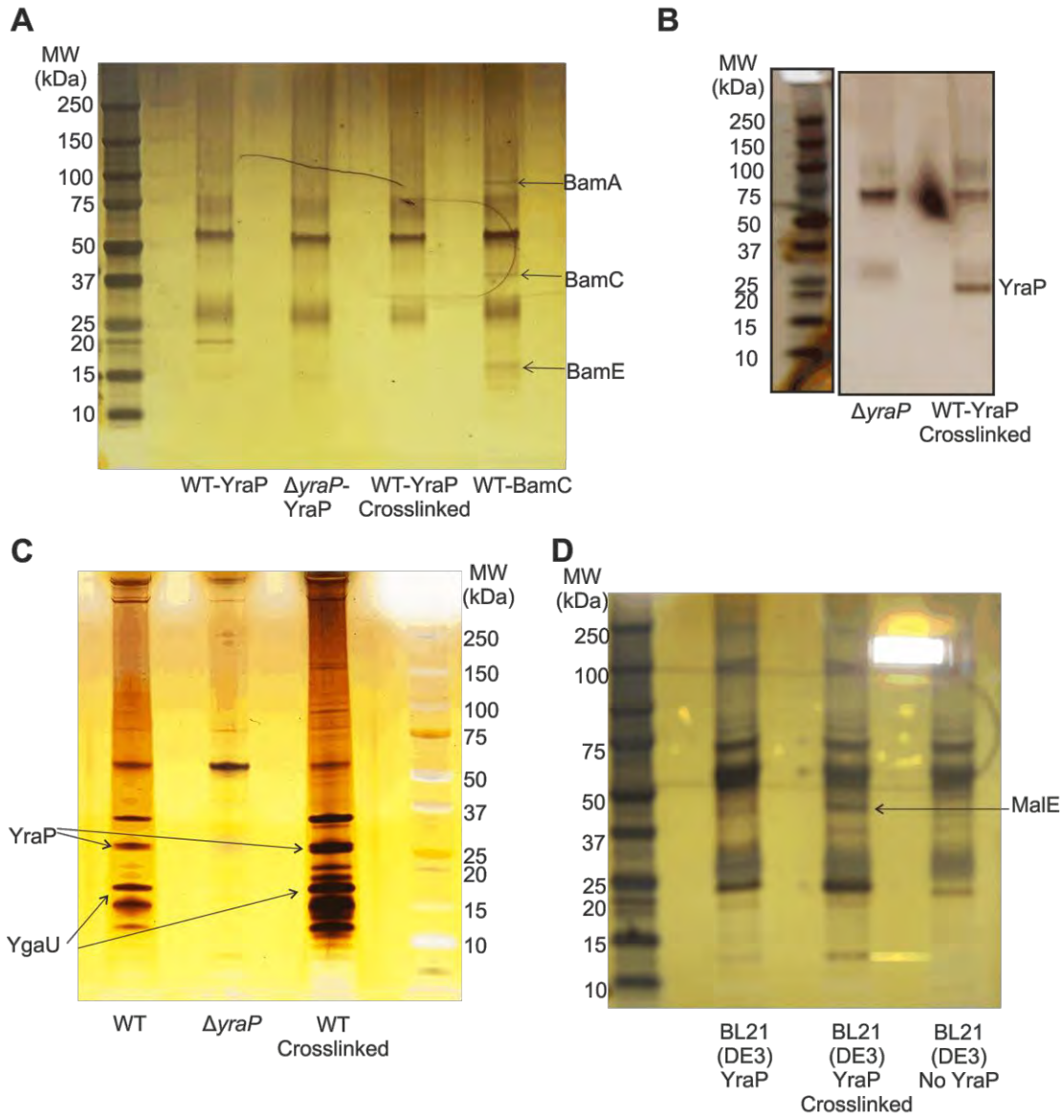


Figure 6.7 YraP seems unlikely to interact with any specific proteins from co-immunoprecipitation experiments. Silver staining SDS-PAGE of Co-IP experiments. **A** Co-IPs using OM fractions from WT and Δ*yraP* BW25113 cells with anti-YraP antibody from 20 µl of eluate (WT-YraP and Δ*yraP*-YraP). DSP cross linker was also used on WT cells (WT-Yrap Crosslinked), which did not work in this gel. A positive control Co-IP was done using anti-BamC against WT OM fractions (WT-BamC). Only the BamC Co-IP has pulled down other proteins, which are likely to be members of the BAM complex, from their sizes. **B** A repeat of the WT-YraP crosslinked sample is shown on the gel to the right. No other additional proteins are seen in this eluate. **C** Co-IPs with whole cells on WT, Δ*yraP* and WT Crosslinked samples using anti-YraP antibody and running 20 µl of eluate. Numerous bands are pulled down in the WT samples, with one band containing YgaU. **D** 20 µl Co-IP eluates from samples using anti-YraP antibody with periplasmic fractions from BL21 (DE3) cells expressing periplasmic YraP as bait

with and without crosslinker [BL21 (DE3) YraP and BL21 (DE3) YraP Crosslinked respectively]. A negative control is used with BL21 (DE3) cells not expressing periplasmic YraP [BL21 (DE3) No YraP]. Numerous bands are present in all of the samples indicating that this approach is not as clean as the other Co-IPs. Only one band is prominently observed in the BL21 (DE3) YraP Crosslinked sample, which was identified as MalE.

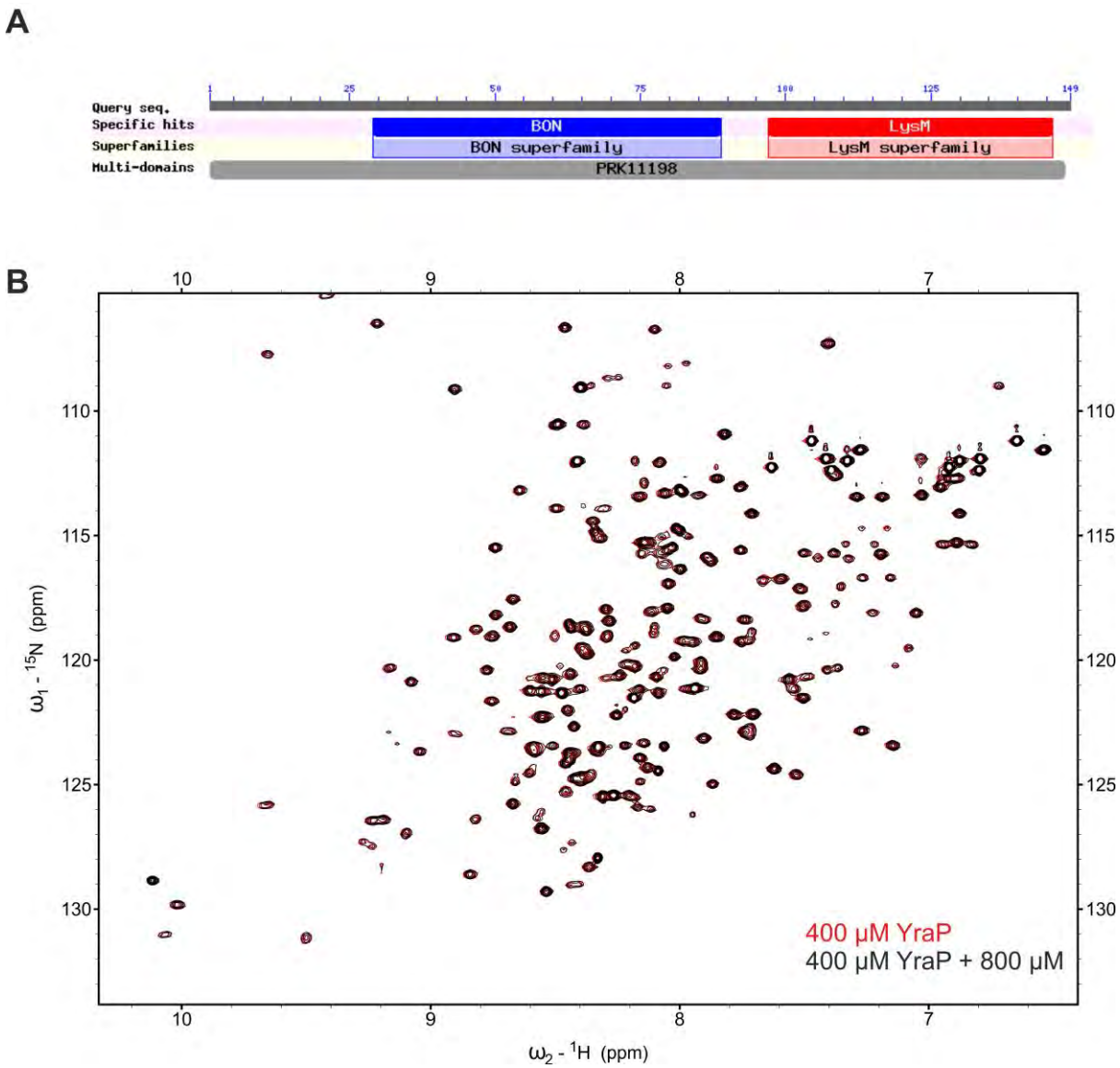


Figure 6.8 YraP does not interact with YgaU. **A** Domain architecture of YgaU. **B** HSQCs performed with 400 µM ^{15}N-labelled YraP (red) and with 800 µM YgaU (black) in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 at 25^{o}C using an Inova Varian 800MHz spectrometer.

6.2.8 YraP does not interact with SurA

It has been shown that a double knockout of *yraP* and *surA* yields a synthetic lethal phenotype (Onufryk *et al*., 2005). As SurA has been identified as a periplasmic chaperone and *ΔsurA* strains display outer membrane defects similar to *ΔyraP* strains, it seems likely that there could be overlapping functional roles.

This finding can indicate two likely possibilities for the role of *yraP* with regards to SurA. Either YraP and SurA function in the same pathway, and the synthetic lethal phenotype is the result of severely disrupting this pathway, or that YraP and SurA are involved in two separate pathways and this phenotype is due to the culmination of the defects in the two pathways. In order to examine if this is the effect of the former possibility we decided to perform binding studies with YraP and SurA using a variety of methods.

We did an NMR two point titration by performing an HSQC using 300 mM ^{15}N YraP with and without 800 mM SurA and examined the possibility of binding by observing changes in peak intensities or peak movement (**Figure 6.9A**). Our analysis from the HSQC spectra showed that there were peak decreases for many of the assignments as a result of line broadening. However, there were a few peaks that showed a similar intensity to the spectrum without SurA, indicating that this was not an effect due to YraP concentration (**Figure 6.9B**). As the nature of peak changes are only due to line broadening, this could be an artefact from the changes in the tumbling rate of the sample, which is in turn due to a change in the viscosity of the sample. It should be noted that concentrated SurA

has the propensity to self aggregate, which could explain this observation in the spectrum.

To further examine the possibility of a YraP and SurA interaction, analytical gel filtration was performed using a Superdex S200 column (GE Healthcare), with 135 μM YraP and 270 μM SurA.

Analysis of the individual elution profiles shows that YraP produces a single peak (**Figure 6.10**), while with SurA, there are two major species present, one corresponding to the SurA monomer and another which elutes much earlier, belonging to aggregated or self associated SurA. In the complex mixture, all of the individual peaks remain with no obvious complex formation. There does seem to be some decrease in the intensity of the SurA aggregate, but all of the peaks still elute off at their individual positions. Running SDS-PAGE on these samples, YraP is monomeric, SurA exists in both a monomeric form at 50 kDa and an aggregated forms at 150 kDa. From gel filtration analysis, it seems that YraP does not interact with SurA.

To finally determine if there is a possible interaction in the cell between YraP and SurA, co-immunoprecipiation experiments were performed using whole cell lysates. Anti-YraP antibody was reacted with a wildtype strain and a *Δyrap* strain as a negative control, as well as with a wildtype strain in the presence of with 0.5 mM DSP (cross linker) and an antibody only negative control. SDS-PAGE followed by western blot analysis was performed on these samples, Western blotting was performed on these samples, using anti-YraP and anti-SurA

antibody. From the western blot, antibody heavy chain is seen in all of the samples (**Figure 6.11**). YraP is present in samples that contain YraP (the wildtype samples), but not in the *ΔyraP* sample. SurA is not present in any of the samples. This result indicates that YraP does not interact with SurA, even in the presence of cross-linker.

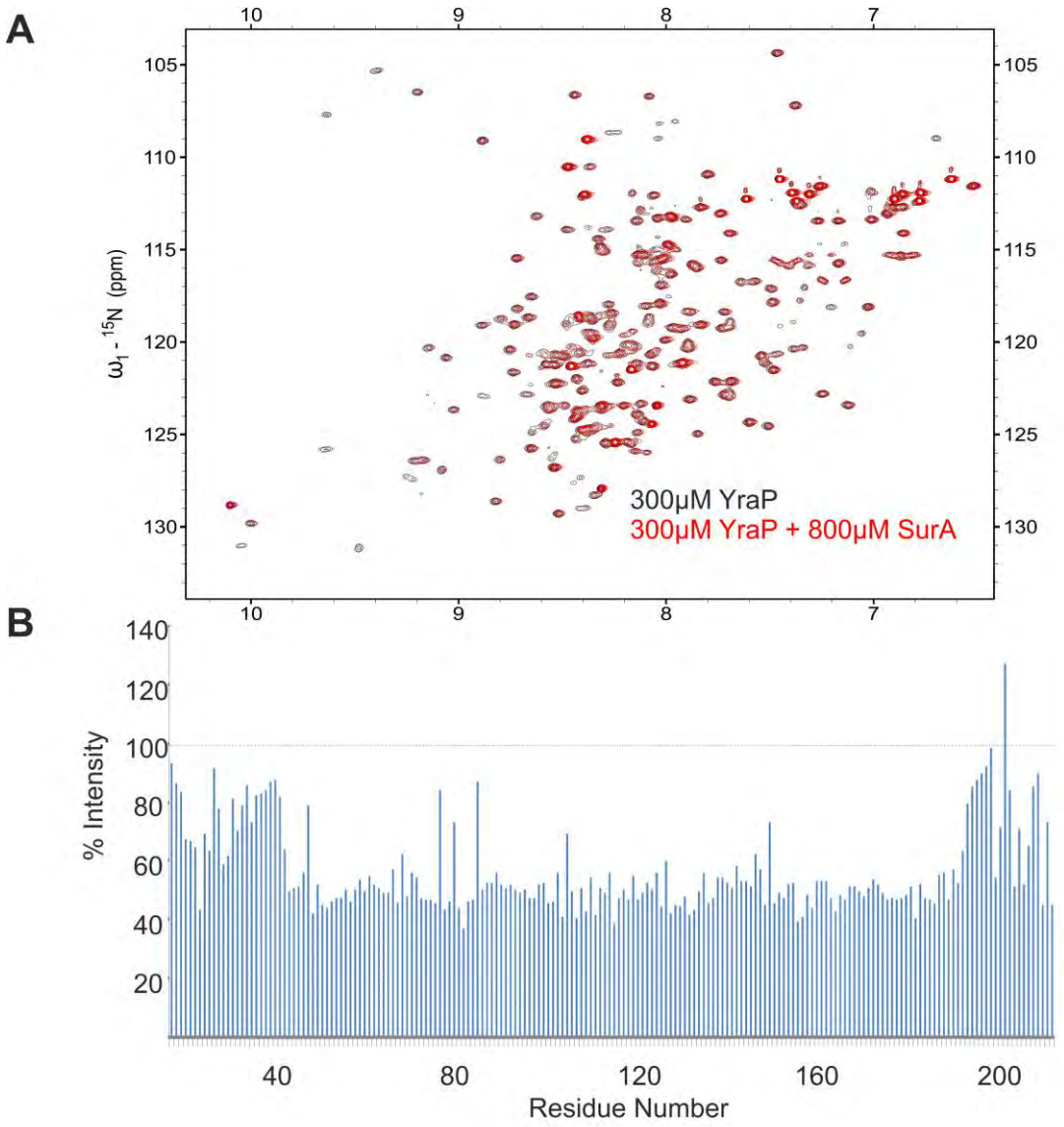


Figure 6.9 SurA does not interact with YraP by NMR A Overlaid HSQC spectrum of 350μM YraP (black) with 800μM SurA (red). The majority of the peaks have shown line broadening. **B** Histogram of the peak intensities as a percentage of YraP without SurA from the HSQC spectra. The data shows that the majority of peaks have decreased in intensity as shown in the HSQC spectra.

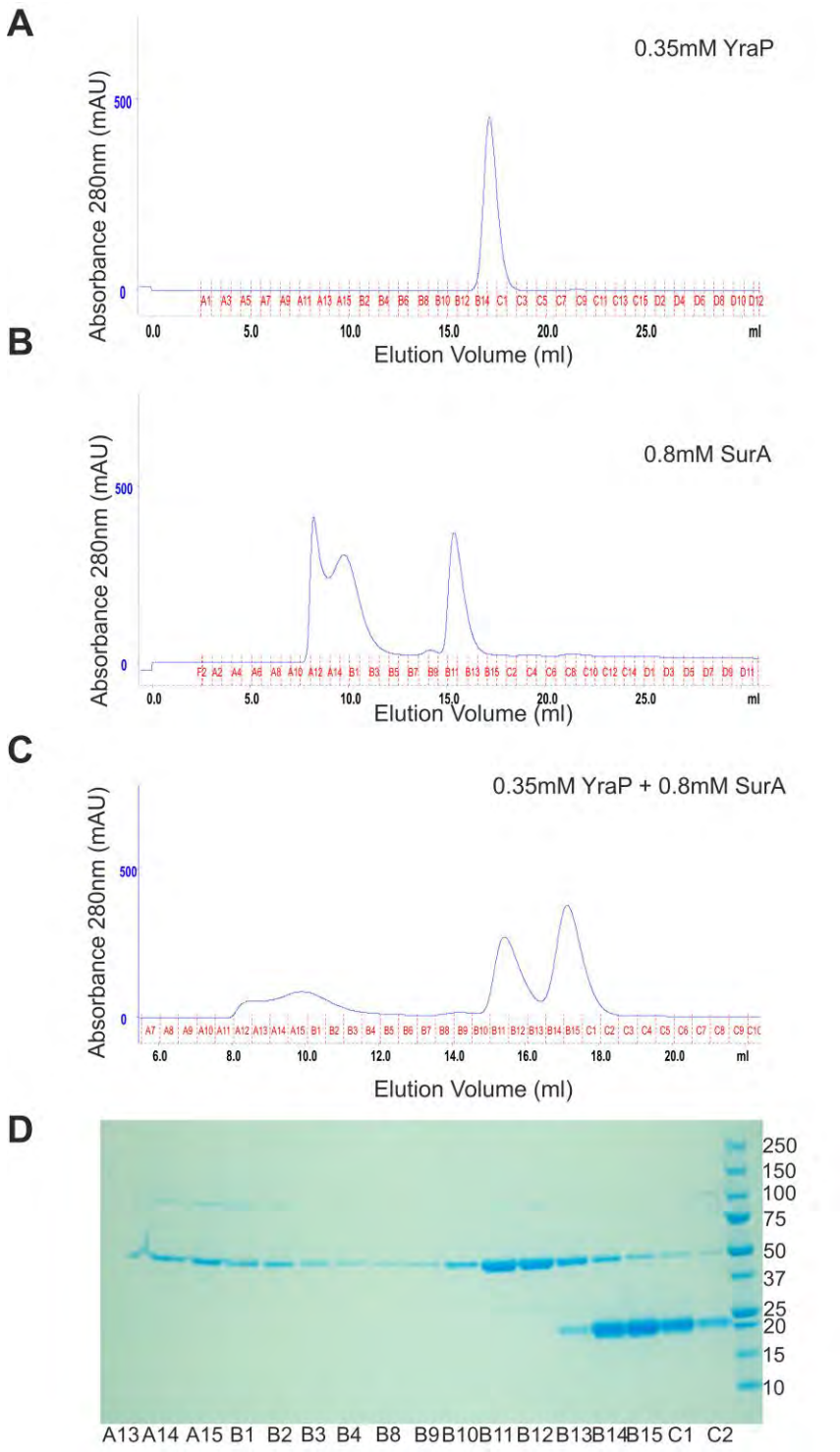


Figure 6.10 SurA does not interact with YraP as determined by gel filtration. Analytical gel filtration performed on a Superdex S200 column using **A** 0.35mM YraP, **B** 0.8mM SurA and **C** 0.35mM YraP and 0.8mM SurA. **D** Coomassie stained SDS-PAGE with samples from the gel filtration. Elution samples are indicated in the figure. The heavier species at 100 kDa is only observed in the SurA samples. No YraP is present in the heavier species.

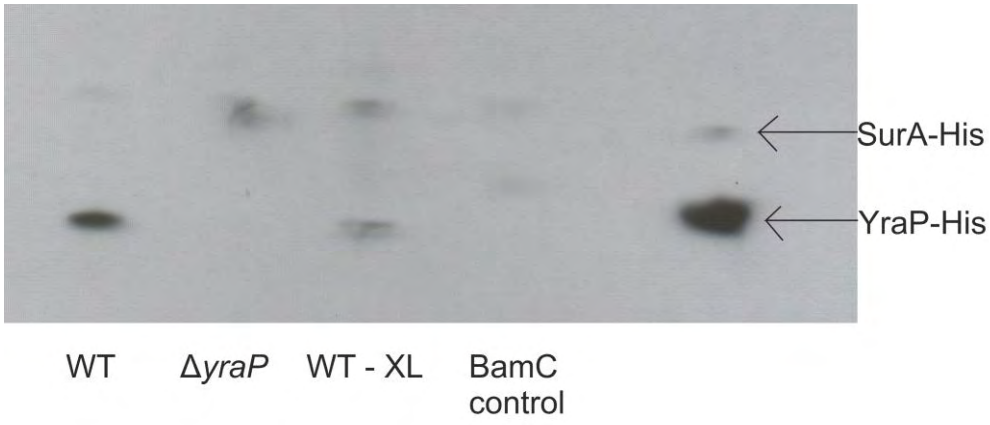


Figure 6.11 YraP and SurA do not interact from Co-IP studies. Western blot of whole cells Co-IPs with WT, Δ*yraP*, WT crosslinked (WT-XL) and BamC, probed with anti-YraP and anti-SurA antibodies. 20 µl of each sample is loaded onto the gel with positive controls are shown on the right, containing 5 µl of 300 µM YraP-His and SurA-His. The blot shows non-specific bands shown that correspond to antibody chains used in the Co-IPs. Only YraP is seen in the WT and WT-XL samples. No SurA is seen in any of the samples.

6.2.9 YraP binds specifically to the OM lipid PG

The structure of YraP has shown the presence of two distinct BON domains (see Chapter 5). BON domains are predicted to bind to phospholipids (Yeats and Bateman, 2003). Although the structure of a BON domain was recently solved (Teriete *et al*., 2010), no actual study with phospholipid binding has been performed. We wanted to examine if YraP is able to bind to phospholipids by virtue of its BON domains, as well as determining if YraP preferentially binds to different phospholipids, such as those found in the OM rather than those in the IM.

Titrations were performed with phosphatidylglycerol (PG), which comprises approximately 20-25% of the total OM phospholipid content. Using HSQC

titration experiments, separate spectra were of ^{15}N-labelled YraP in the presence of 0, 8, 16, 24, 32 and 40 mM PG. By monitoring chemical shift perturbations and the changes in intensity of the assigned peaks in the spectra, possible PG binding residues can be identified (**Figure 6.12**).

Analysis of the spectra from the titration clearly shows interaction of YraP with PG as both significant chemical shift perturbations and in some cases line broadening of residues occurs. These types of chemical shift perturbations indicate a combination of fast and intermediate exchange. By using $\Delta\delta > 50$ as a limit (approximately 1.2 standard deviations) on the data, these residues have been identified as those able to bind to PG, W127, I128, K131, R133, Q135, L137, S144, S178 and V180. W127, I128, K131, R133, Q135, and L137 map to a single region on the YraP structure; they are found to be present in the first helix of BON domain 2 (α3) (**Figure 6.13A and B**). With the number of residues that have been identified to a particular region and the size of chemical shift observed, strongly suggests that this helix binds to PG. It should be noted that S144, although not present in this helix, is very close to it in a flexible loop region. It is most likely that the chemical shifts observed for this residue are the result of structural re-arrangement on binding which is likely to occur on binding PG, however direct interaction with PG cannot be ruled out. S178 and V180 are present at the end of α4.

We wanted to repeat this experiment with phosphatidylethanolamine (PE) and cardiolipin, which are the two other major species of lipid present in the OM.

However, we were unable to solubilise these lipids at the concentrations required for NMR analysis and so the titrations could not be performed.

Finally, we wanted to see if YraP is able to bind to phosphatidylcholine (PC), which was used as a phospholipid control. Following the same titration parameters used for PG, we observed no chemical shift perturbations, suggesting that YraP may only interact with OM phospholipids (**Figure 6.13C**).

The titration data demonstrates that YraP is able to bind to PG, a major component of the OM. As Δ*yraP* cells have an OM permeability defect, this binding data could therefore possibly hint at a functional role for YraP in phospholipid maintenance in the OM.

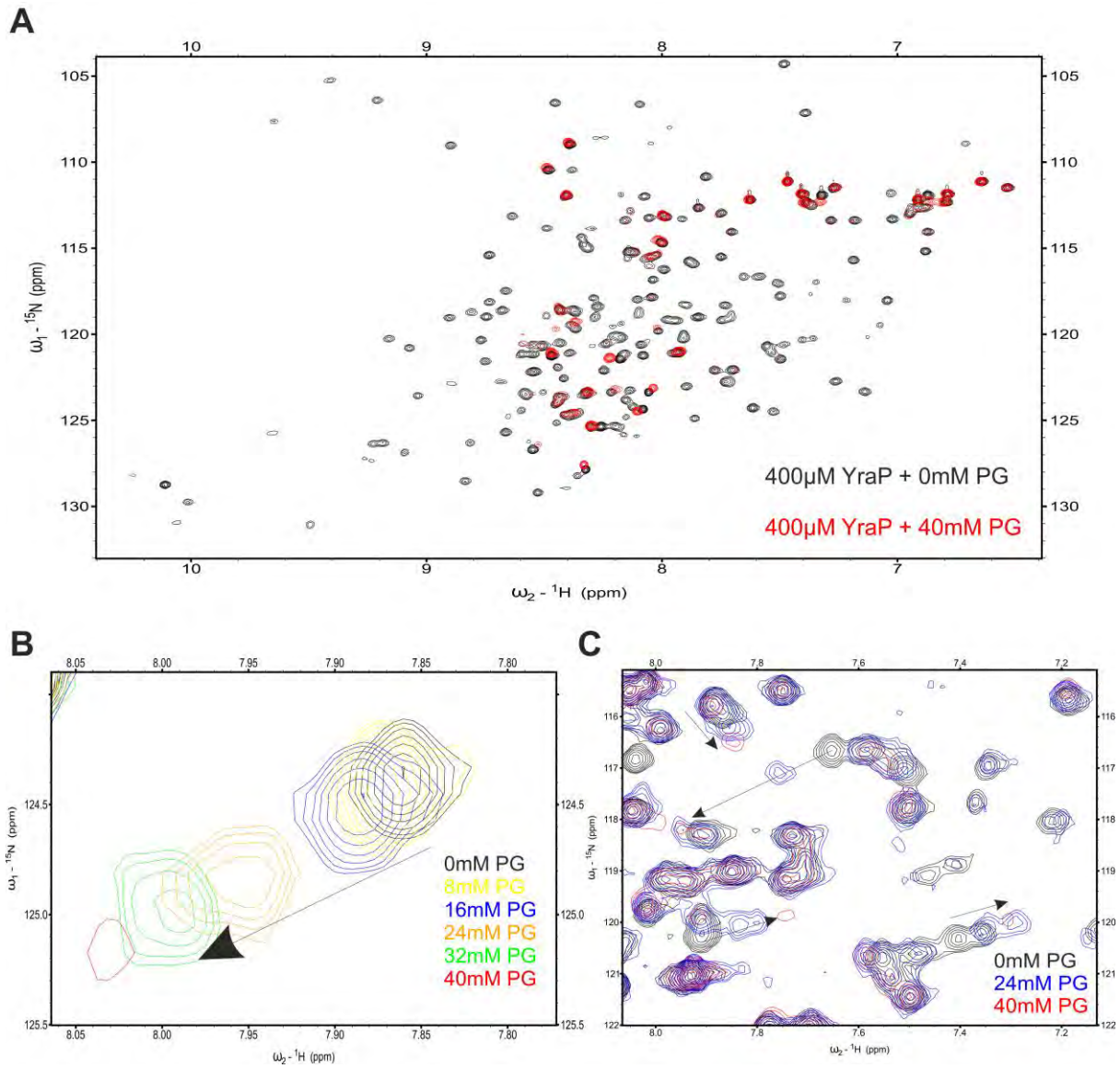


Figure 6.12 YraP binds to PG A Overlaid HSQC spectra of 400 µM YraP (black) and 400 µM YraP with 40 mM PG (red) in 50 mM sodium phosphate, 50 mM NaCl, pH 6.0 . Chemical shift perturbations and line broadening can be observed for some peaks. **B** Overlaid HSQC spectrum of residue W127, with the different concentrations of PG used in the titration. Chemical shift perturbation as well as line broadening can be observed from the starting point of 0 mM PG (black) to the final titration point of 40 mM PG (red). Arrow indicates the direction of movement with increasing PG concentration. **C** Overlaid HSQC spectrum of YraP (black), with 24 mM PG (blue) and 40 mM PG (red). The effect of PG of has caused some peaks to have major chemical shift perturbations and line broadening, which are highlighted by arrows on the spectrum, whereas other peaks have remained in the same position in the presence of PG.

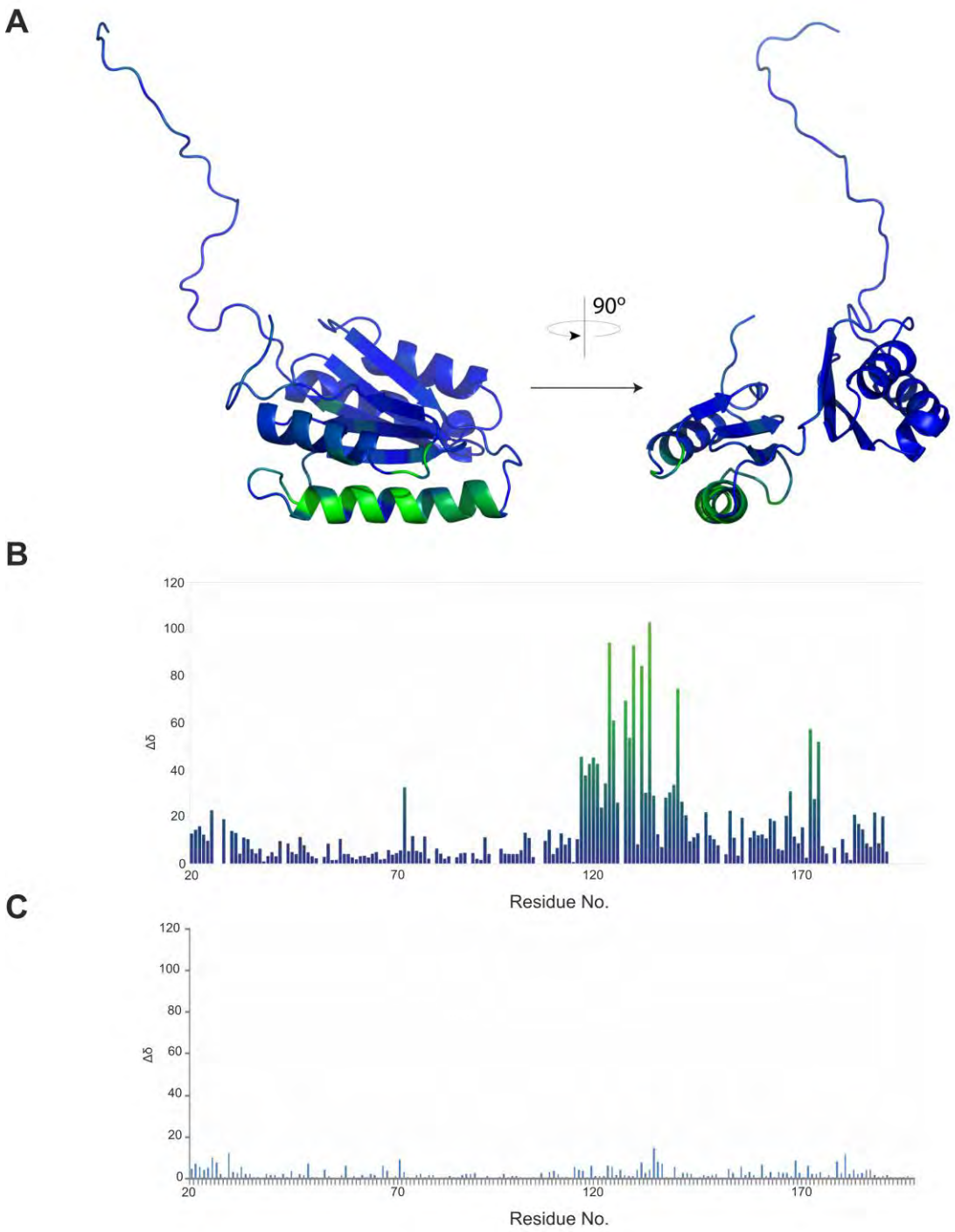


Figure 6.13 PG binding located on helix α3 in YraP A Cartoon model of YraP coloured according to the degree of chemical shift perturbation on PG binding. Colouring according to **B.** Green areas correspond to strongly PG binding areas. **B** Histogram showing the normalised chemical shift perturbations on binding 40mM PG. Normalisation of chemical shift perturbations. PG binding residues are coloured from blue (Δδ = 0ppm) to green (Δδ = 50 ppm) **C** Histogram of the chemcial shift pertubation data from the 40 mM PC titration. The same scale is used and negligible differences are observed.

6.2.10 Mutagenesis of YraP

The structure of YraP reveals the presence of 2 BON domains, which agrees with the domain predictions. From the structure it seems that the BON domains are very similar, with 2 alpha helices and 3 beta sheets each, as well as the sequences of the two domains having 69% homology to each other. We wanted to examine if YraP could function with only one of the BON domains present.

Truncated constructs of YraP lacking the second BON domain were made in the complementing vector pET20b-YraP, by replacing individual codons 112 - 118, in the loop region between the two BON domains, with two stop codons each, to prevent stop codon suppression (see **Table 6.1** for names and description of constructs).

To further the analysis, we also wanted to examine if the presence of the PG binding alpha helix was the essential part of YraP for function. Constructs were made in a similar manner as before, with residues 142 - 146, present in a loop region after the PG binding helix, being substituted with two stop codons each (see **Figure 6.14** for description).

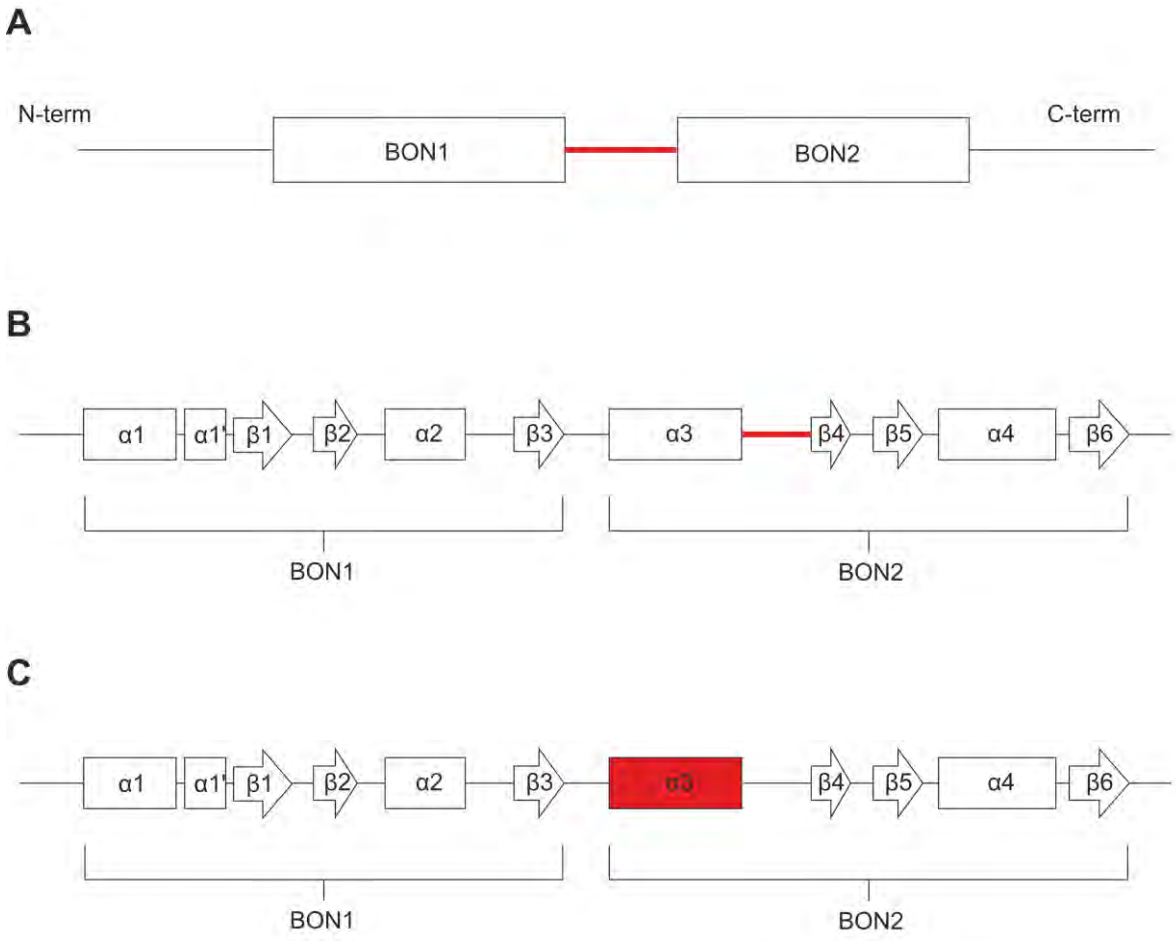


Figure 6.14 Schematics of mutational analysis of YraP. A Mutations to remove BON domain 2. The red line indicates where two stop codons were introduced to replace residues 112-118. **B** Mutations to determine if the presence helix α3 is needed for function. The red line indicates where two stop codons were introduced to replace residues 142-146. **C** Mutations to determine if the PG binding residues in α3 were essential for function. Mutants that were produced in this region are described in **Table 6.1**

As YraP is predicted to be a lipoprotein, we also mutated the N-terminal cysteine (residue 19), that becomes acylated during lipoprotein biogenesis, to an alanine. The effect of this mutation should cause mislocalisation of the protein, as it would not be able to be acylated and inserted into the OM (Narita *et al*., 2004).

These constructs were transformed into *ΔyraP* strains and a plating complementation assay was performed on 0.5% SDS with 0.5 mM EDTA. The plating assay showed that none of these constructs were able to complement the defective phenotype. To ascertain whether this result was due to YraP becoming unstable, or whether it was due to loss of a particular function, western blot analysis using anti-YraP antibody, of whole cell overnight cultures of these transformants was performed (**Figure 6.17**). The western blot showed that apart from C19A which had a trace of detectable YraP, no YraP was being produced from the other constructs, indicating that the proteins were unstable and hence complementation could not be achieved, and so maintaining the *ΔyraP* phenotype.

To further characterise YraP, we investigated whether there were any critical residues important for function. The PG binding data had identified 7 potentially important residues, present in the helix α3 (W127, I128, K131, R133, Q135, L137 and S144), that were candidates for mutagenesis. For each of these amino acids, single point mutations to alanine were performed.

From previous alignments (Teriete *et al*., 2010), the presence of a conserved glycine residue in BON domains was identified. Aligning the consensus BON domain to the individual BON domains in YraP, identified G83 and G160, as potential conserved residues. Constructs were produced, substituting each of the glycines into valines and a further construct was made which had both of the glycines substituted to valine residues.

From the NMR structure, Y75 was identified as a possible residue that might mediate the interaction between the 2 BON domains. This could be an important residue, with the role of keeping both BON domains orientated in this manner and maintaining the overall fold of YraP. To see if this was the case, Y75 was also mutated to an alanine.

These constructs were tested with the plating complementation assay. Y75A, was able to complement the OM permeability defect, strongly suggesting that Y75 does not have a major role in YraP function. The individual alanine mutations in the PG binding helix all showed growth on the test plate, indicating that they also complemented the defective phenotype. G83V and G160V also grew on the test plate, however, the double G83V/G160V construct did not complement. To analyse this result, SDS-PAGE and Western blotting was performed on overnight cultures of all the transformants.

The western blot showed that YraP was produced at similar levels to the leaky pET20b-*yraP* plasmid for the PG binding mutants, Y75A and G83V. However, G160V produced YraP at very low levels and YraP was absent in the double G83V + G160V mutant. SDS-PAGE gel indicates that this was not the effect of a loading artefact, as all lanes had similar levels of protein in them. This result indicates that only low levels of YraP are needed to complement on the plating assay, as shown by G160V and that the removal of the conserved glycine residues in the BON domains causes structural instability, suggesting the conserved nature of these residues is important for fold rather than function.

We also wanted to see if the whole PG binding helix was essential for function. As the individual alanine mutations of the residues shown to bind PG all complemented and produced stable proteins, we produced constructs that had multiple mutations in these residues. Constructs were produced that had all of these residues, or only 3 residues at each end of the helix, substituted to alanines or oppositely charged residues. Not all of the constructs were able to be generated and tested (**Table 6.1**). Out of all of the constructs generated, only the R133-L137 Ala and All Ala constructs were able to complement. W127-K131 Charged, R133-L137 Charged and All Charged constructs were unable to complement.

Western blot analysis of these constructs showed that W127-K131 Charged and All Charged constructs were unstable. However, constructs R133-L137 Ala, R133-L137-Charged and All Ala, were able to produce stable YraP. As the R133-L137 Charged mutant is able to produce YraP and is unable to complement the defective phenotype, it strongly suggests that the PG binding helix is essential for function.

Table 6.1 List of mutants of YraP produced in this study

Name	Comments
C19A	Substitution of C19 to alanine
Q113*	Substitution of Q113 and G114 to 2 stop codons
G114*	Substitution of G114 and Q115 to 2 stop codons
Q115*	Substitution of Q115 and P116 to 2 stop codons
P116*	Substitution of P116 and I117 to 2 stop codons
L117*	Substitution of I117 and G118 to 2 stop codons
G118*	Substitution of G118 and L119 to 2 stop codons
L119*	Substitution of L119 and G120 to 2 stop codons
V142*	Substitution of V142 and K143 to 2 stop codons
K143*	Substitution of K143 and S144 to 2 stop codons
S144*	Substitution of S144 and S145 to 2 stop codons
S145*	Substitution of S145 and N146 to 2 stop codons
N146*	Substitution of N146 and V147 to 2 stop codons
W127A	Substitution of W127 to alanine
I128A	Substitution of I128 to alanine
K131A	Substitution of K131 to alanine
Q135A	Substitution of Q135 to alanine
L137A	Substitution of L137 to alanine
S144A	Substitution of S144 to alanine
G83V	Substitution of G83 to valine
G160V	Substitution of G160 to valine
G83V + G160V	Substitution of G83 and G160 to valines
Y75A	Substitution of Y75 to alanine
W127-K131 Charged	Substitutions of W127Q, I128E, K131E, S144A
R133-L137 Ala	Substitutions of R133A, Q135A, L137A, S144A
R133-L137 Charged	Substitutions of R133E, Q135E, L137D, S144A
All Ala	YraP with W127A, I128A, K131A, R133A, Q135A and L137A
All Charged	YraP with W127Q, I128E, K131E, R133E, Q135E and L137D

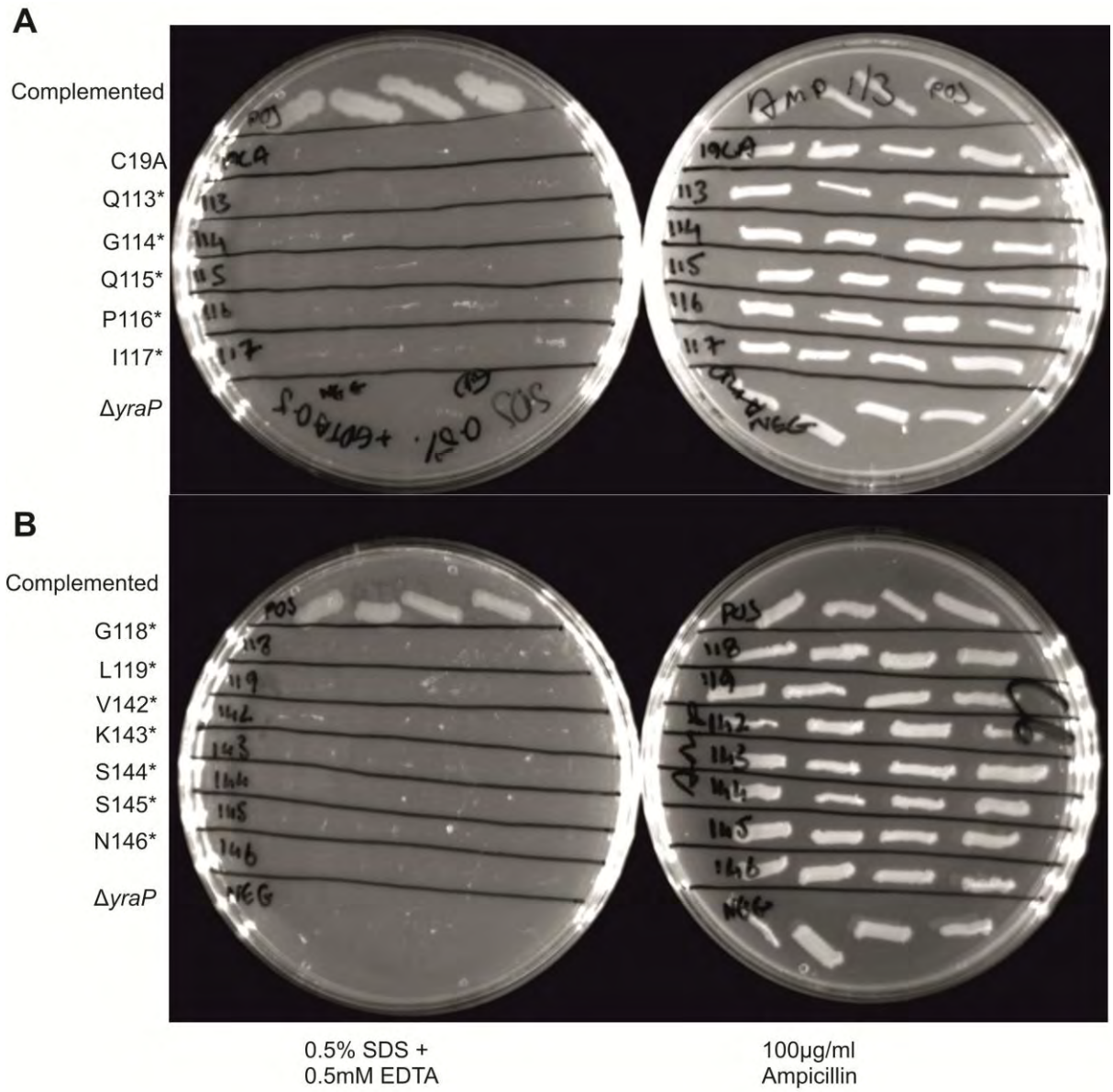


Figure 6.15 Complementation assay performed on 0.5% SDS + 0.5 mM EDTA plates and 100 µg/ml ampicillin used as a positive control. **A** Mutants of C19A and some of the ΔBON2 mutants. **B** The rest of the ΔBON2 mutants and Δα3 mutants. The presence of stop codons at these positions have abolished YraP function.

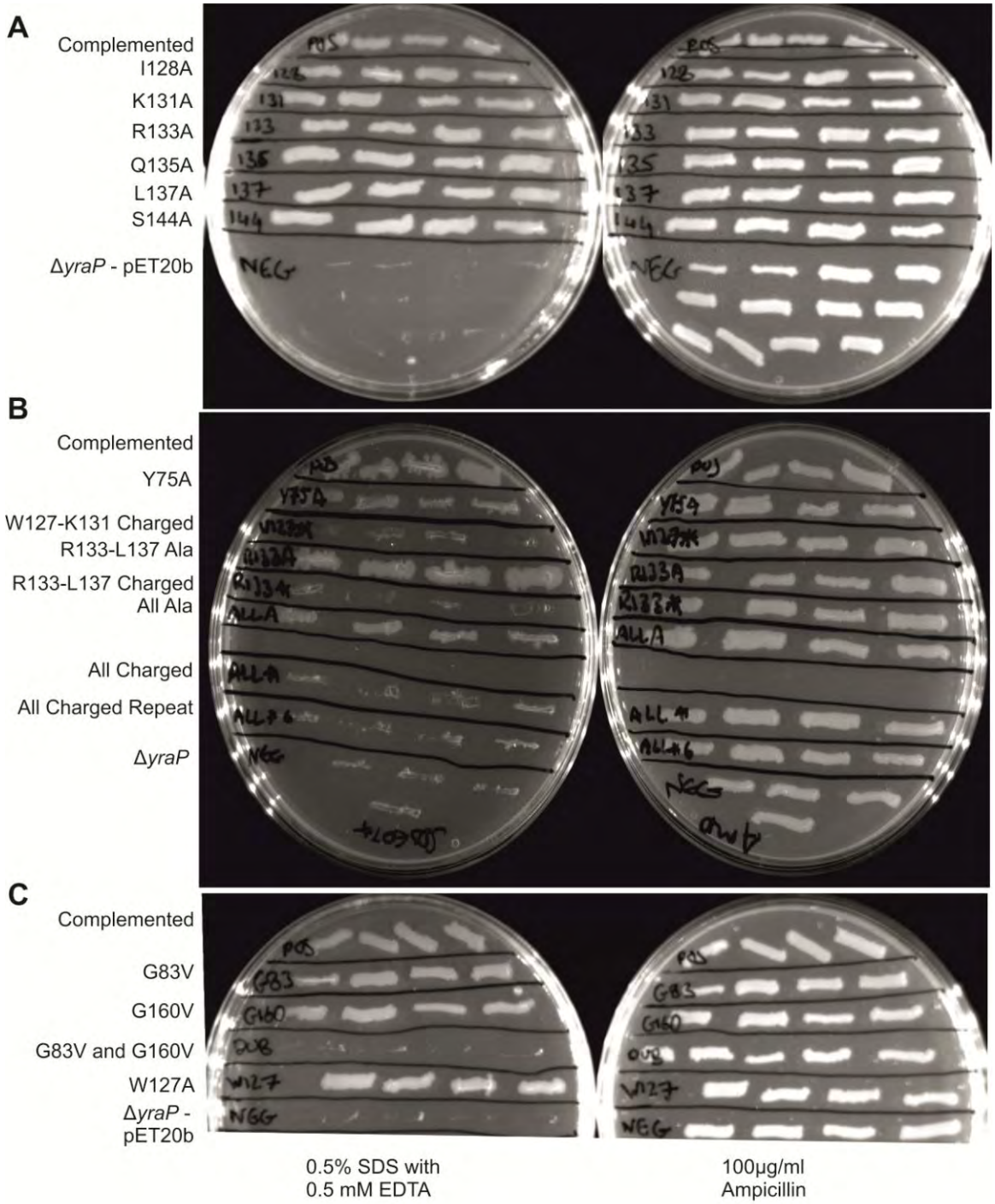


Figure 6.16 Plating assay performed with mutants generated onto 0.5% SDS + 0.5 mM EDTA plates and 100 µg/ml ampicillin, as a positive control, LB-Agar plates. Mutants of pET20b-*yraP* were transformed into Δ*yraP*, tested on those conditions and growth was observed. Each plate has a positive control Δ*yraP* with pET20b-*yraP* and negative control Δ*yraP* with pET20b. **A** Plates with single alanine mutations in α3 of YraP. **B** Plates with Y75A and multiple alanine and charged mutants in α3. **C** Plates testing the conserved glycine and W127A mutants.

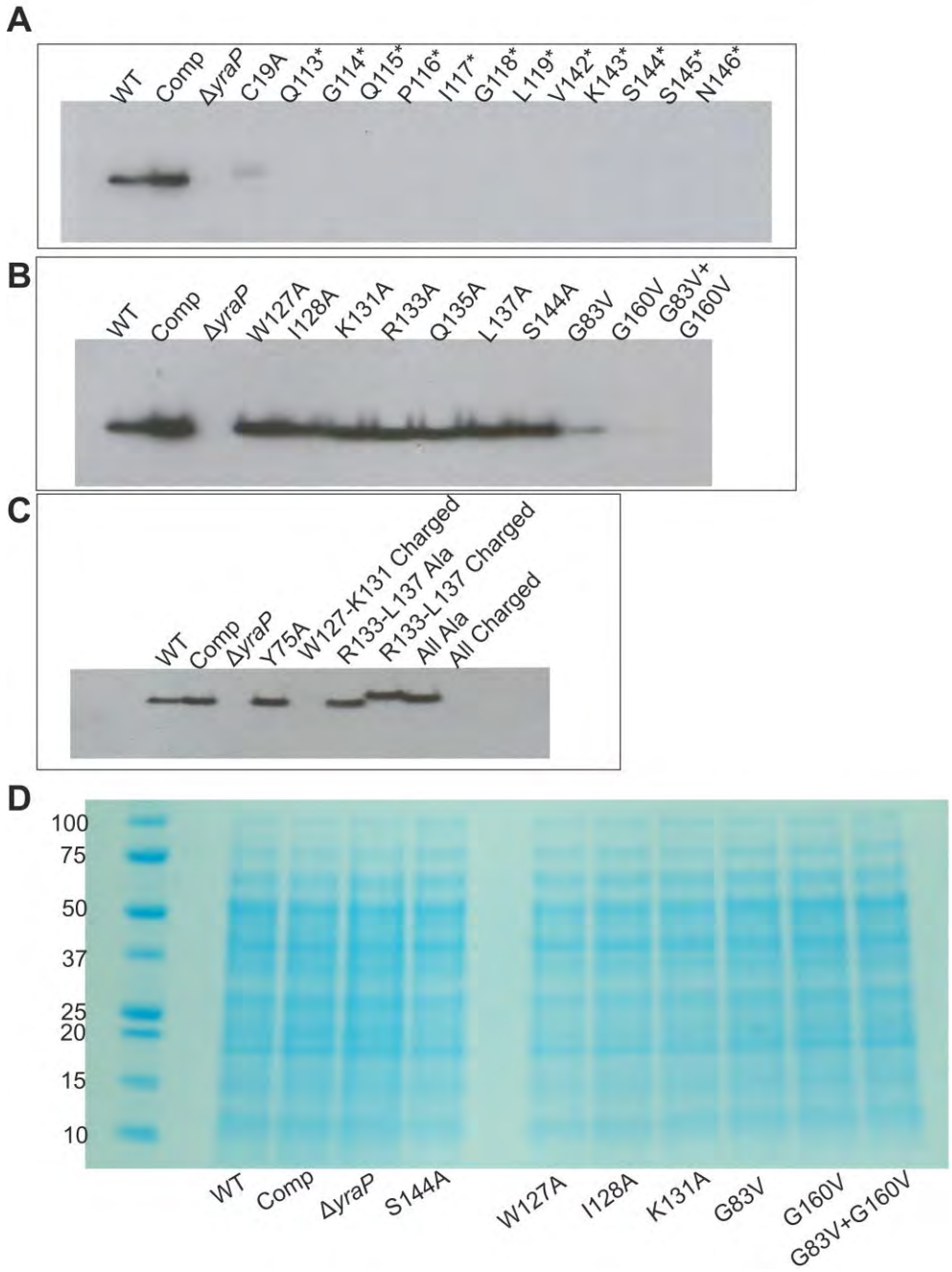


Figure 6.17 Mutants of the conserved glycines of YraP produce unstable protein
A Western blots of normalised cell lysates of the mutants, to indicate if stable YraP is produced from the constructs. 20 µl of each lysate was loaded and anti-YraP antibody (from rabbit) was used. The mutants in this blot are C19A, the stop codon insertions mutants to remove BON2 and stop codon insertions to keep α3. **B** This blot contains mutants from individual alanine substitutions from PG binding residues of α3 and the conserved glycines present in each BON domain. **C** This blot contains mutants Y75A and multiple or charged alanine substitutions in the PG binding residues of α3. **D**

Coomassie stained SDS-PAGE of 20 µl of cell lysates to indicate that there are no loading discrepancies.

To find additional areas of YraP which are crucial for its function, a transposon mutagenesis kit was used (D. Browning, C. Wardius). Using MuA transposase kit, an artificial transposon (Entrancesposon) can be inserted into the coding region of the gene. This results in the generation of a 15 bp in-frame insertion into the target gene, which corresponds to an extra 5 translated amino acids. The function of these mutants could then be determined using the complementation plating assay.

The transposon mutagenesis covered a wide range of YraP. Although unfinished, mutants were generated in the flexible N-terminus, secondary structure elements, PG binding helix and flexible loop regions of YraP (**Table 6.2**).

Screening of these mutants, non-functional YraP mutants were present at the N-terminal flexible region and at the start of the α1 helix (**Figure 6.18**). Interestingly, transposon insertion at the α3 helix (residues D125, W127, S134 and Q135) did not affect the ability to grow on the test plates. There is the possibility that PG binding is dependent on the helix rather than the residues that comprise it and that transposon insertions in this region do not disrupt the helix.

Table 6.2 Mutants produced from the transposon screen. Mutants highlighted in bold indicate transposon insertions that abolish YraP function.

Position	Location
M1	Lipobox
S5	Lipobox
A13	Lipobox
L16	**Lipobox**
Q17	**Lipobox**
A22	N-term flexible
A29	N-term flexible
A35	**N-term flexible**
S40	**N-term flexible**
G42	**N-term flexible**
L50	**α1**
R53	α1
S56	α1
V72	β1
Q76	Flexible
A88	α2
D102	Flexible
N105	Flexible
N109	β4
Q113	Flexible
D125	α3
W127	α3
S134	α3
Q135	α3
V142	Flexible
M159	Flexible
E166	α4
K169	α4

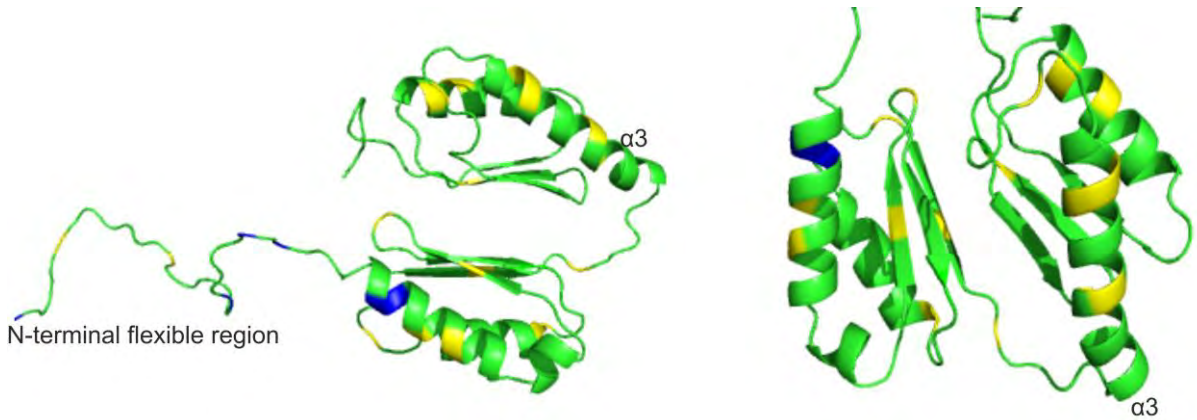


Figure 6.18 The N-terminal region of YraP is important for function. Mapping of the transposon mutants onto the structure of YraP. Yellow regions indicate a transposon insertion and blue regions indicate a transposon insertion which abolished YraP function.

6.2.11 Generation of suppressor mutants

Δ*yraP* cells, within 16 hours, stop growing in the presence of vancomycin, SDS and EDTA. However, allowing growth for longer time periods, allows the generation of suppressor mutations. Those cells which are able to grow on conditions that should be lethal, have acquired mutations in the genomic DNA which restore the OM barrier function and therefore allow them to grow. Analysing these mutations can yield the identification of genes that are involved in coping with the test conditions as well as elucidating the interplay between the genetic pathways that are involved in OM homeostasis.

Two sets of suppressor mutants were generated from plates containing 250 µg/ml vancomycin and 0.5% SDS with 0.5 mM EDTA LB-Agar. Any growth that was observed on these test conditions, were streaked again to single colonies onto the same conditions to ensure that they were suppressor mutants. A final set of streaks to single colonies were performed onto the test conditions as well as a streak onto kanamycin (100 µg/ml) LB-Agar plates, to ensure that the kanamycin resistance cassette, that had replaced the *yraP* gene, was still present and that the observed growth was not from contamination (**Figure 6.19**). In total, 6 suppressor mutants were generated, with each one originating from a different streak.

Genomic DNA was extracted from these strains and whole genome sequencing was performed (Dr. D. Rasko and Dr. T. Lawley, unpublished data). As of this time, only the vancomycin suppressor mutation data has been analysed. The

mutations were identified by sequence alignment between the genomic data of the mutants and BW25113.

The data has identified approximately 20 genes that have acquired mutations (**Table 6.3**). Analysing this dataset, the genes can be grouped into 2 major categories. The first category, genes that have the same mutations in all of the strains, are most likely not involved in suppression, but are mutations that have just occurred from working with the strain. The second category, which consists of mutations in genes that are present in all of the strains but the mutations are in different locations and mutations in genes that are not present in all of the strains, are most likely involved in phenotype suppression.

Although not all of the data has been analysed, some genes have been identified that are likely candidates for suppression. This list consists of *nanT* (a putative sialic acid transporter), *vacJ* (a putative OM lipoprotein), *pgm* (phosphoglucomutase), *ychF* (a GTP and nucleic acid binding protein), and *envZ* (an osmolarity sensing histidine kinase) amongst several other genes. Ideally, further analysis of this data, alongside experimental data, is needed to justify any hypotheses formed from this data.

Table 6.3 List of genes that harboured mutations that suppressed the Δ*yraP* sensitivity to vancomycin

Gene Name/Annotation	**Possible function**
nanT	Sialic acid transporter (permease)
vacJ (*mlaA*)	Phospholipid transport protein
pgm	Phosphoglucomutase
ychF	GTP binding protein
envZ	Sensory histidine kinase
feoB	Ferrous iron transport protein B
fliQ	Flaggellar biosynthesis protein
ybdO	Putative LysR-family trancriptional regulator
hycE	Formate hydrogenlyase subunit 5
FIG00638289	Hypothetical protein
hycC	Formate hydrogenlyase subunit 3
FaeE	Beta fimbriae chaperone protein
FIG00639456	Hypothetical protein
mdoH	Glucans biosynthesis glucosyltransferase H (EC 2.4.1.-)
betT	High-affinity choline uptake protein
yddA	Hypothetical ABC transporter ATP-binding protein
sucP	Putative sucrose phosphorylase (EC 2.4.1.7)
cybB	Cytochrome b(561)
mocA	CTP:molybdopterin cytidylyltransferase
eco	Proteinase inhibitor I11, ecotin precursor
yadM	Fimbrial protein
FIG00639374	Hypothetical protein
rpoD	RNA polymerase sigma factor

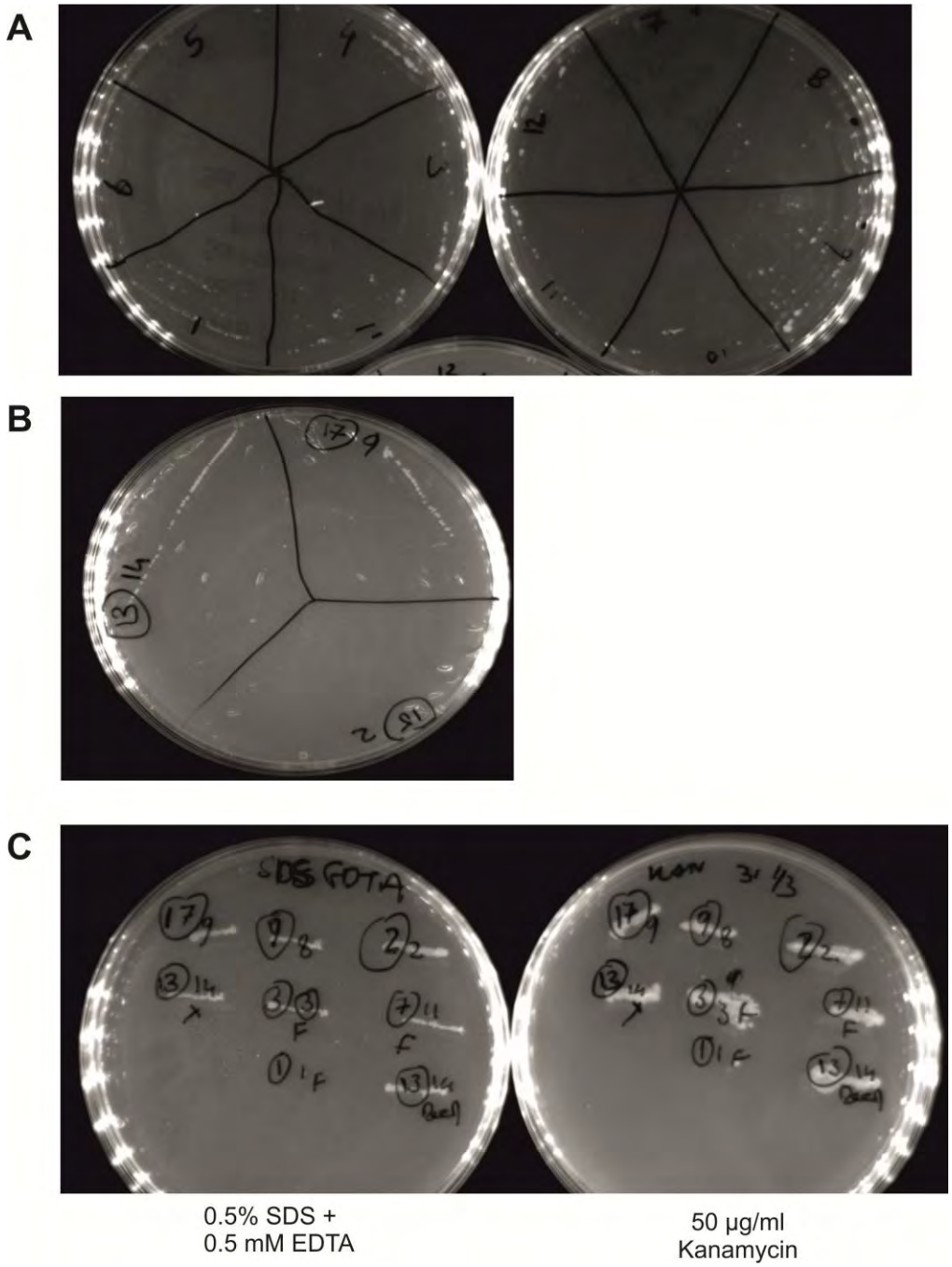


Figure 6.19 Δ*yraP* strains are able to generate suppressor mutations and be selected for 0.5% SDS and 0.5mM EDTA resistance. A Δ*yraP* cells are streaked onto 0.5% SDS with 0.5mM EDTA plates. With enough time, growth can be observed and individual colonies can be picked. **B** Further selection of potential suppressor mutants onto another 0.5% SDS with 0.5mM EDTA plate. Growth is observed with some of the samples and individual colonies can be observed on this plate. **C** Final selection of suppressor mutants. Streaks are performed on 0.5% SDS with 0.5mM EDTA, to confirm resistance and are also streaked onto 100 µg/ml kanamycin plates, to ensure that the kanamycin resistance cassette is still present, indicating that these are still likely to be Δ*yraP* strains.

6.2.12 *S. enterica* Δ*yraP* cells have a similar phenotype to BW25113 Δ*yraP* cells

Knockouts of *yraP* show an outer membrane permeability defect phenotype in *E. coli*. We wanted to examine if a similar phenotype would be observed in the closely related species, *Salmonella entercia*. This approach would give an indication if *yraP* functions in a similar manner in *Salmonella* as it does in *E. coli*. If this was the case, then it would be feasible that *E. coli yraP* might be able to complement the similar defects that would be observed with Δ*yraP Salmonella* strains. From a sequence alignment between the proteins, YraP from Salmonella and *E. coli* share 96% identity (**Figure 6.20A**).

ΔyraP:*kan* strains were generated in *Salmonella enterica serovar* Typhimurium strain SL1344, by Dr. D. Squire. This strain was made competent and transformed with pET20b and pET20b-*yraP* and subjected to the previously described plate screening assay on 4% SDS and 0.5% SDS with 0.5 mM EDTA, with BW25113, Δ*yraP* and complemented *ΔyraP* cells.

The assay shows that *E. coli* and *Salmonella* wildtype and Δ*yraP* strains behave in a similar manner (**Figure 6.21**). Both wildtype strains are able to grow on the test conditions whereas the *ΔyraP* strains are unable to do so. The effect of pET20b-*yraP* shows only partial complementation. This plasmid is able to fully complement the *ΔyraP E. coli* strain, with growth on both test conditions, but in the case of *ΔyraP Salmonella* strains, growth is only observed on 4% SDS and not on 0.5% SDS with 0.5 mM EDTA. Although there is a great sequence identity between *Salmonella* and *E. coli*, the data suggests that *E. coli* YraP is only able

to complement some of the defects in a *Salmonella ΔyraP* strain. By mapping the where the differences in the multiple sequence alignment onto the structure of YraP (**Figure 6.20B**). Residue differences are seen in flexible loop regions as well as secondary structure elements such as α1, α2, α3, α4 and β5. These regions could have important roles in YraP structure and these differences may prevent *E. coli* YraP from complementing the defective phenotype in *S. enterica*.

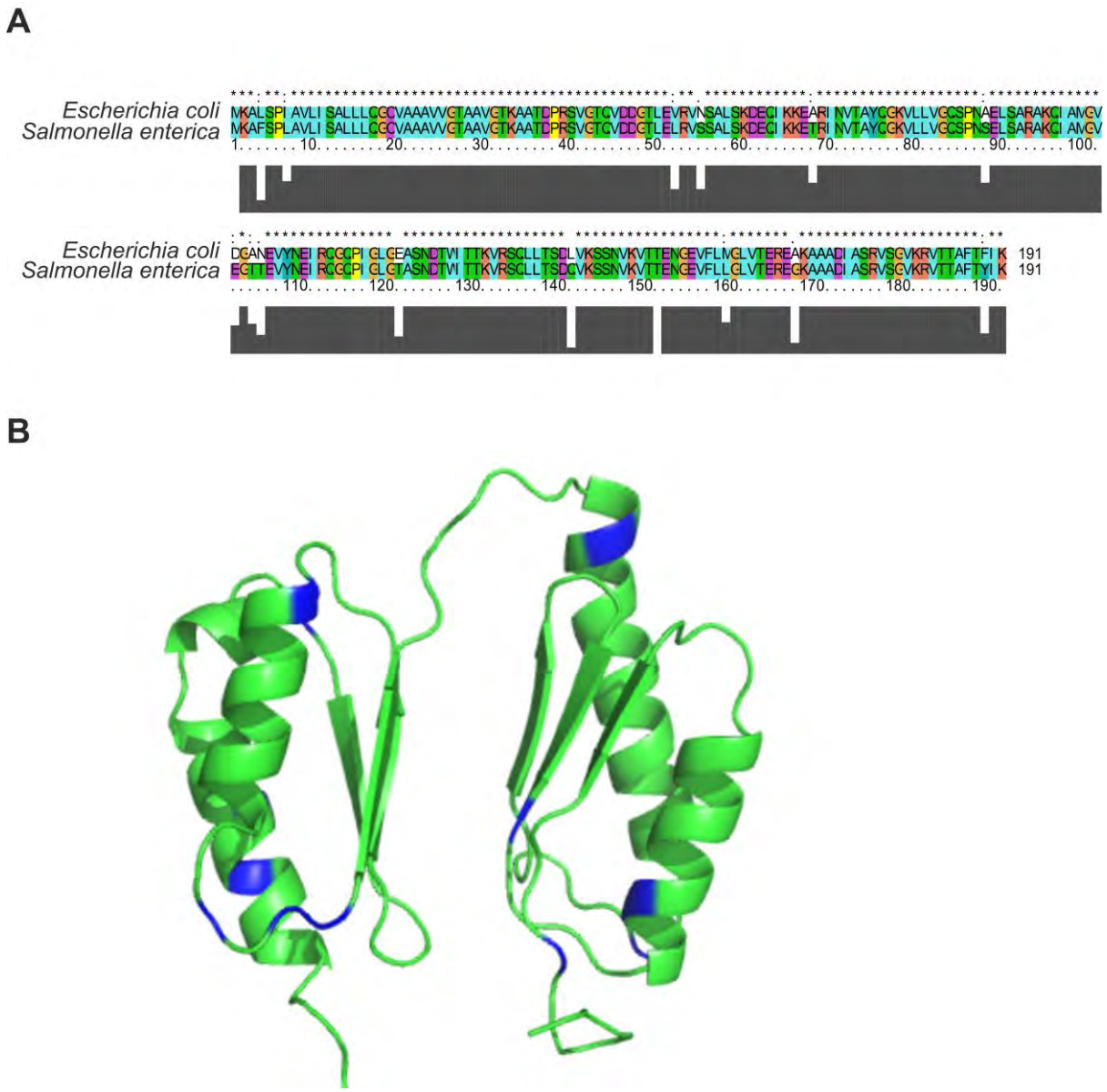


Figure 6.20 *Salmonella enterica* Δ*yraP* cells display a simlar phenotype to BW25113 Δ*yraP* cells.

A Sequence alignment between *E. coli* and *Salmonella enterica* YraP proteins determined using ClustalX. The alignment shows that YraP between the species are heavily conserved, with a calculated 96% identity between the two proteins. **B** Structure of YraP from *E. coli* which residues that are different to YraP from *S. enterica* highlighted in blue.

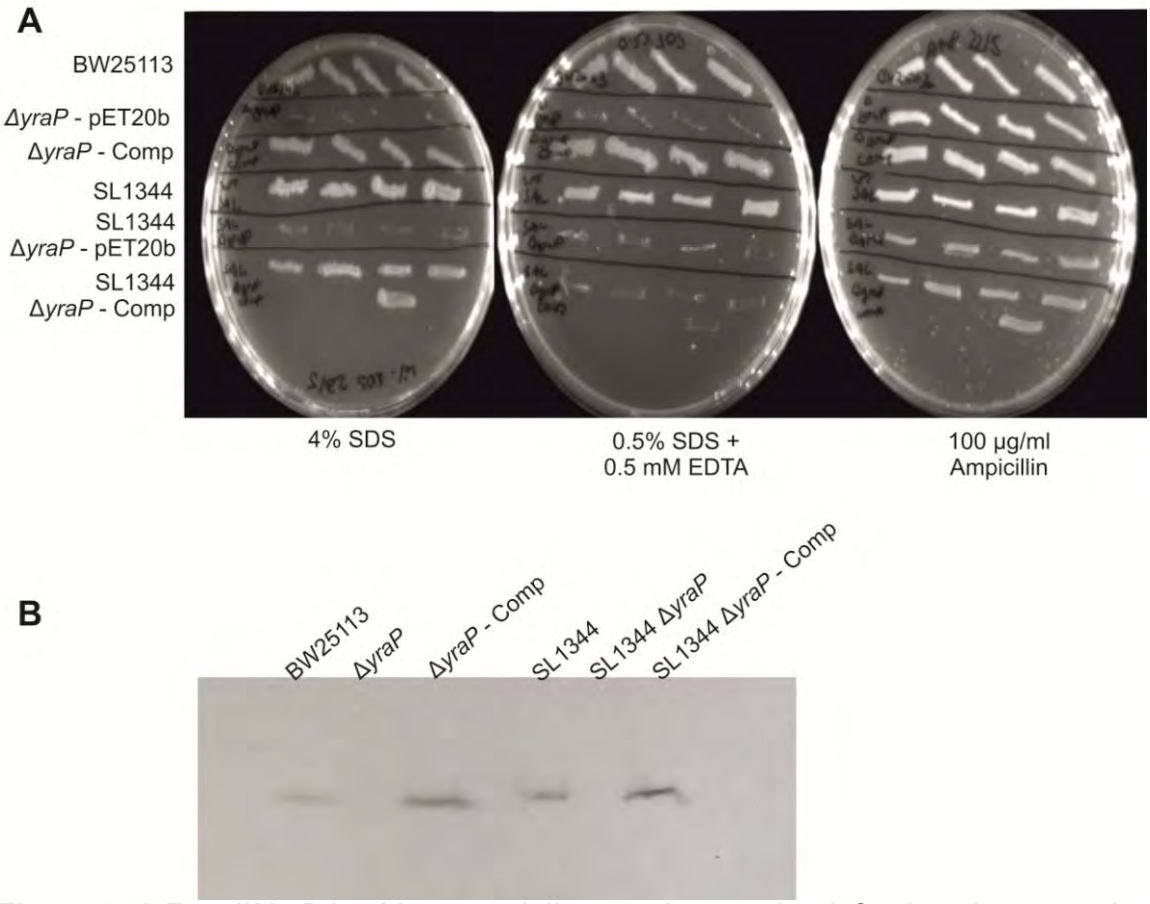


Figure 6.21 *E. coli* YraP is able to partially complement the defective phenotype in Δ*yraP S. enterica* strains. A Plating assay with *E. coli* Δ*yraP* and SL1344 Δ*yraP* complemented with *E. coli* pET20b-*yraP*. SL1344 Δ*yraP* strains have the same phenotype as *E. coli* Δ*yraP* cells in that they show sensitivity to both 4% SDS and 0.5% SDS with 0.5 mM EDTA plates (labelled SL1344 Δ*yraP*-pET20b and Δ*yraP*-pET20b respectively). Transforming *E. coli* pET20b-*yraP* into *S*L1344 Δ*yraP* cells, growth is observed on 4% SDS plates and not on 0.5% SDS with 0.5 mM EDTA plates, indicating that *E. coli yraP* is only able to partially complement the defect in SL1344 Δ*yraP* cells (labelled Δ*yraP*-Comp and SL1344 Δ*yraP*-comp respectively). **B** Western blot of the WT, Δ*yraP* and complemented strains of BW25113 and SL1344, probed with *E. coli* anti-YraP antibody, using 20 µl of normalised cell lysates. This antibody is able to detect SL1344 YraP, demonstrating the large homology between the two proteins. No YraP is observed in the SL1344 Δ*yraP*. However with SL1344 Δ*yraP* strain complemented with *E. coli* pET20b-YraP, *E. coli* YraP is detected, indicating that this plasmid is expressing in SL1344.

6.2.13 Δ*yraP* cells have an altered LPS profile

One aspect of the OM we hadn't previously investigated that could be responsible for the Δ*yraP* phenotype was a possible defect in LPS assembly. BW25113 is a strain that is unable to assemble O-antigen. Because of this, changes in LPS composition cannot be observed in these strains. To solve this problem, the gene *wbbL* was cloned into pET20b. *wbbL* is a gene that encodes an arabinose transferase, which allows transfer of arabinose onto the second sugar group of the growing O-antigen polymer. When pET20b-*wbbL* is transformed into BW25113 cells, leaky expression of WbbL from the vector, allows O-antigen to be synthesised, which ultimately leads to the restoration of observable LPS in BW25113 (work performed by Dr. D. F. Browning) and therefore allows this strain to be used to identify any potential LPS defect in Δ*yraP* cells.

There is a potential drawback with this approach. Expression of *wbbL* is from a leaky vector and hence cannot be regulated and so could lead to excessive LPS being formed. Thus if there is a subtle change in LPS composition with Δ*yraP* cells, it may not be observed. To address this issue, LPS composition was also examined in *Salmonella enterica serovar* Typhimurium strain, SL1344, which is able to produce O-antigen and has been previously shown to have a similar phenotype to *E. coli* with Δ*yraP* cells.

LPS was extracted from WT, Δ*bamE* and Δ*yraP* strains from BW25113, BW25113 transformed with pET20b-*wbbL* and SL1344. Silver staining shows

that BW25113 WT, Δ*bamE* and Δ*yraP* all look identical and lack O-antigen. With the presence of pET20b-*wbbL*, WT and Δ*bamE* LPS looks similar, with O-antigen produced (**Figure 6.22**). However, with Δ*yraP*, the O-antigen 'ladder' is completely absent. In SL1344, a similar result is seen. WT and Δ*bamE* have a similar LPS profile whereas Δ*yraP* has a completely different LPS profile, with O-antigen also absent. This result strongly suggests that there in an LPS defect associated with Δ*yraP* cells.

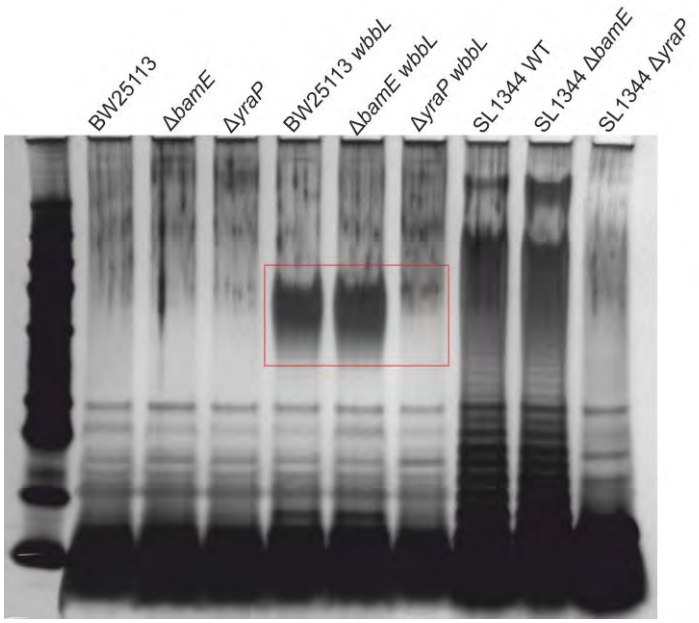


Figure 6.22 Δ*yraP* cells have an altered LPS profile. Silver stain gel of LPS prepared from WT BW25113, Δ*bamE* and Δ*yraP* cells with and without pET20b-*wbbL* and from *Salmonella enterica* SL1344 strains harbouring the same knockout genes. pET20b-*wbbL* allows the production of O-antigen, that is not present in BW25113. Δ*yraP* pET20b-*wbbL* cells display a defect in O-antigen assembly as shown by the absence of O-antigen polymers compared to BW25113 and Δ*bamE* pET20b-*wbbL* cells (indicated in the red box). The last three lanes are LPS from *Salmonella enterica serovar* Typhymirium strain SL1344, which is able to assemble O-antigen. With the Δ*yraP* SL1344, O-antigen polymerisation is effectively absent in comparison to WT and Δ*bamE* SL1344.

6.2.14 OM proteomic comparison of WT and Δ*yraP* cells

Previous work by (Sklar *et al*., 2007a) have shown that Δ*bamE* cells have reduced levels OMPs compared to WT cells. We wanted to examine if this was the case with Δ*yraP* cells, and that altered levels of protein in the OM would explain the associated defect.

In order to determine if this was the case, a mass spectrometry approach was used. Approximately 200 µg of protein was taken from OM fractions taken from sucrose density centrifugation separations from WT and Δ*yraP* cells. Peptides were generated using urea and trypsin and extracted. WT peptides and Δ*yraP* peptides were labelled differently which would distinguish them by 4 Da. Equal volumes of extracted peptides were mixed and ESI-MS was performed and peptides were identified by the use of the MASCOT database (Dr. D. Ward, Dr. N. Shimwell).

The identified peptide list was robust, in that it allowed the identification of the majority of the OM species with a 5% false discovery rate and the presence of some IM protein contaminants. The relative ratios of WT and Δ*yraP* paired peptides were determined by quantifying the peak intensity from the elution profiles of peptides. As the list of identified proteins was quite large (300+ proteins), only the most likely OM proteins were examined this way (e.g. BAM complex proteins, LptD, LptE, porins).

The data demonstrates that relative protein composition between WT and Δ*yraP* cells are similar, with approximately a 1:1 ratio observed between the samples

(**Table 6.4**). None of the major proteins that are involved in OM biogenesis, such as the BAM complex genes, have altered levels. This result suggests that *yraP* does not function in OM protein insertion.

Table 6.4 Elution profile counts of identified peptides

Protein	WT	Δ*yraP*	Ratio (Δ*yraP*:WT)
BamA	128664	130251	1.012334453
BamB	123736	125213	1.011936704
BamC	256912	263692	1.026390359
BamD	65740	67423	1.025600852
BamE	83748	87284	1.042221904
LolB	129252	138620	1.072478569
LptD	68316	74556	1.091340242
LptE	47932	52732	1.100141868
OmpA	643124	653032	1.015406049
OmpC	28692	29392	1.024397044
OmpW	24088	25846	1.072982398
TolQ	27084	28500	1.052281790

6.2.15 The OM of *Δyrap* cells have a different phospholipid composition to WT cells

Δ*yraP* cells show an OM defect. This defect is likely to be due to altered composition of the OM. The most likely candidates for an observed change are phospholipids, lipopolysaccharide or proteins, which are the three major constituents of the OM.

In order to determine possible changes in the phospholipid composition of Δ*yraP* cells, lipids were extracted from OM fractions of WT and Δ*yraP* cells and analysed using mass spectrometry, gauging the relative quantification of phospholipids between the samples (performed by Dr. J. Kirwan). Using PCA (principal component analysis), comparing WT and Δ*yraP* samples, two distinct populations are observed with tight clustering (**Figure 6.23**). This result is reproducible with a repeated set of samples showing a similar pattern. Further mass spectrometry experimentation is required to determine which species of phospholipids have been altered, further experimentation is needed, but the likely candidates are phosphatidylethanolamine (PE), phosphatidylglycerol (PG), cardiolipin and lipid A.

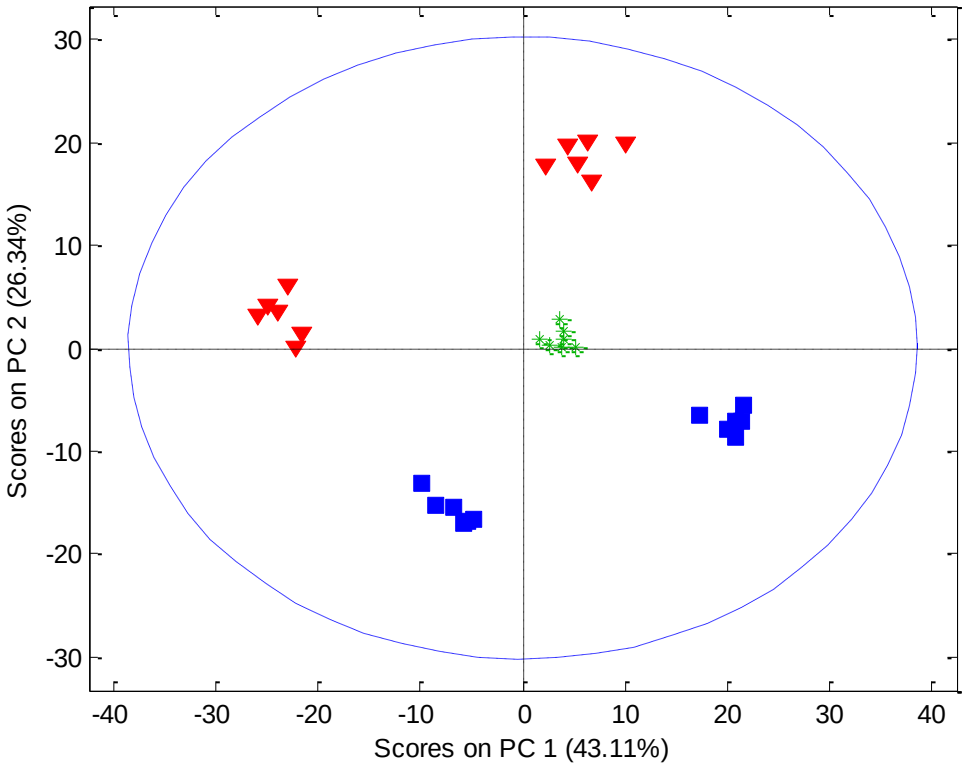


Figure 6.23 Δ*yraP* has a different OM phospholipid composition compared to WT cells. PCA scores plot of all samples [WT (blue), Δ*yraP* (red) and quality control (green) samples]. The data is originally plotted out as peaks in a 3D co-ordinate system. By observing linear trends between the samples in two dimensions (PC1 and PC2), trends in populations, similarities between groups of samples can be observed. Within a group (WT or Δ*yraP* phospholipids), there is tight clustering from the individual samples, but clear differences between groups are observed, with WT samples occupying the bottom half of PC2 and Δ*yraP* samples present in the top half of PC2. Separation across PC1 is due to the two different preparation batches, but WT samples are at the lower half of PC2 and Δ*yraP* samples at the upper half of PC2 is still observed.

6.3 Discussion

The data produced in this study strongly suggests that *yraP* is involved in OM biogenesis. Δ*yraP* cells display an OM defect by failing to grow on LB-Agar plates containing either SDS with EDTA or vancomycin. By cloning *yraP* into pET20b and using the leaky expression from this construct, complementation of the defective phenotype has successfully been achieved. This assay has been a powerful tool in the functional studies of *yraP*.

The studies in *Salmonella* have shown that the same phenotype is associated with knockouts of *yraP* in those strains. It seems likely therefore that *yraP* may function in a similar manner in closely related species. Using the pET20b-*yraP* vector, YraP from *E. coli* was only able to partially complement Δ*yraP* defect in *Salmonella*. Growth was only observed on 4% SDS plates and not on 0.5% SDS with 0.5 mM EDTA. This result suggests that the different plating conditions are probing different aspects of the OM permeability defect. For example, EDTA is a known divalent cation cheltator, which disrupts LPS interactions, thus weakening the OM, which may not happen on the 4% SDS plates. Whatever the reason is, it is apparent that even with 96% homology to *Salmonella yraP*, *E. coli yraP* is unable to fully complement the OM defect, even in conditions where it is the only source of *yraP* present in *Salmonella* Δ*yraP* strains.

YraP is predicted to be a lipoprotein. Both SignalP and LipPred prediction have suggested that a lipobox signal sequence is present within the first 19 residues, with C19 being the cysteine residue used for diacylation for lipoprotein insertion. Examining the signal sequence shows that a Leu-Gln-Gly is present directly

before C19, which is a common lipobox motif. The +2 residue following C19 is a valine residue, suggesting that YraP is localised to the OM (Tokuda and Matsuyama, 2004). Mutagenesis to C19A has shown that this mutant is unable to complement the OM defect in Δ*yraP* cells. It is likely that this mutant is unable to complement due to mislocalisation of YraP into the periplasm, where it is unable to function properly. Western blotting has shown that YraP is produced from this mutant, but at much lower levels than WT. This could be due to the effects of proteases in the periplasm, such as DegP, which are targeting and degrading mislocalised and misfolded proteins, causing the lower levels of YraP observed in the C19A mutant. Although further experimentation could be performed, such as the effect of globomycin on YraP function, the mutagenesis data, along with the predictions, strongly suggest that YraP is an OM lipoprotein.

From sucrose density centrifugation separations of the OM and IM followed by western blot analysis with both anti-YraP and anti-BamC antibodies, YraP is present in the same fractions as BamC, suggesting that YraP is localised to the OM. Using fluorescence microscopy, YraP is not a surface localised lipoprotein and therefore must be localised to the inner leaflet of the OM. However, how YraP functions in the OM is still not understood. Co-IP experiments with OM fractions have shown that is not part of any OM complex and is not interacting with any periplasmic species or at least not stable enough for Co-IP with cross-linking to pick it up. A whole cell Co-IP indentified YgaU as a possible interacting partner, but HSQC experiments showed this was a false positive as no interaction could be observed between the two. The pulldown of YgaU is most likely due to the fact that YgaU contains a BON domain. Due to similar

sequences of this BON domain and those of YraP, the anti-YraP antibody may have possibly bound to YgaU during the Co-IP process. Also, it seems that YraP functions as a monomer, as gel filtration and analytical centrifugation have shown that YraP does not multimerise.

HSQC experiments with PG have demonstrated that YraP is able to bind to phospholipids. By mapping the residues involved onto the structure, we have identified one face of the first α-helix of BON domain 2. This has confirmed the predictions that BON domains can bind phospholipids.

The PG binding residues in α3 are present in two distinct faces of the α-helix (**Figure 6.24**). Residues W127, I128 and K131 are located close together along with Q135 on one face and residues R133 and L137 are present on another face. Both faces contain a positively charged residue, W127 and R133, which could aid in binding to the negatively charged PG head group.

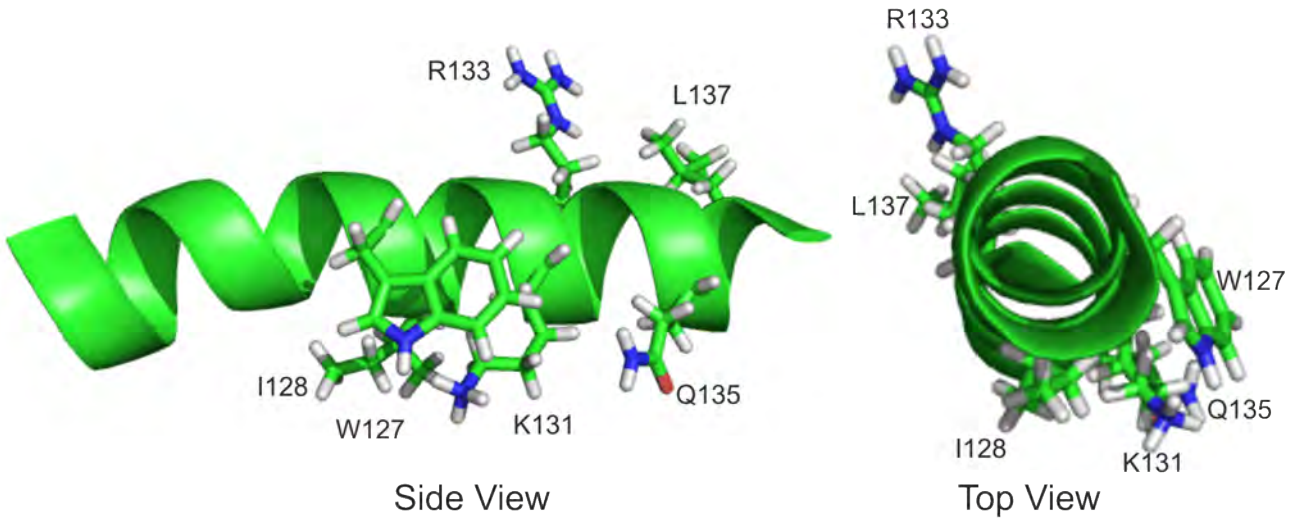


Figure 6.24 Position of the PG binding residues in α3. Carbon atoms are coloured green, nitrogen atoms blue, hydrogen atoms white and oxygen atoms red.

Comparing α1 of BON1 to α3 of BON2, there are some notable differences. Firstly, α1 contains a small flexible region present within its α-helix, which splits it into two α-helices, α1 and α1‟. This is not present in α3, which may explain why no phospholipid binding was observed with α1.

Figure 6.25 shows the helical wheel projections of α1 and α3 of YraP. Both helices contain a mixture of hydrophobic and both positively and negatively charged residues. With α1, there is a small patch of positive charge (R53 and K60) and of negative charge (D47 and E51). The rest of the residues are hydrophobic in nature, with the exception of N55.

With α3, a similar effect is observed, with a mixture of hydrophobic and charged residues, although there are more charged regions than in α1. A notable region is observed with the hydrophilic W127 in close proximity to the positively charged K131, which is next to hydrophilic residues N124 and Q135. W127, K131 and Q135 have been shown to bind to PG by NMR. R133 is present at the opposite of the helix to the other PG binding residues, providing a positive charge that could aid PG binding. Additionally, there is a negative patch composed of E121 and D125 in close proximity to each other.

It should be noted that this projection only takes into the account the helix α1 and omits the shorter α1‟ helix. Within α1‟, there is also the presence of additional positively charged residues (K60, K65 and K60) as well as negatively charged

residues (D61 and E62), which could have had a role in phospholipid binding, were it not for the presence of a small flexible region breaking the α-helix.

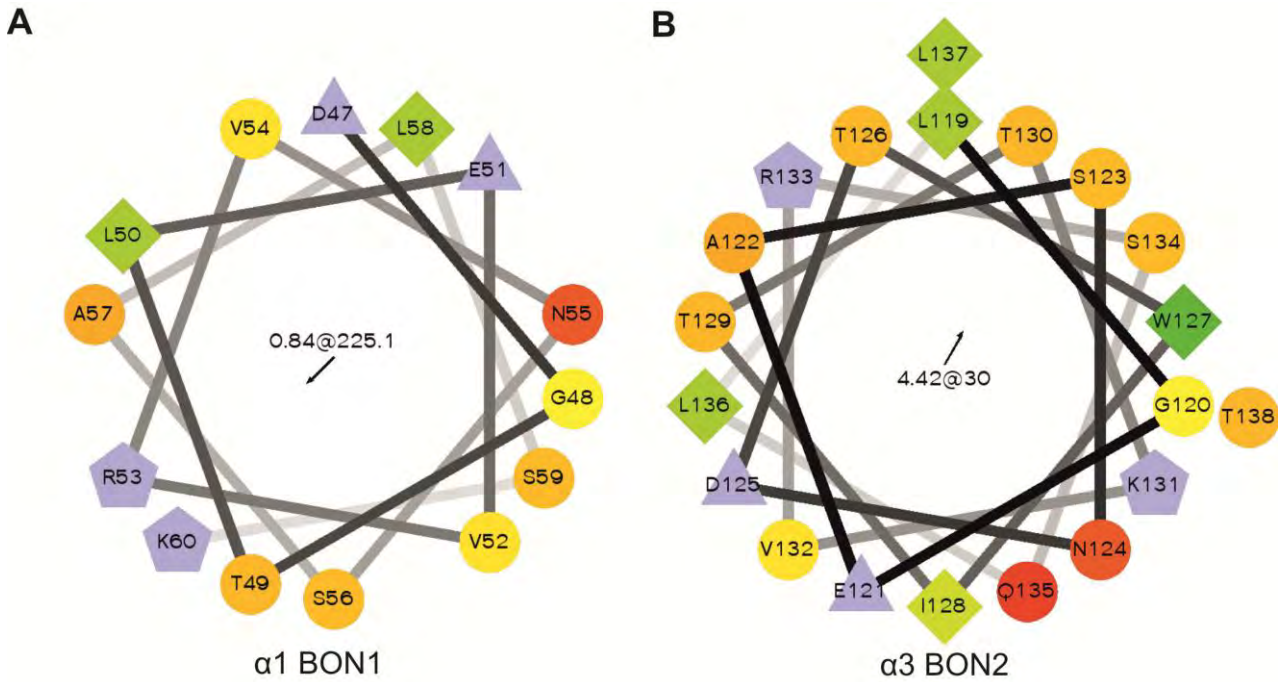


Figure 6.25 Helical wheel projections of A α1 of BON1 (residues 47-60) and B α3 of BON2 (residues 119-138) in YraP. Helical wheels were generated using the online tool found at http://rzlab.ucr.edu/scripts/wheel/. Hydrophilic residues are displayed as circles, hydrophobic residues as diamonds, potentially negatively charged residues as triangles, and potentially positively charged residues as pentagons. Hydrophobicity is colour coded, with the most hydrophobic residues being green, and the amount of green decreasing proportionally to the hydrophobicity, with zero hydrophobicity coloured as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue, and the amount of red decreasing proportionally to the hydrophilicity. The potentially charged residues are coloured grey.

The electrostatic surface of α3 in YraP has been examined (**Figure 6.26**). There is a surface of positive charge within this α-helix mediated by K131 and R133. This surface could be the major interaction point between YraP and PG, whereby these positive residues could interact with the negative charge from the PG head group. α3 also contains an N-terminal negatively charged residue (E121), which could be used to stabilise the helix dipole.

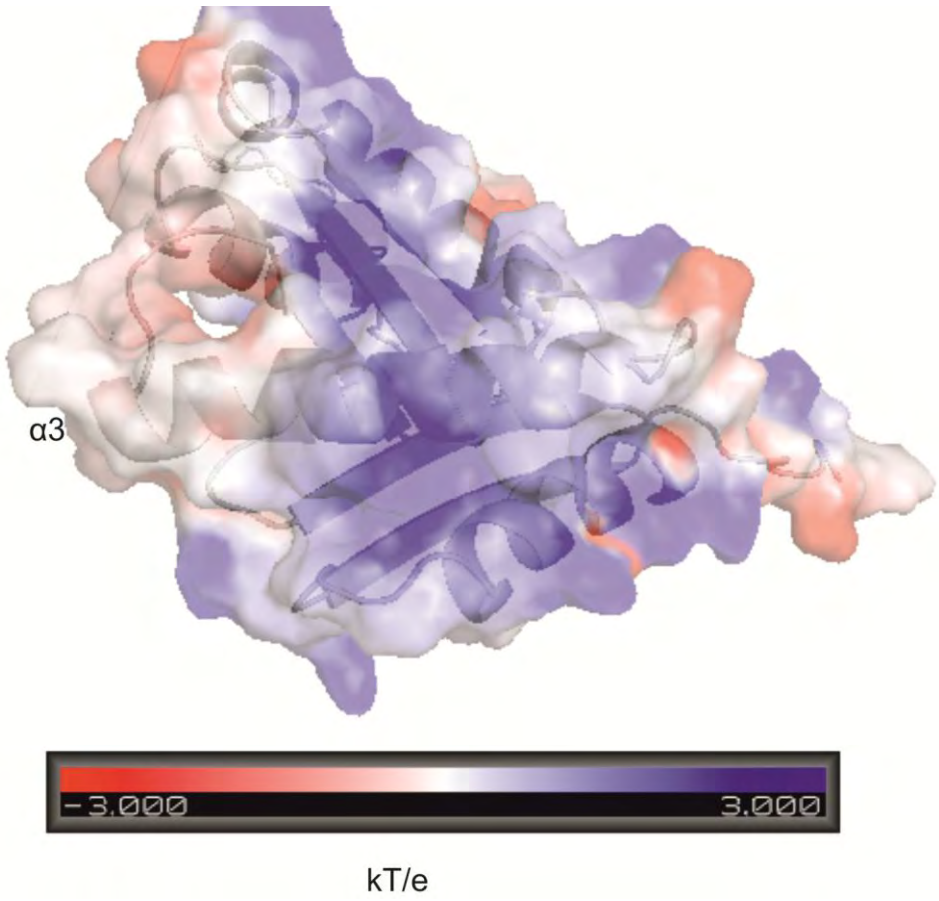


Figure 6.26 Electrostatic model of YraP. Electrostatic potential is define by a scale where areas in red are negatively charged and positively charged surfaces are coloured blue. α3 of BON2 is displayed in the centre of the figure.

Mutagenesis analysis of this helix has in part confirmed that PG binding is essential for *yraP* function. We have mutated this helix substituting the PG binding residues for alanine or oppositely charged residues, either mutating each half individually or the whole helix. By performing our complementation assay on these helix mutants with western blot analysis, many of the mutants were unstable, as no detectable YraP was seen on the blots. Only one mutant, R133-L137 Charged, was not able to complement but produced stable YraP, suggesting that this helix is important for *yraP* function. We hypothesis that its loss of function is due to the negative charges of the aspartate and glutamate residues, that act to repel the PG head group interactions and thus prevent membrane binding. Ideally a thorough characterisation of this mutant is needed, to ensure that this protein is correctly localised and folded. A possible approach to this, would be repeat the OM separations and western blot against YraP, to ensure that it is in the correct location. Using circular dichroism and NMR would be an ideal way to gauge the fold of this mutant.

We have only examined one OM phospholipid. Ideally, these titrations should be repeated to examine if YraP is able to bind to PE and Cardiolipin. These experiments would aid considerably in determining if *yraP* function OM function is PG specific, or if YraP is able to bind most or all of the OM phospholipids.

Our mutagenesis screen has also elucidated the role of the conserved glycine residues in BON domains. From multiple sequence alignments, we have identified two conserved glycine residues, G83 and G160, which are present in each of the BON domains. These glycines are positioned at the end of the

second β-strand in each domain, which agrees with the observations of Teriete *et al.* 2010. Mutating these glycines to valines individually and together and performing the complementation assay, has shown that only the double G83V/G160V mutant was unable complement. By performing western blot analysis, G160V was present at barely visible levels and G83/G160V was absent. This result suggests that these glycines are needed in a structural or mechanical capacity, rather than having a functional role.

Ramachandran plots of the conserved glycine residues G83 and G160 demonstrate that they occupy a flexible region in the protein (**Figure 6.27**). This is consistent with the structure of both YraP and Rv0899, in that are present in the loop region immediately after β2 of each BON domain. Because of this flexibility, the effect of introducing valine residues could cause possible steric clashes, from the bulkier side chains and less rotationally flexible backbone. This could have an effect in unfolding the protein and reducing stability, which is observed in the mutagenesis analysis.

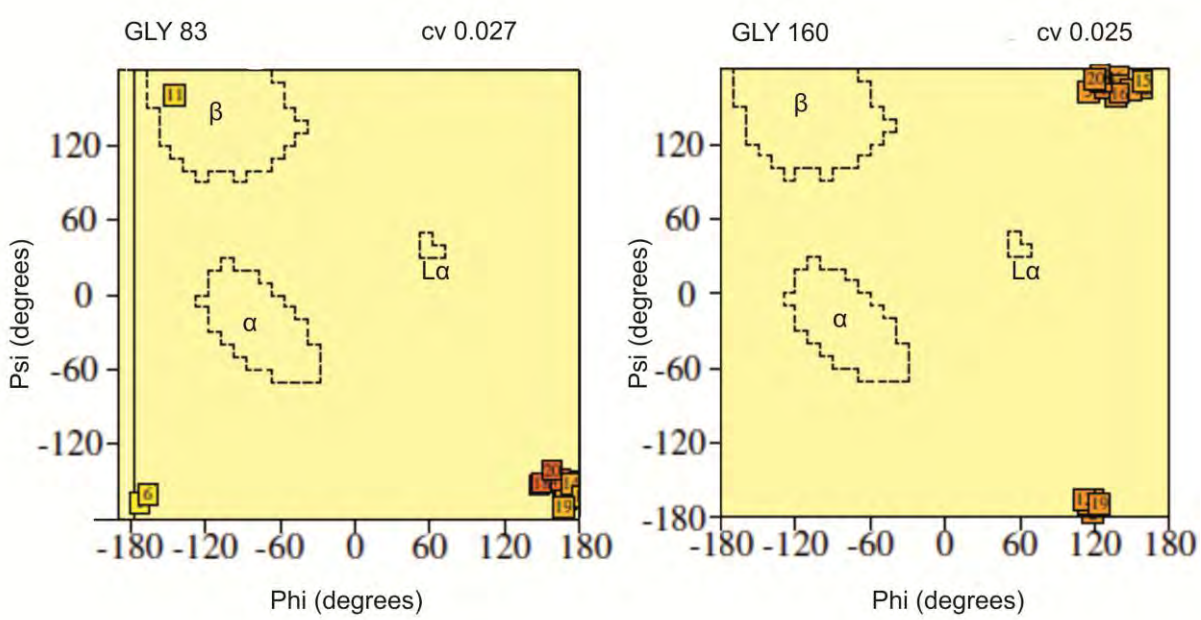


Figure 6.27 Conserved residues G83 and G160 are present in a flexible region of YraP. Ramachandran plots of the 20 structures of the conserved Glycine residues G83 and G160 determined using the NMR structure validation software PROCHECK. Likely areas for secondary structure are shown as α = α-helix, β = β-strand and Lα = left-handed helix. A cv (Circular Variance) value is shown, which indicates the clustering of the 20 of the data points, with low values (< 0.1), indicating good clustering.

The transposon mutagenesis has also demonstrated that the N-terminal region of YraP is essential for its function. Two of the mutants have (L16, Q17) have insertions that disrupt the cysteine acylation site. These mutants fail to complement, possibly due to the failure of YraP to be processed as a lipoprotein. Additionally, transposon screening has also demonstrated that insertions at A37, S40, T49 and L50 also abolish YraP function. This area has been shown to be an unstructured region in the YraP structure. Possible insertions could extend the flexible region, which may position YraP too far away from the membrane for it to carry out its function. Finally, transposon insertions in other areas of YraP were still able to function, particularly the PG binding helix α3, which has been shown to be essential for function. It is possible that PG binding of YraP is mediated by

the helix itself and that transposon insertion has not disrupted the helix. Ideally, structural studies of these mutants are needed to confirm if this is the case.

We have generated suppressor mutants, whereby Δ*yraP* cells have acquired mutations that have restored the OM barrier function. Whole genome sequencing has led to the identification of genes that may have a role in OM biogenesis. This data could give some insight into possible genetic interactions that lead to OM maintenance that are not easily evident as well as shedding light on potential coping strategies that are present in *E. coli* in regards to OM stress. Our analysis has identified *nanT*, *vacJ*, *pgm*, *ychF* and *envZ* as potential areas for further study.

As Δ*yraP* strains exhibit an OM defect, some of the experiments have focused on determining the nature of this defect. The OM consists of phospholipids, proteins and LPS. We hypothesised that at least one of these major components must be perturbed.

Using mass spectrometry approach to quantify the relative abundance of protein species between WT and Δ*yraP* membranes, our analysis demonstrated that there were no major changes between the two samples. Major OMPs, such as OMPA, OMPC and OMPW and lipoproteins such as pal and lpp, were found to be in approximately the same ratio and proteins that are involved in OM biogenesis, such as Lpt and BAM complex proteins, had similar abundances. This result suggests that *yraP* not likely to be involved in protein biogenesis in the OM.

We have been able to investigate the LPS profiles of Δ*yraP* cells. BW25113 is unable to synthesise O-antigen and thus any changes in O-antigen composition could not be examined directly. To circumvent this problem, we used pET20b-*wbbL*, which was transformed into BW25113 Δ*yraP*. This vector has been shown to restore O-antigen in BW25113 (Dr. D. F. Browning), which would allow LPS disruption to be observed. It should be noted that pET20b-*wbbL* is expressing leakily, which may mean that LPS O-antigen production may be at a level greater than WT, thus any subtle changes in LPS composition may not be observed and that only drastic changes could be measured. As a further proof, we examined this in a *Salmonella* Δ*yraP* strain, which is not deficient for LPS production and displays a similar phenotype to our BW25113 Δ*yraP* strain. SDS-PAGE with silver staining demonstrated that there was a major change in the LPS profile in Δ*yraP* strains of both BW25113 pET20b-*wbbL* and *Salmonella*. In the BW25113 Δ*yraP* strain, the O-antigen polymer is absent, when compared to BW25113 WT and Δ*bamE* strains. A similar result is seen in the *Salmonella* strains, where O-antigen is absent and a different banding pattern is observed in comparison to the WT and Δ*bamE* samples. This data suggests that YraP may have role in the transport/insertion of LPS to the OM. As the Co-IP experiments have not shown any interacting partners, this would suggest that *yraP* performs this function in a separate manner to *lptD* and *lptE*, which are major genes involved in LPS insertion. The OM proteomic results have shown that there are no changes in LptD and LptE levels between WT and Δ*yraP* cells, which may therefore support the possibility of another insertion pathway.

The relative quantification of lipids using mass spectrometry and yielded a very interesting result in regards to *yraP* function. This analysis has demonstrated that there are clear differences in the OM phospholipid compositions between these cells. Although further experimentation is required to determine the cause of this difference, so far this result could suggest a possible role in phospholipid assembly in the OM.

mlaA has been shown to have a role in retrograde transport of phospholipids from the OM (Malinverni *et al*., 2009). *yraP* could have a possible role in anterograde transport of phospholipids. Both Δ*mlaA* and Δ*yraP* cells have a similar OM defect, in that they are sensitive to SDS and EDTA conditions, which could be caused by phospholipid mislocalisation at the OM. Δ*yraP* suppressor mutation data has indicated that stop codon mutations present in *mlaA* (*vacJ*), suppress the Δ*yraP* defect on vancomycin. Furthermore, a strain harbouring a double Δ*yraP* and Δ*mlaA* knockout, also grows on vancomycin (F. Morris, unpublished data). This data suggests that *yraP* has a role in phospholipid maintenance of the OM, which could be shared with *mlaA*. In Δ*mlaA* strains, excess phospholipid could be present in the outer leaflet of OM which is caused by the *yraP* activity. Likewise, in Δ*yrap* cells, less phospholipid could be trafficked to the OM, which causes the observed OM defect. There is also the possibility that lipid A levels may be absent or severely reduced in Δ*yraP* cells, which would explain the different LPS profiles. Ideally, further experimentation is needed to confirm this, which would fill in the major gap of OM biogenesis of phospholipid assembly.

Chapter 7

Final Discussion

7.1 Final Discussion

The findings from this study have provided some insights into OM biogenesis. With BamE, a structure-function analysis has been performed, which has highlighted some functions of this protein that were not immediately apparent. With YraP, a structural and partial functional characterisation has helped suggest possible functions that this protein does in regards to OM biogenesis. The details of these findings and the impact that they have on the current understanding of OM biogenesis are discussed in the following sections.

7.2 BamE

Since the discovery of the Bam complex as the machinery involved in OMP biogenesis (Wu *et al*., 2005), a wealth of information regarding Bam complex function has emerged. So far, the structures of accessory lipoproteins are known as well as the N-terminal POTRA domains of BamA. There are also structural and functional insights into how the complex is ordered (Kim *et al*., 2011a). The major outstanding discoveries are the structure of BamA and the functional mechanism of how OMPs are assembled.

Our published study of BamE has contributed to both the structural and functional aspects of the Bam complex. This study represented the first structure of BamE, which had a fold reminiscent of a POTRA domain, suggesting that BamE might have the propensity to perform β-augmentation with precursor OMP substrates. In addition to this, BamE was shown to bind exclusively to the OM lipid PG. From this finding, it could be hypothesised that this aspect of BamE function is to bind to OM lipids and cause local perturbations to aid OMP insertion. Ideally, further

experimentation demonstrating that other Bam complex lipoproteins also have this ability would add weight to this hypothesis. In addition to this, the BamD binding site was shown to overlap with the PG binding site. This could have further functional implications, such as the possibility of temporary dissociation of BamE from the Bam complex during OMP insertion, for its PG binding function.

This study has also obtained functional information on BamE. By exploiting the OM defect of Δ*bamE* cells, we have been able to develop a screen that gauges BamE function. By producing a set of Gly/Cys point mutations in almost every amino acid of BamE and screening them, areas of functional importance in BamE have been identified. We have identified mutants that disrupt the stability of the protein which are present in the hydrophobic core of the protein. We also identified mutants that hindered the function without altering the stability. The majority of this particular set of mutants were localised to the PG and BamD binding sites, highlighting the importance of this site for function.

Although this particular study has focussed on BamE, its overall contribution to the Bam complex is not well understood. It has been shown that BamE binds to BamD and is thought to stabilise the interaction between BamD and BamA (Malinverni *et al*., 2006, Sklar *et al*., 2007a). BamE has also been implicated in modulating the functional state of BamA (Rigel *et al*., 2011). Ideally further experimentation is needed to address the contribution of BamE in OMP assembly. The library of cysteine mutants produced from this study could be used in potential cross linking experiments, as well as the library of BamE functional mutants, which could be used to probe potential genetic interactions.

7.3 YraP

The major part of this study has been the structural and functional work performed on the uncharacterised protein YraP. Starting from a point where YraP was shown to be a lipoprotein and a member of the σ^E regulon (Dartigalongue *et al*., 2001, Onufryk *et al*., 2005) and from various bioinformatic analyses, we have been able to confirm many of these predictions. Numerous other experiments have been performed, gauging some of the phenotypic characteristics of Δ*yraP* cells. Furthermore, we have shown that Δ*yraP* strains exhibit an OM defect, tried to characterise this defect by observing changes in the composition of the OM and suggest that YraP may have a major function in OM biogenesis. In addition to this, the structure of YraP, has been solved using NMR, which has been an invaluable tool in our functional analysis of YraP.

The structure of YraP is the first structure determined of a 2 BON domain architecture protein. Each BON domain contains 2 α-helices and 3 β-strands that are orientated parallel to each other, which is in agreement to the solved structure of rv0899 of *M. tuberculosis* (Teriete *et al*., 2010). The major novelty in the structure of YraP, shows that the 2 BON domains are distinct and are orientated parallel to each other. This particular fold could occur in other multiple BON domain architecture protein and the solved structure of YraP could possibly be used for structural modelling of those proteins.

We have also deduced function of BON domains that have been predicted using bioinformatics. Yeats and Batemam 2003,have predicted that BON domains bind

to phospholipids. NMR titrations of YraP have shown that it is able to bind to PG. As of yet, no actual experimental study has demonstrated that BON domains bind lipids.

This study has also elucidated a possible role critical glycine residue that is conserved in BON domains. Teriete *et al.* 2010 predicted that it had a mechanical role. This study has finally confirmed this hypothesis, where substitution of each glycine residue to a valine led to severely decreased levels of YraP.

The pulldown experiments have not been able to determine protein binding partners to YraP. This has suggested that YraP is not part of a complex and functions on its own, compared to other the genes that exhibit similar OM defects when knocked out (e.g. *bamE*). Furthermore, AUC and gel filtration experiments have demonstrated that YraP does not multimerise, suggesting that it may perform its role in a monomeric state. There is the possibility that YraP might bind lipid and another protein in a complex, but these pulldown approaches have not been able to determine this.

By performing analyses of the OM of Δ*yraP* cells, we have identified two possible OM components that are severely perturbed, LPS and phospholipids. Analysis of LPS extracted from Δ*yraP* strains of BW25113 and SL1344 demonstrated a severe deficiency in O-antigen, suggesting a role in LPS assembly. Our OM proteomic analysis indicated that there were no changes in levels of LptD and LptE, which adds weight to this hypothesis.

The major finding in this project is that Δ*yraP* strains have an altered phospholipid composition in the OM. Although the nature of alteration is still under investigation, it represents a possible major finding in OM biogenesis.

The factors involved in phospholipid assembly have not been identified. There have been the identification of enzymes involved in removing phospholipids from the outer leaflet of the OM (PagP, PldA) (Bishop, 2008), and the identification a possible pathway involved in retrograde phospholipid transport (Malinverni and Silhavy, 2009). Our hypothesis is that YraP is involved in anterograde phospholipid transport. The evidence for this is that firstly, in the generation of suppressor mutants against vancomycin conditions, there were two independent mutants that had generated stop codons in the *mlaA* gene. Additionally, we were able to generate a double *mlaA* and *yraP* knockout strain, which was able to grow on the vancomycin condition. From this, a hypothesis is formed whereby the absence of YraP function of transporting phospholipids to the OM is balanced by the effect of removing the MlaA function of removing phospholipids from the OM. There is also the possibility YraP might be part of the Mla pathway as an unknown factor. Ideally, further experimentation is needed to confirm that there is possible genetic interplay between *mlaA* and *yraP*, but the finding of an *mlaA* suppressor mutant in Δ*yraP* cells and the double knockout is an ideal starting point.

The mass spectrometry approach of the relative quantification of lipids of the OM in Δ*yraP* cells is still under investigation. There is also the possibility that the lipid

perturbation is due to decreased levels of Lipid A in the OM. This would explain the difference in LPS profiles observed in Δ*yraP* cells.

If the hypothesis that YraP is involved in anterograde phospholipid transport, this will represent a major finding as an undiscovered pathway in the poorly understood area of phospholipid assembly in the OM. As *yraP* is not an essential gene, this may suggest that there are other genes involved in phospholipid transport. With further study, the identification of these other factors would aid immensely in understanding OM biogenesis.

REFERENCES

ABDALLAH, J., CALDAS, T., KTHIRI, F., KERN, R. & RICHARME, G. (2007) YhbO protects cells against multiple stresses. *J Bacteriol,* 189**,** 9140-4.

ABE, S., OKUTSU, T., NAKAJIMA, H., KAKUDA, N., OHTSU, I. & AONO, R. (2003) n-Hexane sensitivity of *Escherichia coli* due to low expression of *imp/ostA* encoding an 87 kDa minor protein associated with the outer membrane. *Microbiology,* 149**,** 1265-73.

ADAMIAN, L., NAVEED, H. & LIANG, J. (2010) Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis. *Biochim Biophys Acta,* 1808**,** 1092-102.

ADEREM, A. & ULEVITCH, R. J. (2000) Toll-like receptors in the induction of the innate immune response. *Nature,* 406**,** 782-7.

AHN, V. E., LO, E. I., ENGEL, C. K., CHEN, L., HWANG, P. M., KAY, L. E., BISHOP, R. E. & PRIVE, G. G. (2004) A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J,* 23**,** 2931-41.

ALBA, B. M., LEEDS, J. A., ONUFRYK, C., LU, C. Z. & GROSS, C. A. (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the sigma(E)-dependent extracytoplasmic stress response. *Genes Dev,* 16**,** 2156-68.

ALBRECHT, R. & ZETH, K. (2010) Crystallization and preliminary X-ray data collection of the *Escherichia coli* lipoproteins BamC, BamD and BamE. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 66**,** 1586-90.

ALBRECHT, R. & ZETH, K. (2011) Structural basis of outer membrane protein biogenesis in bacteria. *J Biol Chem,* 286**,** 27792-803.

ANDERSON, M. S. & RAETZ, C. R. (1987) Biosynthesis of lipid A precursors in *Escherichia coli*. A cytoplasmic acyltransferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. *J Biol Chem,* 262**,** 5159-69.

ANDERSON, M. S., ROBERTSON, A. D., MACHER, I. & RAETZ, C. R. (1988) Biosynthesis of lipid A in *Escherichia coli*: identification of UDP-3-O-[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine as a precursor of UDP-N2,O3-bis[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine. *Biochemistry,* 27**,** 1908-17.

ANTONOAEA, R., FURST, M., NISHIYAMA, K. & MULLER, M. (2008) The periplasmic chaperone PpiD interacts with secretory proteins exiting from the SecYEG translocon. *Biochemistry,* 47**,** 5649-56.

ANWARI, K., POGGIO, S., PERRY, A., GATSOS, X., RAMARATHINAM, S. H., WILLIAMSON, N. A., NOINAJ, N., BUCHANAN, S., GABRIEL, K., PURCELL, A. W., JACOBS-WAGNER, C. & LITHGOW, T. (2010) A modular BAM complex in the outer membrane of the alpha-proteobacterium *Caulobacter crescentus*. *PLoS One,* 5**,** e8619.

ANWARI, K., WEBB, C. T., POGGIO, S., PERRY, A. J., BELOUSOFF, M., CELIK, N., RAMM, G., LOVERING, A., SOCKETT, R. E., SMIT, J., JACOBS-WAGNER, C. & LITHGOW, T. (2012) The evolution of new lipoprotein subunits of the bacterial outer membrane BAM complex. *Mol Microbiol,* 84**,** 832-844.

ARIE, J. P., SASSOON, N. & BETTON, J. M. (2001) Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of *Escherichia coli*. *Mol Microbiol,* 39**,** 199-210.

ARNOLD, T., ZETH, K. & LINKE, D. (2010) Omp85 from the thermophilic cyanobacterium *Thermosynechococcus elongatus* differs from proteobacterial Omp85 in structure and domain composition. *J Biol Chem,* 285**,** 18003-15.

AUDET, A., NANTEL, G. & PROULX, P. (1974) Phospholipase A activity in growing *Escherichia coli* cells. *Biochim Biophys Acta,* 348**,** 334-43.

BAADEN, M., MEIER, C. & SANSOM, M. S. (2003) A molecular dynamics investigation of mono and dimeric states of the outer membrane enzyme OMPLA. *J Mol Biol,* 331**,** 177-89.

BABA, T., ARA, T., HASEGAWA, M., TAKAI, Y., OKUMURA, Y., BABA, M., DATSENKO, K. A., TOMITA, M., WANNER, B. L. & MORI, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol,* 2**,** 2006 0008.

BABU, M. M., PRIYA, M. L., SELVAN, A. T., MADERA, M., GOUGH, J., ARAVIND, L. & SANKARAN, K. (2006) A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol,* 188**,** 2761-73.

BADER, M. W., NAVARRE, W. W., SHIAU, W., NIKAIDO, H., FRYE, J. G., MCCLELLAND, M., FANG, F. C. & MILLER, S. I. (2003) Regulation of *Salmonella typhimurium* virulence gene expression by cationic antimicrobial peptides. *Mol Microbiol,* 50**,** 219-30.

BADER, M. W., SANOWAR, S., DALEY, M. E., SCHNEIDER, A. R., CHO, U., XU, W., KLEVIT, R. E., LE MOUAL, H. & MILLER, S. I. (2005) Recognition of antimicrobial peptides by a bacterial sensor kinase. *Cell,* 122**,** 461-72.

BARANOVA, N. & NIKAIDO, H. (2002) The *baeSR* two-component regulatory system activates transcription of the *yegMNOB* (*mdtABCD*) transporter gene cluster in *Escherichia coli* and increases its resistance to novobiocin and deoxycholate. *J Bacteriol,* 184**,** 4168-76.

BATCHELOR, R. A., HARAGUCHI, G. E., HULL, R. A. & HULL, S. I. (1991) Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *J Bacteriol,* 173**,** 5699-704.

BAYER, M. E. (1968) Areas of adhesion between wall and membrane of *Escherichia coli*. *J Gen Microbiol,* 53**,** 395-404.

BAYER, M. E. (1991) Zones of membrane adhesion in the cryofixed envelope of *Escherichia coli*. *J Struct Biol,* 107**,** 268-80.

BECHTLUFT, P., VAN LEEUWEN, R. G., TYREMAN, M., TOMKIEWICZ, D., NOUWEN, N., TEPPER, H. L., DRIESSEN, A. J. & TANS, S. J. (2007) Direct observation of chaperone-induced changes in a protein folding pathway. *Science,* 318**,** 1458-61.

BEHRENS, S., MAIER, R., DE COCK, H., SCHMID, F. X. & GROSS, C. A. (2001) The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J,* 20**,** 285-94.

BELL, R. M., MAVIS, R. D., OSBORN, M. J. & VAGELOS, P. R. (1971) Enzymes of phospholipid metabolism: localization in the cytoplasmic and outer membrane of the cell envelope of *Escherichia coli* and *Salmonella typhimurium*. *Biochim Biophys Acta,* 249**,** 628-35.

BENNING, C. (2008) A role for lipid trafficking in chloroplast biogenesis. *Prog Lipid Res,* 47**,** 381-9.

BENNION, D., CHARLSON, E. S., COON, E. & MISRA, R. (2010) Dissection of beta-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of *Escherichia coli*. *Mol Microbiol,* 77**,** 1153-71.

BERNADAC, A., GAVIOLI, M., LAZZARONI, J. C., RAINA, S. & LLOUBES, R. (1998) *Escherichia coli* tol-pal mutants form outer membrane vesicles. *J Bacteriol,* 180**,** 4872-8.

BERNARD, G. R., VINCENT, J. L., LATERRE, P. F., LAROSA, S. P., DHAINAUT, J. F., LOPEZ-RODRIGUEZ, A., STEINGRUB, J. S., GARBER, G. E., HELTERBRAND, J. D., ELY, E. W. & FISHER, C. J., JR. (2001) Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med,* 344**,** 699-709.

BEUTLER, B. & CERAMI, A. (1988) Tumor necrosis, cachexia, shock, and inflammation: a common mediator. *Annu Rev Biochem,* 57**,** 505-18.

BISHOP, P. (2000) Recent Advances in the Medicinal Chemistry of Anti-Infective Agents--Third International Symposium. 23-26 July 2000, Cambridge, UK. *IDrugs,* 3**,** 1146-8.

BISHOP, R. E. (2005) The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. *Mol Microbiol,* 57**,** 900-12.

BISHOP, R. E. (2008) Structural biology of membrane-intrinsic beta-barrel enzymes: sentinels of the bacterial outer membrane. *Biochim Biophys Acta,* 1778**,** 1881-96.

BITTO, E. & MCKAY, D. B. (2002) Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure,* 10**,** 1489-98.

BITTO, E. & MCKAY, D. B. (2003) The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J Biol Chem,* 278**,** 49316-22.
BITTO, E. & MCKAY, D. B. (2004) Binding of phage-display-selected peptides to the periplasmic chaperone protein SurA mimics binding of unfolded outer membrane proteins. *FEBS Lett,* 568**,** 94-8.
BODELON, G., MARIN, E. & FERNANDEZ, L. A. (2009) Role of periplasmic chaperones and BamA (YaeT/Omp85) in folding and secretion of intimin from enteropathogenic *Escherichia coli* strains. *J Bacteriol,* 191**,** 5169-79.
BOS, M. P., ROBERT, V. & TOMMASSEN, J. (2007a) Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol,* 61**,** 191-214.
BOS, M. P., ROBERT, V. & TOMMASSEN, J. (2007b) Functioning of outer membrane protein assembly factor Omp85 requires a single POTRA domain. *EMBO Rep,* 8**,** 1149-54.
BOS, M. P. & TOMMASSEN, J. (2004) Biogenesis of the Gram-negative bacterial outer membrane. *Curr Opin Microbiol,* 7**,** 610-6.
BOS, M. P. & TOMMASSEN, J. (2011) The LptD chaperone LptE is not directly involved in lipopolysaccharide transport in *Neisseria meningitidis*. *J Biol Chem,* 286**,** 28688-96.
BOTHMANN, H. & PLUCKTHUN, A. (2000) The periplasmic *Escherichia coli* peptidylprolyl cis,trans-isomerase FkpA. I. Increased functional expression of antibody fragments with and without cis-prolines. *J Biol Chem,* 275**,** 17100-5.
BOWERS, C. W., LAU, F. & SILHAVY, T. J. (2003) Secretion of LamB-LacZ by the signal recognition particle pathway of *Escherichia coli*. *J Bacteriol,* 185**,** 5697-705.
BRABETZ, W., MULLER-LOENNIES, S. & BRADE, H. (2000) 3-Deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferase (WaaA) and kdo kinase (KdkA) of Haemophilus influenzae are both required to complement a *waaA* knockout mutation of *Escherichia coli*. *J Biol Chem,* 275**,** 34954-62.
BRAUN, M. & SILHAVY, T. J. (2002) Imp/OstA is required for cell envelope biogenesis in *Escherichia coli*. *Mol Microbiol,* 45**,** 1289-302.
BROZEK, K. A., BULAWA, C. E. & RAETZ, C. R. (1987) Biosynthesis of lipid A precursors in *Escherichia coli*. A membrane-bound enzyme that transfers a palmitoyl residue from a glycerophospholipid to lipid X. *J Biol Chem,* 262**,** 5170-9.
BULIERIS, P. V., BEHRENS, S., HOLST, O. & KLEINSCHMIDT, J. H. (2003) Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide. *J Biol Chem,* 278**,** 9092-9.
BULLMANN, L., HAARMANN, R., MIRUS, O., BREDEMEIER, R., HEMPEL, F., MAIER, U. G. & SCHLEIFF, E. (2010) Filling the gap, evolutionarily conserved Omp85 in plastids of *chromalveolates*. *J Biol Chem,* 285**,** 6848-56.
BURGESS, N. K., DAO, T. P., STANLEY, A. M. & FLEMING, K. G. (2008) Beta-barrel proteins that reside in the *Escherichia coli* outer membrane in vivo demonstrate varied folding behavior in vitro. *J Biol Chem,* 283**,** 26748-58.
BURY-MONE, S., NOMANE, Y., REYMOND, N., BARBET, R., JACQUET, E., IMBEAUD, S., JACQ, A. & BOULOC, P. (2009) Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet,* 5**,** e1000651.
CAMPBELL, J. A., DAVIES, G. J., BULONE, V. & HENRISSAT, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem J,* 326 (Pt 3)**,** 929-39.
CARLSON, J. H. & SILHAVY, T. J. (1993) Signal sequence processing is required for the assembly of LamB trimers in the outer membrane of *Escherichia coli*. *J Bacteriol,* 175**,** 3327-34.
CASALI, N. & RILEY, L. W. (2007) A phylogenomic analysis of the Actinomycetales *mce* operons. *BMC Genomics,* 8**,** 60.

CASCALES, E., BUCHANAN, S. K., DUCHE, D., KLEANTHOUS, C., LLOUBES, R., POSTLE, K., RILEY, M., SLATIN, S. & CAVARD, D. (2007) Colicin biology. *Microbiol Mol Biol Rev,* 71**,** 158-229.

CASTILLOKELLER, M. & MISRA, R. (2003) Protease-deficient DegP suppresses lethal effects of a mutant OmpC protein by its capture. *J Bacteriol,* 185**,** 148-54.

CHARLSON, E. S., WERNER, J. N. & MISRA, R. (2006) Differential effects of *yfgL* mutation on *Escherichia coli* outer membrane proteins and lipopolysaccharide. *J Bacteriol,* 188**,** 7186-94.

CHEN, R. & HENNING, U. (1996) A periplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. *Mol Microbiol,* 19**,** 1287-94.

CHIANG, L. W., KOVARI, I. & HOWE, M. M. (1993) Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. *PCR Methods Appl,* 2**,** 210-7.

CHIMALAKONDA, G., RUIZ, N., CHNG, S. S., GARNER, R. A., KAHNE, D. & SILHAVY, T. J. (2011) Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 108**,** 2492-7.

CHNG, S. S., GRONENBERG, L. S. & KAHNE, D. (2010a) Proteins required for lipopolysaccharide assembly in *Escherichia coli* form a transenvelope complex. *Biochemistry,* 49**,** 4565-7.

CHNG, S. S., RUIZ, N., CHIMALAKONDA, G., SILHAVY, T. J. & KAHNE, D. (2010b) Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. *Proc Natl Acad Sci U S A,* 107**,** 5363-8.

CHO, U. S., BADER, M. W., AMAYA, M. F., DALEY, M. E., KLEVIT, R. E., MILLER, S. I. & XU, W. (2006) Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J Mol Biol,* 356**,** 1193-206.

CLANTIN, B., DELATTRE, A. S., RUCKTOOA, P., SAINT, N., MELI, A. C., LOCHT, C., JACOB-DUBUISSON, F. & VILLERET, V. (2007) Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science,* 317**,** 957-61.

CLANTIN, B., LEYRAT, C., WOHLKONIG, A., HODAK, H., RIBEIRO EDE, A., JR., MARTINEZ, N., BAUD, C., SMET-NOCCA, C., VILLERET, V., JACOB-DUBUISSON, F. & JAMIN, M. (2009) Structure and plasticity of the peptidyl-prolyl isomerase Par27 of *Bordetella pertussis* revealed by X-ray diffraction and small-angle X-ray scattering. *J Struct Biol,* 169**,** 253-65.

CLAVEL, T., GERMON, P., VIANNEY, A., PORTALIER, R. & LAZZARONI, J. C. (1998) TolB protein of *Escherichia coli* K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Mol Microbiol,* 29**,** 359-67.

CLEMENTZ, T., BEDNARSKI, J. J. & RAETZ, C. R. (1996) Function of the *htrB* high temperature requirement gene of *Escherchia coli* in the acylation of lipid A: HtrB catalyzed incorporation of laurate. *J Biol Chem,* 271**,** 12095-102.

CLEMENTZ, T., ZHOU, Z. & RAETZ, C. R. (1997) Function of the *Escherichia coli* msbB gene, a multicopy suppressor of htrB knockouts, in the acylation of lipid A. Acylation by MsbB follows laurate incorporation by HtrB. *J Biol Chem,* 272**,** 10353-60.

COLLIN, S., GUILVOUT, I., NICKERSON, N. N. & PUGSLEY, A. P. (2011) Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin. *Mol Microbiol,* 80**,** 655-65.

COLLINS, R. F. & DERRICK, J. P. (2007) Wza: a new structural paradigm for outer membrane secretory proteins? *Trends Microbiol,* 15**,** 96-100.

CORNILESCU, G., DELAGLIO, F. & BAX, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR,* 13**,** 289-302.

COWLES, C. E., LI, Y., SEMMELHACK, M. F., CRISTEA, I. M. & SILHAVY, T. J. (2011) The free and bound forms of Lpp occupy distinct subcellular locations in *Escherichia coli*. *Mol Microbiol,* 79**,** 1168-81.

CREEGER, E. S., CHEN, J. F. & ROTHFIELD, L. I. (1979) Cloning of genes for bacterial glycosyltransferases. II. Selection of a hybrid plasmid carrying the *rfah* gene. *J Biol Chem,* 254**,** 811-5.

CRONAN, J. E., JR. & WULFF, D. L. (1969) A role for phospholipid hydrolysis in the lysis of *Escherichia coli* infected with bacteriophage T4. *Virology,* 38**,** 241-6.

DALEY, D. O., RAPP, M., GRANSETH, E., MELEN, K., DREW, D. & VON HEIJNE, G. (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science,* 308**,** 1321-3.

DANESE, P. N. & SILHAVY, T. J. (1997) The sigma(E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes Dev,* 11**,** 1183-93.

DANESE, P. N. & SILHAVY, T. J. (1998) Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. *Annu Rev Genet,* 32**,** 59-94.

DANESE, P. N., SNYDER, W. B., COSMA, C. L., DAVIS, L. J. & SILHAVY, T. J. (1995) The Cpx two-component signal transduction pathway of *Escherichia coli* regulates transcription of the gene specifying the stress-inducible periplasmic protease, DegP. *Genes Dev,* 9**,** 387-98.

DANIELS, C. & MORONA, R. (1999) Analysis of *Shigella flexneri* wzz (Rol) function by mutagenesis and cross-linking: *wzz* is able to oligomerize. *Mol Microbiol,* 34**,** 181-94.

DARTIGALONGUE, C., MISSIAKAS, D. & RAINA, S. (2001) Characterization of the *Escherichia coli* sigma E regulon. *J Biol Chem,* 276**,** 20866-75.

DARTIGALONGUE, C. & RAINA, S. (1998) A new heat-shock gene, *ppiD*, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. *EMBO J,* 17**,** 3968-80.

DATSENKO, K. A. & WANNER, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A,* 97**,** 6640-5.

DE COCK, H., STRUYVE, M., KLEEREBEZEM, M., VAN DER KRIFT, T. & TOMMASSEN, J. (1997) Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of *Escherichia coli* K-12. *J Mol Biol,* 269**,** 473-8.

DE GEUS, P., VAN DIE, I., BERGMANS, H., TOMMASSEN, J. & DE HAAS, G. (1983) Molecular cloning of *pldA*, the structural gene for outer membrane phospholipase of *E. coli* K12. *Mol Gen Genet,* 190**,** 150-5.

DE LEEUW, E., TE KAAT, K., MOSER, C., MENESTRINA, G., DEMEL, R., DE KRUIJFF, B., OUDEGA, B., LUIRINK, J. & SINNING, I. (2000) Anionic phospholipids are involved in membrane association of FtsY and stimulate its GTPase activity. *EMBO J,* 19**,** 531-41.

DEKKER, N. (2000) Outer-membrane phospholipase A: known structure, unknown biological function. *Mol Microbiol,* 35**,** 711-7.

DEKKER, N., TOMMASSEN, J., LUSTIG, A., ROSENBUSCH, J. P. & VERHEIJ, H. M. (1997) Dimerization regulates the enzymatic activity of *Escherichia coli* outer membrane phospholipase A. *J Biol Chem,* 272**,** 3179-84.

DELAGLIO, F., GRZESIEK, S., VUISTER, G. W., ZHU, G., PFEIFER, J. & BAX, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR,* 6**,** 277-93.

DELATTRE, A. S., CLANTIN, B., SAINT, N., LOCHT, C., VILLERET, V. & JACOB-DUBUISSON, F. (2010) Functional importance of a conserved sequence motif in FhaC, a prototypic member of the TpsB/Omp85 superfamily. *FEBS J,* 277**,** 4755-65.

DEN BLAAUWEN, T., LINDQVIST, A., LOWE, J. & NANNINGA, N. (2001) Distribution of the *Escherichia coli* structural maintenance of chromosomes (SMC)-like protein MukB in the cell. *Mol Microbiol,* 42**,** 1179-88.

DINARELLO, C. A. (1991) Interleukin-1 and interleukin-1 antagonism. *Blood,* 77**,** 1627-52.

DOERRLER, W. T., REEDY, M. C. & RAETZ, C. R. (2001) An *Escherichia coli* mutant defective in lipid export. *J Biol Chem,* 276**,** 11461-4.

DONOHUE-ROLFE, A. M. & SCHAECHTER, M. (1980) Translocation of phospholipids from the inner to the outer membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 77**,** 1867-71.
DRAKE, T. A., CHENG, J., CHANG, A. & TAYLOR, F. B., JR. (1993) Expression of tissue factor, thrombomodulin, and E-selectin in baboons with lethal *Escherichia coli* sepsis. *Am J Pathol,* 142**,** 1458-70.
DU PLESSIS, D. J., NOUWEN, N. & DRIESSEN, A. J. (2006) Subunit a of cytochrome o oxidase requires both YidC and SecYEG for membrane insertion. *J Biol Chem,* 281**,** 12248-52.
DU PLESSIS, D. J., NOUWEN, N. & DRIESSEN, A. J. (2010) The Sec translocase. *Biochim Biophys Acta,* 1808**,** 851-65.
EHRMANN, M., EHRLE, R., HOFMANN, E., BOOS, W. & SCHLOSSER, A. (1998) The ABC maltose transporter. *Mol Microbiol,* 29**,** 685-94.
ERICKSON, J. W. & GROSS, C. A. (1989) Identification of the sigma E subunit of *Escherichia coli* RNA polymerase: a second alternate sigma factor involved in high-temperature gene expression. *Genes Dev,* 3**,** 1462-71.
ESMON, C. T. (2000) Regulation of blood coagulation. *Biochim Biophys Acta,* 1477**,** 349-60.
EVANICS, F., HWANG, P. M., CHENG, Y., KAY, L. E. & PROSSER, R. S. (2006) Topology of an outer-membrane enzyme: Measuring oxygen and water contacts in solution NMR studies of PagP. *J Am Chem Soc,* 128**,** 8256-64.
FEKKES, P. & DRIESSEN, A. J. (1999) Protein targeting to the bacterial cytoplasmic membrane. *Microbiol Mol Biol Rev,* 63**,** 161-73.
FLYNN, J. M., LEVCHENKO, I., SAUER, R. T. & BAKER, T. A. (2004) Modulating substrate choice: the SspB adaptor delivers a regulator of the extracytoplasmic-stress response to the AAA+ protease ClpXP for degradation. *Genes Dev,* 18**,** 2292-301.
FRASER, C. M., CASJENS, S., HUANG, W. M., SUTTON, G. G., CLAYTON, R., LATHIGRA, R., WHITE, O., KETCHUM, K. A., DODSON, R., HICKEY, E. K., GWINN, M., DOUGHERTY, B., TOMB, J. F., FLEISCHMANN, R. D., RICHARDSON, D., PETERSON, J., KERLAVAGE, A. R., QUACKENBUSH, J., SALZBERG, S., HANSON, M., VAN VUGT, R., PALMER, N., ADAMS, M. D., GOCAYNE, J., WEIDMAN, J., UTTERBACK, T., WATTHEY, L., MCDONALD, L., ARTIACH, P., BOWMAN, C., GARLAND, S., FUJI, C., COTTON, M. D., HORST, K., ROBERTS, K., HATCH, B., SMITH, H. O. & VENTER, J. C. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature,* 390**,** 580-6.
FREINKMAN, E., CHNG, S. S. & KAHNE, D. (2011) The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proc Natl Acad Sci U S A,* 108**,** 2486-91.
GAMA-CASTRO, S., JIMENEZ-JACINTO, V., PERALTA-GIL, M., SANTOS-ZAVALETA, A., PENALOZA-SPINOLA, M. I., CONTRERAS-MOREIRA, B., SEGURA-SALAZAR, J., MUNIZ-RASCADO, L., MARTINEZ-FLORES, I., SALGADO, H., BONAVIDES-MARTINEZ, C., ABREU-GOODGER, C., RODRIGUEZ-PENAGOS, C., MIRANDA-RIOS, J., MORETT, E., MERINO, E., HUERTA, A. M., TREVINO-QUINTANILLA, L. & COLLADO-VIDES, J. (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res,* 36**,** D120-4.
GATSOS, X., PERRY, A. J., ANWARI, K., DOLEZAL, P., WOLYNEC, P. P., LIKIC, V. A., PURCELL, A. W., BUCHANAN, S. K. & LITHGOW, T. (2008) Protein secretion and outer membrane assembly in Alphaproteobacteria. *FEMS Microbiol Rev,* 32**,** 995-1009.
GATZEVA-TOPALOVA, P. Z., WALTON, T. A. & SOUSA, M. C. (2008) Crystal structure of YaeT: conformational flexibility and substrate recognition. *Structure,* 16**,** 1873-81.
GENNITY, J. M. & INOUYE, M. (1991) The protein sequence responsible for lipoprotein membrane localization in *Escherichia coli* exhibits remarkable specificity. *J Biol Chem,* 266**,** 16458-64.

GENTLE, I., GABRIEL, K., BEECH, P., WALLER, R. & LITHGOW, T. (2004) The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J Cell Biol,* 164**,** 19-24.

GENTLE, I. E., BURRI, L. & LITHGOW, T. (2005) Molecular architecture and function of the Omp85 family of proteins. *Mol Microbiol,* 58**,** 1216-25.

GIUSEPPE, P. O., VON ATZINGEN, M., NASCIMENTO, A. L., ZANCHIN, N. I. & GUIMARAES, B. G. (2010) The crystal structure of the leptospiral hypothetical protein LIC12922 reveals homology with the periplasmic chaperone SurA. *J Struct Biol,* 173**,** 312-22.

GRIGOROVA, I. L., CHABA, R., ZHONG, H. J., ALBA, B. M., RHODIUS, V., HERMAN, C. & GROSS, C. A. (2004) Fine-tuning of the *Escherichia coli* sigmaE envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA. *Genes Dev,* 18**,** 2686-97.

GUILVOUT, I., CHAMI, M., ENGEL, A., PUGSLEY, A. P. & BAYAN, N. (2006) Bacterial outer membrane secretin PulD assembles and inserts into the inner membrane in the absence of its pilotin. *EMBO J,* 25**,** 5241-9.

GUNTERT, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol,* 278**,** 353-78.

HAGAN, C. L. & KAHNE, D. (2011) The reconstituted *Escherichia coli* Bam complex catalyzes multiple rounds of beta-barrel assembly. *Biochemistry,* 50**,** 7444-6.

HAGAN, C. L., KIM, S. & KAHNE, D. (2010) Reconstitution of outer membrane protein assembly from purified components. *Science,* 328**,** 890-2.

HAGAN, C. L., SILHAVY, T. J. & KAHNE, D. (2011) beta-Barrel membrane protein assembly by the Bam complex. *Annu Rev Biochem,* 80**,** 189-210.

HANAHAN, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol,* 166**,** 557-80.

HARA, T., MATSUYAMA, S. & TOKUDA, H. (2003) Mechanism underlying the inner membrane retention of *Escherichia coli* lipoproteins caused by Lol avoidance signals. *J Biol Chem,* 278**,** 40408-14.

HARDAWAY, K. L. & BULLER, C. S. (1979) Effect of ethylenediaminetetraacetate on phospholipids and outer membrane function in *Escherichia coli*. *J Bacteriol,* 137**,** 62-8.

HARMS, N., KONINGSTEIN, G., DONTJE, W., MULLER, M., OUDEGA, B., LUIRINK, J. & DE COCK, H. (2001) The early interaction of the outer membrane protein phoe with the periplasmic chaperone Skp occurs at the cytoplasmic membrane. *J Biol Chem,* 276**,** 18804-11.

HAYASHI, S. & WU, H. C. (1990) Lipoproteins in bacteria. *J Bioenerg Biomembr,* 22**,** 451-71.

HEINRICHS, D. E., MONTEIRO, M. A., PERRY, M. B. & WHITFIELD, C. (1998a) The assembly system for the lipopolysaccharide R2 core-type of *Escherichia coli* is a hybrid of those found in *Escherichia coli* K-12 and *Salmonella enterica*. Structure and function of the R2 WaaK and WaaL homologs. *J Biol Chem,* 273**,** 8849-59.

HEINRICHS, D. E., YETHON, J. A. & WHITFIELD, C. (1998b) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol Microbiol,* 30**,** 221-32.

HELBIG, J. H., KONIG, B., KNOSPE, H., BUBERT, B., YU, C., LUCK, C. P., RIBOLDI-TUNNICLIFFE, A., HILGENFELD, R., JACOBS, E., HACKER, J. & FISCHER, G. (2003) The PPIase active site of *Legionella pneumophila* Mip protein is involved in the infection of eukaryotic host cells. *Biol Chem,* 384**,** 125-37.

HELBIG, S., PATZER, S. I., SCHIENE-FISCHER, C., ZETH, K. & BRAUN, V. (2010) Activation of colicin M by the FkpA prolyl cis-trans isomerase/chaperone. *J Biol Chem,* 286**,** 6280-90.

HENNECKE, G., NOLTE, J., VOLKMER-ENGERT, R., SCHNEIDER-MERGENER, J. & BEHRENS, S. (2005) The periplasmic chaperone SurA exploits two features characteristic of integral outer membrane proteins for selective substrate recognition. *J Biol Chem,* 280**,** 23540-8.

HEUCK, A., SCHLEIFFER, A. & CLAUSEN, T. (2011) Augmenting beta-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. *J Mol Biol,* 406**,** 659-66.

HORNE, S. M., KOTTOM, T. J., NOLAN, L. K. & YOUNG, K. D. (1997) Decreased intracellular survival of an *fkpA* mutant of *Salmonella typhimurium* Copenhagen. *Infect Immun,* 65**,** 806-10.

HORREVOETS, A. J., HACKENG, T. M., VERHEIJ, H. M., DIJKMAN, R. & DE HAAS, G. H. (1989) Kinetic characterization of *Escherichia coli* outer membrane phospholipase A using mixed detergent-lipid micelles. *Biochemistry,* 28**,** 1139-47.

HOSHINO, K., TAKEUCHI, O., KAWAI, T., SANJO, H., OGAWA, T., TAKEDA, Y., TAKEDA, K. & AKIRA, S. (1999) Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J Immunol,* 162**,** 3749-52.

HSU, J. L., HUANG, S. Y., CHOW, N. H. & CHEN, S. H. (2003) Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem,* 75**,** 6843-52.

HUIJBREGTS, R. P., DE KROON, A. I. & DE KRUIJFF, B. (2000) Topology and transport of membrane lipids in bacteria. *Biochim Biophys Acta,* 1469**,** 43-61.

HULLMANN, J., PATZER, S. I., ROMER, C., HANTKE, K. & BRAUN, V. (2008) Periplasmic chaperone FkpA is essential for imported colicin M toxicity. *Mol Microbiol,* 69**,** 926-37.

HUMPHREYS, S., ROWLEY, G., STEVENSON, A., KENYON, W. J., SPECTOR, M. P. & ROBERTS, M. (2003) Role of periplasmic peptidylprolyl isomerases in *Salmonella enterica* serovar Typhimurium virulence. *Infect Immun,* 71**,** 5386-8.

HWANG, P. M., BISHOP, R. E. & KAY, L. E. (2004) The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc Natl Acad Sci U S A,* 101**,** 9618-23.

HWANG, P. M., CHOY, W. Y., LO, E. I., CHEN, L., FORMAN-KAY, J. D., RAETZ, C. R., PRIVE, G. G., BISHOP, R. E. & KAY, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci U S A,* 99**,** 13560-5.

IEVA, R. & BERNSTEIN, H. D. (2009) Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc Natl Acad Sci U S A,* 106**,** 19120-5.

ITO, Y., KANAMARU, K., TANIGUCHI, N., MIYAMOTO, S. & TOKUDA, H. (2006) A novel ligand bound ABC transporter, LolCDE, provides insights into the molecular mechanisms underlying membrane detachment of bacterial lipoproteins. *Mol Microbiol,* 62**,** 1064-75.

JACKMAN, J. E., RAETZ, C. R. & FIERKE, C. A. (1999) UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase of *Escherichia coli* is a zinc metalloenzyme. *Biochemistry,* 38**,** 1902-11.

JACKMAN, J. E., RAETZ, C. R. & FIERKE, C. A. (2001) Site-directed mutagenesis of the bacterial metalloamidase UDP-(3-O-acyl)-N-acetylglucosamine deacetylase (LpxC). Identification of the zinc binding site. *Biochemistry,* 40**,** 514-23.

JACOB-DUBUISSON, F., VILLERET, V., CLANTIN, B., DELATTRE, A. S. & SAINT, N. (2009) First structural insights into the TpsB/Omp85 superfamily. *Biol Chem,* 390**,** 675-84.

JANSEN, C., HEUTINK, M., TOMMASSEN, J. & DE COCK, H. (2000) The assembly pathway of outer membrane protein PhoE of *Escherichia coli*. *Eur J Biochem,* 267**,** 3792-800.

JARCHOW, S., LUCK, C., GORG, A. & SKERRA, A. (2008) Identification of potential substrate proteins for the periplasmic *Escherichia coli* chaperone Skp. *Proteomics,* 8**,** 4987-94.

JIA, W., EL ZOEIBY, A., PETRUZZIELLO, T. N., JAYABALASINGHAM, B., SEYEDIRASHTI, S. & BISHOP, R. E. (2004) Lipid trafficking controls endotoxin acylation in outer membranes of *Escherichia coli*. *J Biol Chem,* 279**,** 44966-75.

JONES, N. C. & OSBORN, M. J. (1977) Translocation of phospholipids between the outer and inner membranes of *Salmonella typhimurium*. *J Biol Chem,* 252**,** 7405-12.

JUNCKER, A. S., WILLENBROCK, H., VON HEIJNE, G., BRUNAK, S., NIELSEN, H. & KROGH, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci,* 12**,** 1652-62.

JUSTICE, S. S., HUNSTAD, D. A., HARPER, J. R., DUGUAY, A. R., PINKNER, J. S., BANN, J., FRIEDEN, C., SILHAVY, T. J. & HULTGREN, S. J. (2005) Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability, but SurA is required for pilus biogenesis in *Escherichia coli*. *J Bacteriol,* 187**,** 7680-6.

KADRMAS, J. L., BROZEK, K. A. & RAETZ, C. R. (1996) Lipopolysaccharide core glycosylation in Rhizobium leguminosarum. An unusual mannosyl transferase resembling the heptosyl transferase I of *Escherichia coli*. *J Biol Chem,* 271**,** 32119-25.

KAMIO, Y. & NIKAIDO, H. (1976) Outer membrane of *Salmonella typhimurium*: accessibility of phospholipid head groups to phospholipase c and cyanogen bromide activated dextran in the external medium. *Biochemistry,* 15**,** 2561-70.

KANAMARU, K., TANIGUCHI, N., MIYAMOTO, S., NARITA, S. & TOKUDA, H. (2007) Complete reconstitution of an ATP-binding cassette transporter LolCDE complex from separately isolated subunits. *FEBS J,* 274**,** 3034-43.

KANEHARA, K., ITO, K. & AKIYAMA, Y. (2002) YaeL (EcfE) activates the sigma(E) pathway of stress response through a site-2 cleavage of anti-sigma(E), RseA. *Genes Dev,* 16**,** 2147-55.

KAROW, M., FAYET, O. & GEORGOPOULOS, C. (1992) The lethal phenotype caused by null mutations in the *Escherichia coli htrB* gene is suppressed by mutations in the *accBC* operon, encoding two subunits of acetyl coenzyme A carboxylase. *J Bacteriol,* 174**,** 7407-18.

KAROW, M. & GEORGOPOULOS, C. (1991) Sequencing, mutational analysis, and transcriptional regulation of the *Escherichia coli htrB* gene. *Mol Microbiol,* 5**,** 2285-92.

KEENLEYSIDE, W. J., PERRY, M., MACLEAN, L., POPPE, C. & WHITFIELD, C. (1994) A plasmid-encoded *rfbO*:54 gene cluster is required for biosynthesis of the O:54 antigen in *Salmonella enterica* serovar Borreze. *Mol Microbiol,* 11**,** 437-48.

KELLENBERGER, E. (1990) The 'Bayer bridges' confronted with results from improved electron microscopy methods. *Mol Microbiol,* 4**,** 697-705.

KELLY, T. M., STACHULA, S. A., RAETZ, C. R. & ANDERSON, M. S. (1993) The firA gene of *Escherichia coli* encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase. The third step of endotoxin biosynthesis. *J Biol Chem,* 268**,** 19866-74.

KEYAMURA, K., FUJIKAWA, N., ISHIDA, T., OZAKI, S., SU'ETSUGU, M., FUJIMITSU, K., KAGAWA, W., YOKOYAMA, S., KURUMIZAKA, H. & KATAYAMA, T. (2007) The interaction of DiaA and DnaA regulates the replication cycle in E. coli by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev,* 21**,** 2083-99.

KIDO, N., TORGOV, V. I., SUGIYAMA, T., UCHIYA, K., SUGIHARA, H., KOMATSU, T., KATO, N. & JANN, K. (1995) Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol,* 177**,** 2178-87.

KIM, K. H., AULAKH, S. & PAETZEL, M. (2011a) Crystal structure of beta-barrel assembly machinery BamCD protein complex. *J Biol Chem,* 286**,** 39116-21.

KIM, K. H., AULAKH, S. & PAETZEL, M. (2012) The bacterial outer membrane beta-barrel assembly machinery. *Protein Sci,* 21**,** 751-68.

KIM, K. H., AULAKH, S., TAN, W. & PAETZEL, M. (2011b) Crystallographic analysis of the C-terminal domain of the *Escherichia coli* lipoprotein BamC. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 67**,** 1350-8.

KIM, K. H., KANG, H. S., OKON, M., ESCOBAR-CABRERA, E., MCINTOSH, L. P. & PAETZEL, M. (2011c) Structural characterization of *Escherichia coli* BamE, a lipoprotein component of the beta-barrel assembly machinery complex. *Biochemistry,* 50**,** 1081-90.

KIM, K. H. & PAETZEL, M. (2010) Crystal structure of *Escherichia coli* BamB, a lipoprotein component of the beta-barrel assembly machinery complex. *J Mol Biol,* 406**,** 667-78.
KIM, S., MALINVERNI, J. C., SLIZ, P., SILHAVY, T. J., HARRISON, S. C. & KAHNE, D. (2007) Structure and function of an essential component of the outer membrane protein assembly machine. *Science,* 317**,** 961-4.
KIMURA, S. & SUZUKI, T. (2009) Fine-tuning of the ribosomal decoding center by conserved methyl-modifications in the *Escherichia coli* 16S rRNA. *Nucleic Acids Res,* 38**,** 1341-52.
KINGMA, R. L., FRAGIATHAKI, M., SNIJDER, H. J., DIJKSTRA, B. W., VERHEIJ, H. M., DEKKER, N. & EGMOND, M. R. (2000) Unusual catalytic triad of *Escherichia coli* outer membrane phospholipase A. *Biochemistry,* 39**,** 10017-22.
KINGMA, R. L., SNIJDER, H. J., DIJKSTRA, B. W., DEKKER, N. & EGMOND, M. R. (2002) Functional importance of calcium binding sites in outer membrane phospholipase A. *Biochim Biophys Acta,* 1561**,** 230-7.
KLEIN, J. R., HENRICH, B. & PLAPP, R. (1991) Molecular analysis and nucleotide sequence of the *envCD* operon of *Escherichia coli*. *Mol Gen Genet,* 230**,** 230-40.
KLEINSCHMIDT, J. H. (2003) Membrane protein folding on the example of outer membrane protein A of *Escherichia coli*. *Cell Mol Life Sci,* 60**,** 1547-58.
KLIMKE, W., AGARWALA, R., BADRETDIN, A., CHETVERNIN, S., CIUFO, S., FEDOROV, B., KIRYUTIN, B., O'NEILL, K., RESCH, W., RESENCHUK, S., SCHAFER, S., TOLSTOY, I. & TATUSOVA, T. (2009) The National Center for Biotechnology Information's Protein Clusters Database. *Nucleic Acids Res,* 37**,** D216-23.
KNIREL' IU, A. & KOCHETKOV, N. K. (1994) [Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides]. *Biokhimiia,* 59**,** 1784-851.
KNOWLES, T. J., BROWNING, D. F., JEEVES, M., MADERBOCUS, R., RAJESH, S., SRIDHAR, P., MANOLI, E., EMERY, D., SOMMER, U., SPENCER, A., LEYTON, D. L., SQUIRE, D., CHAUDHURI, R. R., VIANT, M. R., CUNNINGHAM, A. F., HENDERSON, I. R. & OVERDUIN, M. (2011) Structure and function of BamE within the outer membrane and the beta-barrel assembly machine. *EMBO Rep,* 12**,** 123-8.
KNOWLES, T. J., JEEVES, M., BOBAT, S., DANCEA, F., MCCLELLAND, D., PALMER, T., OVERDUIN, M. & HENDERSON, I. R. (2008) Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Mol Microbiol,* 68**,** 1216-27.
KNOWLES, T. J., MCCLELLAND, D. M., RAJESH, S., HENDERSON, I. R. & OVERDUIN, M. (2009) Secondary structure and (1)H, (13)C and (15)N backbone resonance assignments of BamC, a component of the outer membrane protein assembly machinery in *Escherichia coli*. *Biomol NMR Assign,* 3**,** 203-6.
KOEBNIK, R. (1996) In vivo membrane assembly of split variants of the *E.coli* outer membrane protein OmpA. *EMBO J,* 15**,** 3529-37.
KOL, M. A., DE KROON, A. I., KILLIAN, J. A. & DE KRUIJFF, B. (2004) Transbilayer movement of phospholipids in biogenic membranes. *Biochemistry,* 43**,** 2673-81.
KOL, M. A., VAN DALEN, A., DE KROON, A. I. & DE KRUIJFF, B. (2003) Translocation of phospholipids is facilitated by a subset of membrane-spanning proteins of the bacterial cytoplasmic membrane. *J Biol Chem,* 278**,** 24586-93.
KOL, S., MAJCZAK, W., HEERLIEN, R., VAN DER BERG, J. P., NOUWEN, N. & DRIESSEN, A. J. (2009) Subunit a of the F(1)F(0) ATP synthase requires YidC and SecYEG for membrane insertion. *J Mol Biol,* 390**,** 893-901.
KORADI, R., BILLETER, M. & WUTHRICH, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph,* 14**,** 51-5, 29-32.
KOREA, C. G., BADOURALY, R., PREVOST, M. C., GHIGO, J. M. & BELOIN, C. (2010) *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. *Environ Microbiol,* 12**,** 1957-77.

KORNDORFER, I. P., DOMMEL, M. K. & SKERRA, A. (2004) Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. *Nat Struct Mol Biol,* 11**,** 1015-20.

KORONAKIS, V., SHARFF, A., KORONAKIS, E., LUISI, B. & HUGHES, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature,* 405**,** 914-9.

KRAMER, R. A., ZANDWIJKEN, D., EGMOND, M. R. & DEKKER, N. (2000) In vitro folding, purification and characterization of *Escherichia coli* outer membrane protease ompT. *Eur J Biochem,* 267**,** 885-93.

KRIECHBAUMER, V., VON LOFFELHOLZ, O. & ABELL, B. M. (2011) Chaperone receptors: guiding proteins to intracellular compartments. *Protoplasma,* 249**,** 21-30.

KROJER, T., SAWA, J., HUBER, R. & CLAUSEN, T. (2010) HtrA proteases have a conserved activation mechanism that can be triggered by distinct molecular cues. *Nat Struct Mol Biol,* 17**,** 844-52.

KROJER, T., SAWA, J., SCHAFER, E., SAIBIL, H. R., EHRMANN, M. & CLAUSEN, T. (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature,* 453**,** 885-90.

KUNTUMALLA, S., ZHANG, Q., BRAISTED, J. C., FLEISCHMANN, R. D., PETERSON, S. N., DONOHUE-ROLFE, A., TZIPORI, S. & PIEPER, R. (2011) In vivo versus in vitro protein abundance analysis of *Shigella dysenteriae* type 1 reveals changes in the expression of proteins involved in virulence, stress and energy metabolism. *BMC Microbiol,* 11**,** 147.

KUTIK, S., STOJANOVSKI, D., BECKER, L., BECKER, T., MEINECKE, M., KRUGER, V., PRINZ, C., MEISINGER, C., GUIARD, B., WAGNER, R., PFANNER, N. & WIEDEMANN, N. (2008) Dissecting membrane insertion of mitochondrial beta-barrel proteins. *Cell,* 132**,** 1011-24.

LAUBACHER, M. E. & ADES, S. E. (2008) The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. *J Bacteriol,* 190**,** 2065-74.

LAZAR, S. W. & KOLTER, R. (1996) SurA assists the folding of *Escherichia coli* outer membrane proteins. *J Bacteriol,* 178**,** 1770-3.

LEE, H. C. & BERNSTEIN, H. D. (2001) The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A,* 98**,** 3471-6.

LEHR, U., SCHUTZ, M., OBERHETTINGER, P., RUIZ-PEREZ, F., DONALD, J. W., PALMER, T., LINKE, D., HENDERSON, I. R. & AUTENRIETH, I. B. (2010) C-terminal amino acid residues of the trimeric autotransporter adhesin YadA of *Yersinia enterocolitica* are decisive for its recognition and assembly by BamA. *Mol Microbiol,* 78**,** 932-46.

LEIVE, L. (1965) Release of lipopolysaccharide by EDTA treatment of E. coli. *Biochem Biophys Res Commun,* 21**,** 290-6.

LEWENZA, S., VIDAL-INGIGLIARDI, D. & PUGSLEY, A. P. (2006) Direct visualization of red fluorescent lipoproteins indicates conservation of the membrane sorting rules in the family Enterobacteriaceae. *J Bacteriol,* 188**,** 3516-24.

LEWIS, C., SKOVIEROVA, H., ROWLEY, G., REZUCHOVA, B., HOMEROVA, D., STEVENSON, A., SHERRY, A., KORMANEC, J. & ROBERTS, M. (2008) Small outer-membrane lipoprotein, SmpA, is regulated by sigmaE and has a role in cell envelope integrity and virulence of *Salmonella enterica* serovar Typhimurium. *Microbiology,* 154**,** 979-88.

LINTON, K. J. & HIGGINS, C. F. (1998) The *Escherichia coli* ATP-binding cassette (ABC) proteins. *Mol Microbiol,* 28**,** 5-13.

LINTON, K. J. & HIGGINS, C. F. (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. *Pflugers Arch,* 453**,** 555-67.

LIU, D., COLE, R. A. & REEVES, P. R. (1996) An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol,* 178**,** 2102-7.

LIU, D., HAASE, A. M., LINDQVIST, L., LINDBERG, A. A. & REEVES, P. R. (1993) Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. *J Bacteriol,* 175**,** 3408-13.

LOPEZ-CAMPISTROUS, A., SEMCHUK, P., BURKE, L., PALMER-STONE, T., BROKX, S. J., BRODERICK, G., BOTTORFF, D., BOLCH, S., WEINER, J. H. & ELLISON, M. J. (2005) Localization, annotation, and comparison of the *Escherichia coli* K-12 proteome under two states of growth. *Mol Cell Proteomics,* 4**,** 1205-9.

LUGTENBERG, E. J. & PETERS, R. (1976) Distribution of lipids in cytoplasmic and outer membranes of *Escherichia coli* K12. *Biochim Biophys Acta,* 441**,** 38-47.

MA, B., REYNOLDS, C. M. & RAETZ, C. R. (2008) Periplasmic orientation of nascent lipid A in the inner membrane of an *Escherichia coli* LptA mutant. *Proc Natl Acad Sci U S A,* 105**,** 13823-8.

MALINVERNI, J. C. & SILHAVY, T. J. (2009) An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. *Proc Natl Acad Sci U S A,* 106**,** 8009-14.

MALINVERNI, J. C., WERNER, J., KIM, S., SKLAR, J. G., KAHNE, D., MISRA, R. & SILHAVY, T. J. (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol Microbiol,* 61**,** 151-64.

MASUDA, K., MATSUYAMA, S. & TOKUDA, H. (2002) Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proc Natl Acad Sci U S A,* 99**,** 7390-5.

MATERN, Y., BARION, B. & BEHRENS-KNEIP, S. (2010) PpiD is a player in the network of periplasmic chaperones in *Escherichia coli*. *BMC Microbiol,* 10**,** 251.

MATSUYAMA, S., TAJIMA, T. & TOKUDA, H. (1995) A novel periplasmic carrier protein involved in the sorting and transport of *Escherichia coli* lipoproteins destined for the outer membrane. *EMBO J,* 14**,** 3365-72.

MATSUYAMA, S., YOKOTA, N. & TOKUDA, H. (1997) A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of *Escherichia coli*. *EMBO J,* 16**,** 6947-55.

MCGRATH, B. C. & OSBORN, M. J. (1991) Localization of the terminal steps of O-antigen synthesis in *Salmonella typhimurium*. *J Bacteriol,* 173**,** 649-54.

MECSAS, J., ROUVIERE, P. E., ERICKSON, J. W., DONOHUE, T. J. & GROSS, C. A. (1993) The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes Dev,* 7**,** 2618-28.

MEDZHITOV, R. & JANEWAY, C., JR. (2000) Innate immunity. *N Engl J Med,* 343**,** 338-44.

MELTZER, M., HASENBEIN, S., MAMANT, N., MERDANOVIC, M., POEPSEL, S., HAUSKE, P., KAISER, M., HUBER, R., KROJER, T., CLAUSEN, T. & EHRMANN, M. (2009) Structure, function and regulation of the conserved serine proteases DegP and DegS of *Escherichia coli*. *Res Microbiol,* 160**,** 660-6.

MERTEN, J. A., SCHULTZ, K. M. & KLUG, C. S. (2012) Concentration-dependent oligomerization and oligomeric arrangement of LptA. *Protein Sci,* 21**,** 211-8.

METZLER, W. J., CONSTANTINE, K. L., FRIEDRICHS, M. S., BELL, A. J., ERNST, E. G., LAVOIE, T. B. & MUELLER, L. (1993) Characterization of the three-dimensional solution structure of human profilin: 1H, 13C, and 15N NMR assignments and global folding pattern. *Biochemistry,* 32**,** 13818-29.

MICHEL, G. P. & STARKA, J. (1979) Phospholipase A activity with integrated phospholipid vesicles in intact cells of an envelope mutant of *Escherichia coli*. *FEBS Lett,* 108**,** 261-5.

MILEYKOVSKAYA, E. & DOWHAN, W. (1997) The Cpx two-component signal transduction pathway is activated in *Escherichia coli* mutant strains lacking phosphatidylethanolamine. *J Bacteriol,* 179**,** 1029-34.

MISRA, R., CASTILLOKELLER, M. & DENG, M. (2000) Overexpression of protease-deficient DegP(S210A) rescues the lethal phenotype of *Escherichia coli* OmpF assembly mutants in a degP background. *J Bacteriol,* 182**,** 4882-8.

MISSIAKAS, D., BETTON, J. M. & RAINA, S. (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol Microbiol,* 21**,** 871-84.

MIYADAI, H., TANAKA-MASUDA, K., MATSUYAMA, S. & TOKUDA, H. (2004) Effects of lipoprotein overproduction on the induction of DegP (HtrA) involved in quality control in the *Escherichia coli* periplasm. *J Biol Chem,* 279**,** 39807-13.

MIYAMOTO, A., MATSUYAMA, S. & TOKUDA, H. (2001) Mutant of LolA, a lipoprotein-specific molecular chaperone of *Escherichia coli*, defective in the transfer of lipoproteins to LolB. *Biochem Biophys Res Commun,* 287**,** 1125-8.

MIYAMOTO, S. & TOKUDA, H. (2007) Diverse effects of phospholipids on lipoprotein sorting and ATP hydrolysis by the ABC transporter LolCDE complex. *Biochim Biophys Acta,* 1768**,** 1848-54.

MOHAN, S. & RAETZ, C. R. (1994) Endotoxin biosynthesis in *Pseudomonas aeruginosa*: enzymatic incorporation of laurate before 3-deoxy-D-manno-octulosonate. *J Bacteriol,* 176**,** 6944-51.

MOHN, W. W., VAN DER GEIZE, R., STEWART, G. R., OKAMOTO, S., LIU, J., DIJKHUIZEN, L. & ELTIS, L. D. (2008) The actinobacterial *mce4* locus encodes a steroid transporter. *J Biol Chem,* 283**,** 35368-74.

MORO, A., RUIZ-CABELLO, F., FERNANDEZ-CANO, A., STOCK, R. P. & GONZALEZ, A. (1995) Secretion by *Trypanosoma cruzi* of a peptidyl-prolyl cis-trans isomerase involved in cell infection. *EMBO J,* 14**,** 2483-90.

MORONA, R., VAN DEN BOSCH, L. & DANIELS, C. (2000) Evaluation of Wzz/MPA1/MPA2 proteins based on the presence of coiled-coil regions. *Microbiology,* 146 (Pt 1)**,** 1-4.

MULLER-LOENNIES, S., RUND, S., ERVELA, E., SKURNIK, M. & HOLST, O. (1999) The structure of the carbohydrate backbone of the core-lipid A region of the lipopolysaccharide from a clinical isolate of *Yersinia enterocolitica* O:9. *Eur J Biochem,* 261**,** 19-24.

NAGAKUBO, S., NISHINO, K., HIRATA, T. & YAMAGUCHI, A. (2002) The putative response regulator BaeR stimulates multidrug resistance of *Escherichia coli* via a novel multidrug exporter system, MdtABC. *J Bacteriol,* 184**,** 4161-7.

NARITA, S., MATSUYAMA, S. & TOKUDA, H. (2004) Lipoprotein trafficking in *Escherichia coli*. *Arch Microbiol,* 182**,** 1-6.

NARITA, S., TANAKA, K., MATSUYAMA, S. & TOKUDA, H. (2002) Disruption of *lolCDE*, encoding an ATP-binding cassette transporter, is lethal for *Escherichia coli* and prevents release of lipoproteins from the inner membrane. *J Bacteriol,* 184**,** 1417-22.

NARITA, S. & TOKUDA, H. (2007) Amino acids at positions 3 and 4 determine the membrane specificity of *Pseudomonas aeruginosa* lipoproteins. *J Biol Chem,* 282**,** 13372-8.

NARITA, S. & TOKUDA, H. (2009) Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS Lett,* 583**,** 2160-4.

NAVEED, H., JACKUPS, R., JR. & LIANG, J. (2009) Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins. *Proc Natl Acad Sci U S A,* 106**,** 12735-40.

NICHOLS, R. J., SEN, S., CHOO, Y. J., BELTRAO, P., ZIETEK, M., CHABA, R., LEE, S., KAZMIERCZAK, K. M., LEE, K. J., WONG, A., SHALES, M., LOVETT, S., WINKLER, M. E., KROGAN, N. J., TYPAS, A. & GROSS, C. A. (2011) Phenotypic landscape of a bacterial cell. *Cell,* 144**,** 143-56.

NIKAIDO, H. (1998) Multiple antibiotic resistance and efflux. *Curr Opin Microbiol,* 1**,** 516-23.

NIKAIDO, H. (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev,* 67**,** 593-656.

NIKAIDO, H. (2005) Restoring permeability barrier function to outer membrane. *Chem Biol,* 12**,** 507-9.
NOINAJ, N., FAIRMAN, J. W. & BUCHANAN, S. K. (2011) The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex. *J Mol Biol,* 407**,** 248-60.
NOUWEN, N., VAN DER LAAN, M. & DRIESSEN, A. J. (2001) SecDFyajC is not required for the maintenance of the proton motive force. *FEBS Lett,* 508**,** 103-6.
OKUDA, S. & TOKUDA, H. (2009) Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proc Natl Acad Sci U S A,* 106**,** 5877-82.
OKUDA, S., WATANABE, S. & TOKUDA, H. (2008) A short helix in the C-terminal region of LolA is important for the specific membrane localization of lipoproteins. *FEBS Lett,* 582**,** 2247-51.
ONUFRYK, C., CROUCH, M. L., FANG, F. C. & GROSS, C. A. (2005) Characterization of six lipoproteins in the sigmaE regulon. *J Bacteriol,* 187**,** 4552-61.
OSBORN, M. J., GANDER, J. E., PARISI, E. & CARSON, J. (1972) Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Isolation and characterization of cytoplasmic and outer membrane. *J Biol Chem,* 247**,** 3962-72.
OSBORN, M. J., RICK, P. D., LEHMANN, V., RUPPRECHT, E. & SINGH, M. (1974) Structure and biogenesis of the cell envelope of gram-negative bacteria. *Ann N Y Acad Sci,* 235**,** 52-65.
PAETZEL, M., DALBEY, R. E. & STRYNADKA, N. C. (2002) Crystal structure of a bacterial signal peptidase apoenzyme: implications for signal peptide binding and the Ser-Lys dyad mechanism. *J Biol Chem,* 277**,** 9512-9.
PANDEY, A. K. & SASSETTI, C. M. (2008) Mycobacterial persistence requires the utilization of host cholesterol. *Proc Natl Acad Sci U S A,* 105**,** 4376-80.
PARADIS-BLEAU, C., MARKOVSKI, M., UEHARA, T., LUPOLI, T. J., WALKER, S., KAHNE, D. E. & BERNHARDT, T. G. (2010) Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell,* 143**,** 1110-20.
PARILLO, V. L. (1993) Systems analysis of an occupational health department: recommendations to increase effectiveness. *AAOHN J,* 41**,** 220-7.
PASTUKHOV, A. V. & ROPSON, I. J. (2003) Fluorescent dyes as probes to study lipid-binding proteins. *Proteins,* 53**,** 607-15.
PATRIARCA, P., BECKERDITE, S. & ELSBACH, P. (1972) Phospholipases and phospholipid turnover in *Escherichia coli* spheroplasts. *Biochim Biophys Acta,* 260**,** 593-600.
PHIZICKY, E. M. & FIELDS, S. (1995) Protein-protein interactions: methods for detection and analysis. *Microbiol Rev,* 59**,** 94-123.
POGLIANO, K., HOFMEISTER, A. E. & LOSICK, R. (1997) Disappearance of the sigma E transcription factor from the forespore and the SpoIIE phosphatase from the mother cell contributes to establishment of cell-specific gene expression during sporulation in Bacillus subtilis. *J Bacteriol,* 179**,** 3331-41.
POLTORAK, A., HE, X., SMIRNOVA, I., LIU, M. Y., VAN HUFFEL, C., DU, X., BIRDWELL, D., ALEJOS, E., SILVA, M., GALANOS, C., FREUDENBERG, M., RICCIARDI-CASTAGNOLI, P., LAYTON, B. & BEUTLER, B. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science,* 282**,** 2085-8.
POOL, M. R. (2009) A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J Cell Biol,* 185**,** 889-902.
POPOFF, M. Y. & LE MINOR, L. (1985) Expression of antigenic factor O:54 is associated with the presence of a plasmid in *Salmonella*. *Ann Inst Pasteur Microbiol,* 136B**,** 169-79.
PORITZ, M. A., BERNSTEIN, H. D., STRUB, K., ZOPF, D., WILHELM, H. & WALTER, P. (1990) An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science,* 250**,** 1111-7.

PUGSLEY, A. P. & SCHWARTZ, M. (1984) Colicin E2 release: lysis, leakage or secretion? Possible role of a phospholipase. *EMBO J,* 3**,** 2393-7.
PURDY, G. E., FISHER, C. R. & PAYNE, S. M. (2007) IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. *J Bacteriol,* 189**,** 5566-73.
QU, J., BEHRENS-KNEIP, S., HOLST, O. & KLEINSCHMIDT, J. H. (2009) Binding regions of outer membrane protein A in complexes with the periplasmic chaperone Skp. A site-directed fluorescence study. *Biochemistry,* 48**,** 4926-36.
QU, J., MAYER, C., BEHRENS, S., HOLST, O. & KLEINSCHMIDT, J. H. (2007) The trimeric periplasmic chaperone Skp of *Escherichia coli* forms 1:1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions. *J Mol Biol,* 374**,** 91-105.
RAETZ, C. R. (1990) Biochemistry of endotoxins. *Annu Rev Biochem,* 59**,** 129-70.
RAETZ, C. R. (1993) Bacterial endotoxins: extraordinary lipids that activate eucaryotic signal transduction. *J Bacteriol,* 175**,** 5745-53.
RAETZ, C. R. & WHITFIELD, C. (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem,* 71**,** 635-700.
RAFFA, R. G. & RAIVIO, T. L. (2002) A third envelope stress signal transduction pathway in *Escherichia coli*. *Mol Microbiol,* 45**,** 1599-611.
RAIVIO, T. L., POPKIN, D. L. & SILHAVY, T. J. (1999) The Cpx envelope stress response is controlled by amplification and feedback inhibition. *J Bacteriol,* 181**,** 5263-72.
RAIVIO, T. L. & SILHAVY, T. J. (1997) Transduction of envelope stress in *Escherichia coli* by the Cpx two-component system. *J Bacteriol,* 179**,** 7724-33.
RAIVIO, T. L. & SILHAVY, T. J. (2001) Periplasmic stress and ECF sigma factors. *Annu Rev Microbiol,* 55**,** 591-624.
RAMESH, G., ALVAREZ, A. L., ROBERTS, E. D., DENNIS, V. A., LASATER, B. L., ALVAREZ, X. & PHILIPP, M. T. (2003) Pathogenesis of Lyme neuroborreliosis: *Borrelia burgdorferi* lipoproteins induce both proliferation and apoptosis in rhesus monkey astrocytes. *Eur J Immunol,* 33**,** 2539-50.
RAMM, K. & PLUCKTHUN, A. (2000) The periplasmic *Escherichia coli* peptidylprolyl cis,trans-isomerase FkpA. II. Isomerase-independent chaperone activity in vitro. *J Biol Chem,* 275**,** 17106-13.
RAY, B. L. & RAETZ, C. R. (1987) The biosynthesis of gram-negative endotoxin. A novel kinase in *Escherichia coli* membranes that incorporates the 4'-phosphate of lipid A. *J Biol Chem,* 262**,** 1122-8.
REMAUT, H. & WAKSMAN, G. (2006) Protein-protein interaction through beta-strand addition. *Trends Biochem Sci,* 31**,** 436-44.
RIBES, V., ROMISCH, K., GINER, A., DOBBERSTEIN, B. & TOLLERVEY, D. (1990) *E. coli* 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle. *Cell,* 63**,** 591-600.
RICK, P. D., HUBBARD, G. L. & BARR, K. (1994) Role of the *rfe* gene in the synthesis of the O8 antigen in *Escherichia coli* K-12. *J Bacteriol,* 176**,** 2877-84.
RIGEL, N. W., SCHWALM, J., RICCI, D. P. & SILHAVY, T. J. (2011) BamE modulates the *Escherichia coli* beta-barrel assembly machine component BamA. *J Bacteriol,* 194**,** 1002-8.
RIZZITELLO, A. E., HARPER, J. R. & SILHAVY, T. J. (2001) Genetic evidence for parallel pathways of chaperone activity in the periplasm of *Escherichia coli*. *J Bacteriol,* 183**,** 6794-800.
ROBERT, V., VOLOKHINA, E. B., SENF, F., BOS, M. P., VAN GELDER, P. & TOMMASSEN, J. (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol,* 4**,** e377.
ROUVIERE, P. E. & GROSS, C. A. (1996) SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev,* 10**,** 3170-82.

ROWLEY, G., SKOVIEROVA, H., STEVENSON, A., REZUCHOVA, B., HOMEROVA, D., LEWIS, C., SHERRY, A., KORMANEC, J. & ROBERTS, M. (2010) The periplasmic chaperone Skp is required for successful *Salmonella Typhimurium* infection in a murine typhoid model. *Microbiology,* 157**,** 848-58.

ROY, A., KUCUKURAL, A. & ZHANG, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc,* 5**,** 725-38.

RUIZ-PEREZ, F., HENDERSON, I. R., LEYTON, D. L., ROSSITER, A. E., ZHANG, Y. & NATARO, J. P. (2009) Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of Enterobacteriaceae. *J Bacteriol,* 191**,** 6571-83.

RUIZ-PEREZ, F., HENDERSON, I. R. & NATARO, J. P. (2011) Interaction of FkpA, a peptidyl-prolyl cis/trans isomerase with EspP autotransporter protein. *Gut Microbes,* 1**,** 339-344.

RUIZ, N., FALCONE, B., KAHNE, D. & SILHAVY, T. J. (2005) Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell,* 121**,** 307-17.

RUIZ, N., GRONENBERG, L. S., KAHNE, D. & SILHAVY, T. J. (2008) Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 105**,** 5537-42.

RYAN, K. R., TAYLOR, J. A. & BOWERS, L. M. (2009) The BAM complex subunit BamE (SmpA) is required for membrane integrity, stalk growth and normal levels of outer membrane {beta}-barrel proteins in *Caulobacter crescentus*. *Microbiology,* 156**,** 742-56.

SAKAMOTO, C., SATOU, R., TOKUDA, H. & NARITA, S. (2010) Novel mutations of the LolCDE complex causing outer membrane localization of lipoproteins despite their inner membrane-retention signals. *Biochem Biophys Res Commun,* 401**,** 586-91.

SANCHEZ-PULIDO, L., DEVOS, D., GENEVROIS, S., VICENTE, M. & VALENCIA, A. (2003) POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem Sci,* 28**,** 523-6.

SANDOVAL, C. M., BAKER, S. L., JANSEN, K., METZNER, S. I. & SOUSA, M. C. (2011) Crystal structure of BamD: an essential component of the beta-Barrel assembly machinery of gram-negative bacteria. *J Mol Biol,* 409**,** 348-57.

SANKARAN, K. & WU, H. C. (1994) Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem,* 269**,** 19701-6.

SARSERO, J. P. & PITTARD, A. J. (1995) Membrane topology analysis of *Escherichia coli* K-12 Mtr permease by alkaline phosphatase and beta-galactosidase fusions. *J Bacteriol,* 177**,** 297-306.

SAUL, F. A., ARIE, J. P., VULLIEZ-LE NORMAND, B., KAHN, R., BETTON, J. M. & BENTLEY, G. A. (2004) Structural and functional studies of FkpA from *Escherichia coli*, a cis/trans peptidyl-prolyl isomerase with chaperone activity. *J Mol Biol,* 335**,** 595-608.

SAURI, A., SOPROVA, Z., WICKSTROM, D., DE GIER, J. W., VAN DER SCHORS, R. C., SMIT, A. B., JONG, W. S. & LUIRINK, J. (2009) The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology,* 155**,** 3982-91.

SAXENA, I. M., BROWN, R. M., JR., FEVRE, M., GEREMIA, R. A. & HENRISSAT, B. (1995) Multidomain architecture of beta-glycosyl transferases: implications for mechanism of action. *J Bacteriol,* 177**,** 1419-24.

SCANDELLA, C. J. & KORNBERG, A. (1971) A membrane-bound phospholipase A1 purified from *Escherichia coli*. *Biochemistry,* 10**,** 4447-56.

SCHAFER, U., BECK, K. & MULLER, M. (1999) Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J Biol Chem,* 274**,** 24567-74.

SCHIEBEL, E., DRIESSEN, A. J., HARTL, F. U. & WICKNER, W. (1991) Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell,* 64**,** 927-39.

SCHLEIFF, E., MAIER, U. G. & BECKER, T. (2011) Omp85 in eukaryotic systems: one protein family with distinct functions. *Biol Chem,* 392**,** 21-7.

SCHULZE, R. J. & ZUCKERT, W. R. (2006) *Borrelia burgdorferi* lipoproteins are secreted to the outer surface by default. *Mol Microbiol,* 59**,** 1473-84.
SCRAGG, I. G., KWIATKOWSKI, D., VIDAL, V., REASON, A., PAXTON, T., PANICO, M., DELL, A. & MORRIS, H. (2000) Structural characterization of the inflammatory moiety of a variable major lipoprotein of *Borrelia recurrentis*. *J Biol Chem,* 275**,** 937-41.
SEIFFER, D., KLEIN, J. R. & PLAPP, R. (1993) EnvC, a new lipoprotein of the cytoplasmic membrane of *Escherichia coli*. *FEMS Microbiol Lett,* 107**,** 175-8.
SESHADRI, K., GAREMYR, R., WALLIN, E., VON HEIJNE, G. & ELOFSSON, A. (1998) Architecture of beta-barrel membrane proteins: analysis of trimeric porins. *Protein Sci,* 7**,** 2026-32.
SEYDEL, A., GOUNON, P. & PUGSLEY, A. P. (1999) Testing the '+2 rule' for lipoprotein sorting in the *Escherichia coli* cell envelope with a new genetic selection. *Mol Microbiol,* 34**,** 810-21.
SILHAVY, T. J., KAHNE, D. & WALKER, S. (2010) The bacterial cell envelope. *Cold Spring Harb Perspect Biol,* 2**,** a000414.
SKLAR, J. G., WU, T., GRONENBERG, L. S., MALINVERNI, J. C., KAHNE, D. & SILHAVY, T. J. (2007a) Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in *Escherichia coli*. *Proc Natl Acad Sci U S A,* 104**,** 6400-5.
SKLAR, J. G., WU, T., KAHNE, D. & SILHAVY, T. J. (2007b) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev,* 21**,** 2473-84.
SKOVIEROVA, H., REZUCHOVA, B., HOMEROVA, D., ROBERTS, M. & KORMANEC, J. (2006) Characterization of the sigmaE-dependent rpoEp3 promoter of *Salmonella enterica* serovar Typhimurium. *FEMS Microbiol Lett,* 261**,** 53-9.
SMITH, P. C., KARPOWICH, N., MILLEN, L., MOODY, J. E., ROSEN, J., THOMAS, P. J. & HUNT, J. F. (2002) ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell,* 10**,** 139-49.
SNIJDER, H. J., KINGMA, R. L., KALK, K. H., DEKKER, N., EGMOND, M. R. & DIJKSTRA, B. W. (2001) Structural investigations of calcium binding and its role in activity and activation of outer membrane phospholipase A from *Escherichia coli*. *J Mol Biol,* 309**,** 477-89.
SNIJDER, H. J., UBARRETXENA-BELANDIA, I., BLAAUW, M., KALK, K. H., VERHEIJ, H. M., EGMOND, M. R., DEKKER, N. & DIJKSTRA, B. W. (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature,* 401**,** 717-21.
SNYDER, W. B., DAVIS, L. J., DANESE, P. N., COSMA, C. L. & SILHAVY, T. J. (1995) Overproduction of NlpE, a new outer membrane lipoprotein, suppresses the toxicity of periplasmic LacZ by activation of the Cpx signal transduction pathway. *J Bacteriol,* 177**,** 4216-23.
SORENSEN, P. G., LUTKENHAUS, J., YOUNG, K., EVELAND, S. S., ANDERSON, M. S. & RAETZ, C. R. (1996) Regulation of UDP-3-O-[R-3-hydroxymyristoyl]-N-acetylglucosamine deacetylase in *Escherichia coli*. The second enzymatic step of lipid a biosynthesis. *J Biol Chem,* 271**,** 25898-905.
SOUTHAM, A. D., PAYNE, T. G., COOPER, H. J., ARVANITIS, T. N. & VIANT, M. R. (2007) Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal Chem,* 79**,** 4595-602.
SPERANDEO, P., CESCUTTI, R., VILLA, R., DI BENEDETTO, C., CANDIA, D., DEHO, G. & POLISSI, A. (2007) Characterization of *lptA* and *lptB*, two essential genes implicated in lipopolysaccharide transport to the outer membrane of *Escherichia coli*. *J Bacteriol,* 189**,** 244-53.
SPERANDEO, P., DEHO, G. & POLISSI, A. (2009) The lipopolysaccharide transport system of Gram-negative bacteria. *Biochim Biophys Acta,* 1791**,** 594-602.
SPERANDEO, P., LAU, F. K., CARPENTIERI, A., DE CASTRO, C., MOLINARO, A., DEHO, G., SILHAVY, T. J. & POLISSI, A. (2008) Functional analysis of the protein machinery required for

transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *J Bacteriol,* 190**,** 4460-9.
SPERANDEO, P., POZZI, C., DEHO, G. & POLISSI, A. (2006) Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the *Escherichia coli yrbG-yhbG* locus. *Res Microbiol,* 157**,** 547-58.
SPERANDEO, P., VILLA, R., MARTORANA, A. M., SAMALIKOVA, M., GRANDORI, R., DEHO, G. & POLISSI, A. (2011) New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *J Bacteriol,* 193**,** 1042-53.
STANLEY, A. M., CHUAWONG, P., HENDRICKSON, T. L. & FLEMING, K. G. (2006) Energetics of outer membrane phospholipase A (OMPLA) dimerization. *J Mol Biol,* 358**,** 120-31.
STANLEY, A. M., TREUBRODT, A. M., CHUAWONG, P., HENDRICKSON, T. L. & FLEMING, K. G. (2007) Lipid chain selectivity by outer membrane phospholipase A. *J Mol Biol,* 366**,** 461-8.
STEGMEIER, J. F. & ANDERSEN, C. (2006) Characterization of pores formed by YaeT (Omp85) from *Escherichia coli*. *J Biochem,* 140**,** 275-83.
STRUYVE, M., MOONS, M. & TOMMASSEN, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J Mol Biol,* 218**,** 141-8.
STYMEST, K. H. & KLAPPA, P. (2008) The periplasmic peptidyl prolyl cis-trans isomerases PpiD and SurA have partially overlapping substrate specificities. *FEBS J,* 275**,** 3470-9.
SUBRINI, O. & BETTON, J. M. (2009) Assemblies of DegP underlie its dual chaperone and protease function. *FEMS Microbiol Lett,* 296**,** 143-8.
SUITS, M. D., SPERANDEO, P., DEHO, G., POLISSI, A. & JIA, Z. (2008) Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J Mol Biol,* 380**,** 476-88.
SUZUKI, T., MURAI, T., FUKUDA, I., TOBE, T., YOSHIKAWA, M. & SASAKAWA, C. (1994) Identification and characterization of a chromosomal virulence gene, *vacJ*, required for intercellular spreading of *Shigella flexneri*. *Mol Microbiol,* 11**,** 31-41.
TAKASE, I., ISHINO, F., WACHI, M., KAMATA, H., DOI, M., ASOH, S., MATSUZAWA, H., OHTA, T. & MATSUHASHI, M. (1987) Genes encoding two lipoproteins in the *leuS-dacA* region of the *Escherichia coli* chromosome. *J Bacteriol,* 169**,** 5692-9.
TAKEDA, K., MIYATAKE, H., YOKOTA, N., MATSUYAMA, S., TOKUDA, H. & MIKI, K. (2003a) Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO J,* 22**,** 3199-209.
TAKEDA, K., MIYATAKE, H., YOKOTA, N., MATSUYAMA, S. I., TOKUDA, H. & MIKI, K. (2003b) Crystallization and preliminary crystallographic study of the outer-membrane lipoprotein receptor LolB, a member of the lipoprotein localization factors. *Acta Crystallogr D Biol Crystallogr,* 59**,** 1224-6.
TAKETO, A. (1974) Sensitivity of *Escherichia coli* to viral nucleic acid. 8. Idiosyncrasy of Ca2+-dependent competence for DNA. *J Biochem,* 75**,** 895-904.
TAM, C. & MISSIAKAS, D. (2005) Changes in lipopolysaccharide structure induce the sigma(E)-dependent response of *Escherichia coli*. *Mol Microbiol,* 55**,** 1403-12.
TAMM, L. K., ARORA, A. & KLEINSCHMIDT, J. H. (2001) Structure and assembly of beta-barrel membrane proteins. *J Biol Chem,* 276**,** 32399-402.
TAMM, L. K., HONG, H. & LIANG, B. (2004) Folding and assembly of beta-barrel membrane proteins. *Biochim Biophys Acta,* 1666**,** 250-63.
TANAKA, K., MATSUYAMA, S. I. & TOKUDA, H. (2001) Deletion of *lolB*, encoding an outer membrane lipoprotein, is lethal for *Escherichia coli* and causes accumulation of lipoprotein localization intermediates in the periplasm. *J Bacteriol,* 183**,** 6538-42.

TANIGUCHI, N., MATSUYAMA, S. & TOKUDA, H. (2005) Mechanisms underlying energy-independent transfer of lipoproteins from LolA to LolB, which have similar unclosed {beta}-barrel structures. *J Biol Chem,* 280**,** 34481-8.

TANIGUCHI, N. & TOKUDA, H. (2008) Molecular events involved in a single cycle of ligand transfer from an ATP binding cassette transporter, LolCDE, to a molecular chaperone, LolA. *J Biol Chem,* 283**,** 8538-44.

TAO, K., WATANABE, S., NARITA, S. & TOKUDA, H. (2010) A periplasmic LolA derivative with a lethal disulfide bond activates the Cpx stress response system. *J Bacteriol,* 192**,** 5657-62.

TEFSEN, B., GEURTSEN, J., BECKERS, F., TOMMASSEN, J. & DE COCK, H. (2005) Lipopolysaccharide transport to the bacterial outer membrane in spheroplasts. *J Biol Chem,* 280**,** 4504-9.

TERADA, M., KURODA, T., MATSUYAMA, S. I. & TOKUDA, H. (2001) Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of *Escherichia coli*. *J Biol Chem,* 276**,** 47690-4.

TERIETE, P., YAO, Y., KOLODZIK, A., YU, J., SONG, H., NIEDERWEIS, M. & MARASSI, F. M. (2010) *Mycobacterium tuberculosis* Rv0899 adopts a mixed alpha/beta-structure and does not form a transmembrane beta-barrel. *Biochemistry,* 49**,** 2768-77.

THOME, B. M., HOFFSCHULTE, H. K., SCHILTZ, E. & MULLER, M. (1990) A protein with sequence identity to Skp (FirA) supports protein translocation into plasma membrane vesicles of *Escherichia coli*. *FEBS Lett,* 269**,** 113-6.

THOME, B. M. & MULLER, M. (1991) Skp is a periplasmic *Escherichia coli* protein requiring SecA and SecY for export. *Mol Microbiol,* 5**,** 2815-21.

TOKUDA, H. (2009) Biogenesis of outer membranes in Gram-negative bacteria. *Biosci Biotechnol Biochem,* 73**,** 465-73.

TOKUDA, H. & MATSUYAMA, S. (2004) Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim Biophys Acta,* 1693**,** 5-13.

TOKUDA, H., TAKAI, S., HANAI, Y., HARADA, A., MATSUSHIMA-NISHIWAKI, R., AKAMATSU, S., OHTA, T. & KOZAWA, O. (2007) Platelet-derived growth factor-BB amplifies PGF2alpha-stimulated VEGF synthesis in osteoblasts: function of phosphatidylinositol 3-kinase. *Prostaglandins Leukot Essent Fatty Acids,* 77**,** 187-93.

TOWBIN, H., STAEHELIN, T. & GORDON, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc Natl Acad Sci U S A,* 76**,** 4350-4.

TRAN, A. X., DONG, C. & WHITFIELD, C. (2010) Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in *Escherichia coli*. *J Biol Chem,* 285**,** 33529-39.

TSUKAHARA, J., MUKAIYAMA, K., OKUDA, S., NARITA, S. & TOKUDA, H. (2009a) Dissection of LolB function--lipoprotein binding, membrane targeting and incorporation of lipoproteins into lipid bilayers. *FEBS J,* 276**,** 4496-504.

TSUKAHARA, J., NARITA, S. & TOKUDA, H. (2009b) Real time analysis of lipoprotein transfer from LolA to LolB by means of surface plasmon resonance. *FEBS Lett,* 583**,** 2987-90.

TSUKAZAKI, T., MORI, H., ECHIZEN, Y., ISHITANI, R., FUKAI, S., TANAKA, T., PEREDERINA, A., VASSYLYEV, D. G., KOHNO, T., MATURANA, A. D., ITO, K. & NUREKI, O. (2011) Structure and function of a membrane component SecDF that enhances protein export. *Nature,* 474**,** 235-8.

TYPAS, A., BANZHAF, M., GROSS, C. A. & VOLLMER, W. (2011) From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol,* 10**,** 123-36.

TYPAS, A., BANZHAF, M., VAN DEN BERG VAN SAPAROEA, B., VERHEUL, J., BIBOY, J., NICHOLS, R. J., ZIETEK, M., BEILHARZ, K., KANNENBERG, K., VON RECHENBERG, M., BREUKINK, E.,

DEN BLAAUWEN, T., GROSS, C. A. & VOLLMER, W. (2010) Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell,* 143**,** 1097-109.
UEHARA, T., DINH, T. & BERNHARDT, T. G. (2009) LytM-domain factors are required for daughter cell separation and rapid ampicillin-induced lysis in *Escherichia coli*. *J Bacteriol,* 191**,** 5094-107.
ULBRANDT, N. D., NEWITT, J. A. & BERNSTEIN, H. D. (1997) The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell,* 88**,** 187-96.
URETA, A. R., ENDRES, R. G., WINGREEN, N. S. & SILHAVY, T. J. (2007) Kinetic analysis of the assembly of the outer membrane protein LamB in *Escherichia coli* mutants each lacking a secretion or targeting factor in a different cellular compartment. *J Bacteriol,* 189**,** 446-54.
VALENT, Q. A., KENDALL, D. A., HIGH, S., KUSTERS, R., OUDEGA, B. & LUIRINK, J. (1995) Early events in preprotein recognition in *E. coli*: interaction of SRP and trigger factor with nascent polypeptides. *EMBO J,* 14**,** 5494-505.
VALENT, Q. A., SCOTTI, P. A., HIGH, S., DE GIER, J. W., VON HEIJNE, G., LENTZEN, G., WINTERMEYER, W., OUDEGA, B. & LUIRINK, J. (1998) The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *EMBO J,* 17**,** 2504-12.
VAN BLOOIS, E., HAAN, G. J., DE GIER, J. W., OUDEGA, B. & LUIRINK, J. (2006) Distinct requirements for translocation of the N-tail and C-tail of the *Escherichia coli* inner membrane protein CyoA. *J Biol Chem,* 281**,** 10002-9.
VAN DEUREN, M., BRANDTZAEG, P. & VAN DER MEER, J. W. (2000) Update on meningococcal disease with emphasis on pathogenesis and clinical management. *Clin Microbiol Rev,* 13**,** 144-66, table of contents.
VAZQUEZ-LASLOP, N., LEE, H., HU, R. & NEYFAKH, A. A. (2001) Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked *Escherichia coli*. *J Bacteriol,* 183**,** 2399-404.
VERTOMMEN, D., RUIZ, N., LEVERRIER, P., SILHAVY, T. J. & COLLET, J. F. (2009) Characterization of the role of the *Escherichia coli* periplasmic chaperone SurA using differential proteomics. *Proteomics,* 9**,** 2432-43.
VOLOKHINA, E. B., BECKERS, F., TOMMASSEN, J. & BOS, M. P. (2009) The beta-barrel outer membrane protein assembly complex of *Neisseria meningitidis*. *J Bacteriol,* 191**,** 7074-85.
VOULHOUX, R., BOS, M. P., GEURTSEN, J., MOLS, M. & TOMMASSEN, J. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science,* 299**,** 262-5.
VUONG, P., BENNION, D., MANTEI, J., FROST, D. & MISRA, R. (2008) Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry. *J Bacteriol,* 190**,** 1507-17.
WALSH, A. G., MATEWISH, M. J., BURROWS, L. L., MONTEIRO, M. A., PERRY, M. B. & LAM, J. S. (2000) Lipopolysaccharide core phosphates are required for viability and intrinsic drug resistance in *Pseudomonas aeruginosa*. *Mol Microbiol,* 35**,** 718-27.
WALSH, N. P., ALBA, B. M., BOSE, B., GROSS, C. A. & SAUER, R. T. (2003) OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell,* 113**,** 61-71.
WALTON, T. A., SANDOVAL, C. M., FOWLER, C. A., PARDI, A. & SOUSA, M. C. (2009) The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc Natl Acad Sci U S A,* 106**,** 1772-7.
WALTON, T. A. & SOUSA, M. C. (2004) Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol Cell,* 15**,** 367-74.
WARD, A., REYES, C. L., YU, J., ROTH, C. B. & CHANG, G. (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A,* 104**,** 19005-10.

WARD, R., ZOLTNER, M., BEER, L., EL MKAMI, H., HENDERSON, I. R., PALMER, T. & NORMAN, D. G. (2009) The orientation of a tandem POTRA domain pair, of the beta-barrel assembly protein BamA, determined by PELDOR spectroscopy. *Structure,* 17**,** 1187-94.

WARNER, L. R., VARGA, K., LANGE, O. F., BAKER, S. L., BAKER, D., SOUSA, M. C. & PARDI, A. (2011) Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set. *J Mol Biol,* 411**,** 83-95.

WATTS, K. M. & HUNSTAD, D. A. (2008) Components of SurA required for outer membrane biogenesis in uropathogenic *Escherichia coli*. *PLoS One,* 3**,** e3359.

WEBB, C. T., GORMAN, M. A., LAZAROU, M., RYAN, M. T. & GULBIS, J. M. (2006) Crystal structure of the mitochondrial chaperone TIM9.10 reveals a six-bladed alpha-propeller. *Mol Cell,* 21**,** 123-33.

WEBB, C. T., SELKRIG, J., PERRY, A. J., NOINAJ, N., BUCHANAN, S. K. & LITHGOW, T. (2012) Dynamic Association of BAM Complex Modules Includes Surface Exposure of the Lipoprotein BamC. *J Mol Biol,* 422**,** 545-55.

WEBB, H. M., RUDDOCK, L. W., MARCHANT, R. J., JONAS, K. & KLAPPA, P. (2001) Interaction of the periplasmic peptidylprolyl cis-trans isomerase SurA with model peptides. The N-terminal region of SurA id essential and sufficient for peptide binding. *J Biol Chem,* 276**,** 45622-7.

WEININGER, U., JAKOB, R. P., KOVERMANN, M., BALBACH, J. & SCHMID, F. X. (2009) The prolyl isomerase domain of PpiD from *Escherichia coli* shows a parvulin fold but is devoid of catalytic activity. *Protein Sci,* 19**,** 6-18.

WEISS, J., BECKERDITE-QUAGLIATA, S. & ELSBACH, P. (1979) Determinants of the action of phospholipases A on the envelope phospholipids of *Escherichia coli*. *J Biol Chem,* 254**,** 11010-4.

WHITE, K. A., LIN, S., COTTER, R. J. & RAETZ, C. R. (1999) A *Haemophilus influenzae* gene that encodes a membrane bound 3-deoxy-D-manno-octulosonic acid (Kdo) kinase. Possible involvement of kdo phosphorylation in bacterial virulence. *J Biol Chem,* 274**,** 31391-400.

WIMLEY, W. C. (2003) The versatile beta-barrel membrane protein. *Curr Opin Struct Biol,* 13**,** 404-11.

WISHART, D. S. & SYKES, B. D. (1994) Chemical shifts as a tool for structure determination. *Methods Enzymol,* 239**,** 363-92.

WRIGHT, G. C., WEISS, J., KIM, K. S., VERHEIJ, H. & ELSBACH, P. (1990) Bacterial phospholipid hydrolysis enhances the destruction of *Escherichia coli* ingested by rabbit neutrophils. Role of cellular and extracellular phospholipases. *J Clin Invest,* 85**,** 1925-35.

WU, H., SOUTHAM, A. D., HINES, A. & VIANT, M. R. (2008) High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal Biochem,* 372**,** 204-12.

WU, T., MALINVERNI, J., RUIZ, N., KIM, S., SILHAVY, T. J. & KAHNE, D. (2005) Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell,* 121**,** 235-45.

WU, T., MCCANDLISH, A. C., GRONENBERG, L. S., CHNG, S. S., SILHAVY, T. J. & KAHNE, D. (2006) Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 103**,** 11754-9.

WYCKOFF, T. J., RAETZ, C. R. & JACKMAN, J. E. (1998) Antibacterial and anti-inflammatory agents that target endotoxin. *Trends Microbiol,* 6**,** 154-9.

XU, X., WANG, S., HU, Y. X. & MCKAY, D. B. (2007) The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *J Mol Biol,* 373**,** 367-81.

YAKUSHI, T., MASUDA, K., NARITA, S., MATSUYAMA, S. & TOKUDA, H. (2000) A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nat Cell Biol,* 2**,** 212-8.

YAKUSHI, T., TAJIMA, T., MATSUYAMA, S. & TOKUDA, H. (1997) Lethality of the covalent linkage between mislocalized major outer membrane lipoprotein and the peptidoglycan of *Escherichia coli*. *J Bacteriol,* 179**,** 2857-62.

YAMAGUCHI, K., YU, F. & INOUYE, M. (1988) A single amino acid determinant of the membrane localization of lipoproteins in *E. coli. Cell,* 53**,** 423-32.

YASUDA, M., IGUCHI-YOKOYAMA, A., MATSUYAMA, S., TOKUDA, H. & NARITA, S. (2009) Membrane topology and functional importance of the periplasmic region of ABC transporter LolCDE. *Biosci Biotechnol Biochem,* 73**,** 2310-6.

YEATS, C. & BATEMAN, A. (2003) The BON domain: a putative membrane-binding domain. *Trends Biochem Sci,* 28**,** 352-5.

YEE, A. A., SAVCHENKO, A., IGNACHENKO, A., LUKIN, J., XU, X., SKARINA, T., EVDOKIMOVA, E., LIU, C. S., SEMESI, A., GUIDO, V., EDWARDS, A. M. & ARROWSMITH, C. H. (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J Am Chem Soc,* 127**,** 16512-7.

YETHON, J. A., HEINRICHS, D. E., MONTEIRO, M. A., PERRY, M. B. & WHITFIELD, C. (1998) Involvement of *waaY*, *waaQ*, and *waaP* in the modification of *Escherichia coli* lipopolysaccharide and their role in the formation of a stable outer membrane. *J Biol Chem,* 273**,** 26310-6.

YETHON, J. A., VINOGRADOV, E., PERRY, M. B. & WHITFIELD, C. (2000) Mutation of the lipopolysaccharide core glycosyltransferase encoded by *waaG* destabilizes the outer membrane of *Escherichia coli* by interfering with core phosphorylation. *J Bacteriol,* 182**,** 5620-3.

YIM, H. H. & VILLAREJO, M. (1992) *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. *J Bacteriol,* 174**,** 3637-44.

YOKOTA, N., KURODA, T., MATSUYAMA, S. & TOKUDA, H. (1999) Characterization of the LolA-LolB system as the general lipoprotein localization mechanism of *Escherichia coli*. *J Biol Chem,* 274**,** 30995-9.

YOUNG, K., SILVER, L. L., BRAMHILL, D., CAMERON, P., EVELAND, S. S., RAETZ, C. R., HYLAND, S. A. & ANDERSON, M. S. (1995) The *envA* permeability/cell division gene of *Escherichia coli* encodes the second enzyme of lipid A biosynthesis. UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase. *J Biol Chem,* 270**,** 30384-91.

YUAN, Y. R., BLECKER, S., MARTSINKEVICH, O., MILLEN, L., THOMAS, P. J. & HUNT, J. F. (2001) The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J Biol Chem,* 276**,** 32313-21.

ZHANG, H. & JAVOR, G. T. (2003) Regulation of the isofunctional genes *ubiD* and *ubiX* of the ubiquinone biosynthetic pathway of *Escherichia coli*. *FEMS Microbiol Lett,* 223**,** 67-72.

ZHOU, Z., WHITE, K. A., POLISSI, A., GEORGOPOULOS, C. & RAETZ, C. R. (1998) Function of *Escherichia coli* MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J Biol Chem,* 273**,** 12466-75.

ZHU, Z. X., CONG, W. T., NI, M. W., WANG, X., MA, W. D., YE, W. J., JIN, L. T. & LI, X. K. (2012) An improved silver stain for the visualization of lipopolysaccharides on polyacrylamide gels. *Electrophoresis,* 33**,** 1220-3.

Appendix I BamE publication by Knowles *et al.* 2011

Appendix II YraP assignment data

1	176.507	0.000	C	20
2	62.132	0.165	CA	20
3	39.799	0.080	CB	20
4	14.065	0.003	CD1	20
5	28.183	0.007	CG1	20
6	18.481	0.013	CG2	20
7	8.098	0.009	H	20
8	4.117	0.005	HA	20
9	1.828	0.006	HB	20
10	1.455	0.007	HG12	20
11	1.189	0.005	HG13	20
12	0.866	0.017	QD1	20
13	0.900	0.009	QG2	20
14	124.353	0.011	N	20
15	177.168	0.000	C	21
16	53.291	0.117	CA	21
17	20.155	0.024	CB	21
18	8.335	0.009	H	21
19	4.324	0.010	HA	21
20	1.370	0.030	QB	21
21	127.789	0.011	N	21
22	177.313	0.000	C	22
23	53.111	0.126	CA	22
24	20.474	0.048	CB	22
25	8.067	0.005	H	22
26	4.288	0.004	HA	22
27	1.318	0.003	QB	22
28	123.346	0.033	N	22
29	177.255	0.000	C	23
30	53.157	0.151	CA	23
31	20.565	0.090	CB	23
32	8.222	0.006	H	23
33	4.290	0.008	HA	23
34	1.317	0.003	QB	23
35	123.325	0.018	N	23
36	175.486	0.000	C	24

37	63.610	0.096	CA	24
38	33.682	0.054	CB	24
39	22.138	0.000	CG1	24
40	22.052	0.041	CG2	24
41	8.024	0.009	H	24
42	3.979	0.011	HA	24
43	2.021	0.018	HB	24
44	0.930	0.006	QG1	24
45	0.924	0.006	QG2	24
46	119.719	0.011	N	24
47	176.474	0.000	C	25
48	62.955	0.247	CA	25
49	33.800	0.324	CB	25
50	22.046	0.052	CG1	25
51	22.056	0.052	CG2	25
52	8.163	0.003	H	25
53	4.105	0.055	HA	25
54	2.012	0.005	HB	25
55	0.925	0.010	QG1	25
56	0.920	0.008	QG2	25
57	124.681	0.012	N	25
58	173.504	0.000	C	26
59	45.708	0.196	CA	26
60	8.708	0.008	H	26
61	4.351	0.001	HA2	26
62	3.907	0.002	HA3	26
63	114.084	0.010	N	26
64	174.608	0.000	C	27
65	62.387	0.155	CA	27
66	71.565	0.016	CB	27
67	22.565	0.010	CG2	27
68	8.144	0.007	H	27
69	4.278	0.009	HA	27
70	4.111	0.010	HB	27
71	1.147	0.005	QG2	27
72	112.911	0.011	N	27
73	177.378	0.000	C	28
74	53.301	0.051	CA	28
75	20.067	0.054	CB	28
76	8.565	0.006	H	28
77	4.361	0.002	HA	28
78	1.394	0.001	QB	28
79	126.827	0.074	N	28
80	177.686	0.000	C	29
81	53.347	0.007	CA	29
82	20.121	0.076	CB	29
83	8.561	0.007	H	29
84	4.358	0.006	HA	29
85	1.395	0.001	QB	29
86	124.929	0.020	N	29
87	176.732	0.000	C	30
88	63.739	0.111	CA	30
89	33.525	0.067	CB	30
90	21.875	0.000	CG1	30
91	21.933	0.040	CG2	30
92	8.199	0.003	H	30
93	4.031	0.009	HA	30
94	2.066	0.004	HB	30
95	1.005	0.004	QG1	30
96	0.997	0.006	QG2	30
97	119.449	0.040	N	30
98	173.927	0.000	C	31
99	45.737	0.105	CA	31
100	8.408	0.008	H	31
101	3.712	0.003	HA2	31
102	4.260	0.008	HA3	31
103	113.570	0.024	N	31
104	174.286	0.000	C	32
105	62.866	0.120	CA	32
106	71.394	0.085	CB	32
107	22.773	0.075	CG2	32
108	8.159	0.004	H	32
109	4.390	0.002	HA	32
110	4.011	0.006	HB	32
111	1.231	0.003	QG2	32
112	115.558	0.023	N	32
113	176.047	0.000	C	33

```
114   57.307  0.132     CA 33
115   34.152  0.021     CB 33
116   29.866  0.048     CD 33
117   42.839  0.017     CE 33
118   25.812  0.074     CG 33
119    8.566  0.006      H 33
120    4.385  0.009     HA 33
121    1.864  0.009    HB2 33
122    1.793  0.004    HB3 33
123    1.704  0.009    HD2 33
124    1.696  0.007    HD3 33
125    3.001  0.011    HE2 33
126    3.006  0.003    HE3 33
127    1.472  0.016    HG2 33
128    1.503  0.014    HG3 33
129  125.475  0.025      N 33
130  177.258  0.000      C 34
131   53.174  0.142     CA 34
132   20.591  0.057     CB 34
133    8.192  0.005      H 34
134    4.292  0.006     HA 34
135    1.401  0.008     QB 34
136  125.792  0.046      N 34
137  177.752  0.000      C 35
138   53.530  0.118     CA 35
139   20.117  0.080     CB 35
140    8.331  0.013      H 35
141    4.361  0.005     HA 35
142    1.395  0.002     QB 35
143  123.435  0.028      N 35
144  174.021  0.000      C 36
145   62.599  0.049     CA 36
146   70.801  0.121     CB 36
147   21.779  0.009    CG2 36
148    8.011  0.006      H 36
149    4.276  0.005     HA 36
150    4.101  0.011     HB 36
151    1.150  0.008    QG2 36
152  113.087  0.024      N 36
153   52.826  0.070     CA 37
154   42.509  0.031     CB 37
155    8.385  0.003      H 37
156    4.897  0.002     HA 37
157    2.813  0.004    HB2 37
158    2.607  0.004    HB3 37
159  124.568  0.007      N 37
160  177.563  0.000      C 38
161   64.966  0.055     CA 38
162   33.137  0.014     CB 38
163   51.960  0.007     CD 38
164   28.121  0.016     CG 38
165    4.406  0.003     HA 38
166    1.977  0.018    HB2 38
167    2.318  0.004    HB3 38
168    3.907  0.010    HD2 38
169    3.889  0.008    HD3 38
170    2.006  0.014    HG2 38
171    2.006  0.012    HG3 38
172  177.031  0.000      C 39
173   57.624  0.100     CA 39
174   31.158  0.079     CB 39
175   44.183  0.020     CD 39
176   28.139  0.006     CG 39
177    8.447  0.005      H 39
178    4.253  0.006     HA 39
179    1.886  0.017    HB2 39
180    1.819  0.015    HB3 39
181    3.191  0.007    HD2 39
182    3.191  0.007    HD3 39
183    1.655  0.003    HG2 39
184    1.658  0.004    HG3 39
185  118.650  0.011      N 39
186  174.646  0.000      C 40
187   59.384  0.113     CA 40
188   64.705  0.090     CB 40
189    8.069  0.006      H 40
190    4.449  0.003     HA 40
```

```
191    3.883  0.003    HB2 40
192    3.881  0.007    HB3 40
193  115.525  0.013      N 40
194  176.664  0.000      C 41
195   63.497  0.073     CA 41
196   33.529  0.060     CB 41
197   22.162  0.078    CG1 41
198   21.337  0.020    CG2 41
199    7.945  0.008      H 41
200    4.135  0.003     HA 41
201    2.118  0.002     HB 41
202    0.936  0.003    QG1 41
203    0.938  0.002    QG2 41
204  121.035  0.013      N 41
205  174.263  0.000      C 42
206   46.257  0.126     CA 42
207    8.417  0.009      H 42
208    3.975  0.000    HA2 42
209    3.975  0.000    HA3 42
210  111.918  0.003      N 42
211  174.296  0.000      C 43
212   62.747  0.162     CA 43
213   70.787  0.016     CB 43
214   22.578  0.000    CG2 43
215    8.022  0.007      H 43
216    4.276  0.007     HA 43
217    4.113  0.011     HB 43
218    1.146  0.004    QG2 43
219  114.612  0.005      N 43
220  175.572  0.000      C 44
221   56.331  0.144     CA 44
222   30.324  0.060     CB 44
223   34.306  0.014     CG 44
224    8.449  0.009      H 44
225    4.363  0.005     HA 44
226    1.998  0.004    HB2 44
227    1.888  0.003    HB3 44
228    2.264  0.003    HG2 44
229    2.265  0.001    HG3 44
230  123.904  0.066      N 44
231  175.015  0.000      C 45
232   62.374  0.135     CA 45
233   34.335  0.023     CB 45
234   22.199  0.083    CG1 45
235   22.194  0.079    CG2 45
236    8.436  0.006      H 45
237    4.157  0.011     HA 45
238    1.933  0.002     HB 45
239    0.934  0.002    QG1 45
240    0.935  0.003    QG2 45
241  123.584  0.056      N 45
242  175.617  0.000      C 46
243   55.283  0.093     CA 46
244   42.855  0.063     CB 46
245    8.319  0.007      H 46
246    4.472  0.007     HA 46
247    2.857  0.001    HB2 46
248    2.667  0.002    HB3 46
249  125.417  0.036      N 46
250  177.746  0.000      C 47
251   58.855  0.007     CA 47
252   41.971  0.049     CB 47
253    8.461  0.005      H 47
254    4.237  0.010     HA 47
255    2.822  0.004    HB2 47
256    2.681  0.001    HB3 47
257  122.000  0.030      N 47
258  176.805  0.000      C 48
259   48.034  0.146     CA 48
260    8.474  0.003      H 48
261    3.863  0.015    HA2 48
262    3.852  0.001    HA3 48
263  106.617  0.005      N 48
264  176.215  0.000      C 49
265   67.179  0.110     CA 49
266   68.770  0.117     CB 49
267   22.462  0.079    CG2 49
```

```
268    7.923  0.009      H 49
269    3.901  0.005     HA 49
270    4.102  0.003     HB 49
271    1.129  0.010    QG2 49
272  119.992  0.046      N 49
273  178.218  0.000      C 50
274   59.300  0.043     CA 50
275   41.767  0.020     CB 50
276   24.332  0.004    CD2 50
277   27.844  0.000     CG 50
278    8.344  0.004      H 50
279    3.935  0.006     HA 50
280    1.883  0.004    HB2 50
281    1.300  0.004    HB3 50
282    1.697  0.006     HG 50
283    0.708  0.004    QD2 50
284  123.588  0.037      N 50
285  178.893  0.000      C 51
286   61.399  0.065     CA 51
287   29.881  0.081     CB 51
288   38.367  0.009     CG 51
289    8.308  0.004      H 51
290    3.559  0.004     HA 51
291    2.366  0.006    HB2 51
292    1.931  0.004    HB3 51
293    2.670  0.002    HG2 51
294    1.984  0.008    HG3 51
295  117.872  0.015      N 51
296  179.479  0.000      C 52
297   67.576  0.029     CA 52
298   32.675  0.102     CB 52
299   21.981  0.003    CG1 52
300   23.949  0.001    CG2 52
301    7.759  0.004      H 52
302    3.750  0.005     HA 52
303    2.282  0.006     HB 52
304    0.991  0.004    QG1 52
305    1.140  0.003    QG2 52
306  119.174  0.011      N 52
307  180.414  0.000      C 53
308   60.850  0.121     CA 53
309   32.440  0.148     CB 53
310   45.059  0.040     CD 53
311   22.317  0.113     CG 53
312    8.619  0.005      H 53
313    4.084  0.010     HA 53
314    2.043  0.007    HB2 53
315    1.860  0.013    HB3 53
316    3.287  0.005    HD2 53
317    2.841  0.007    HD3 53
318    1.125  0.010    HG2 53
319    0.942  0.007    HG3 53
320  121.213  0.015      N 53
321  177.448  0.000      C 54
322   67.747  0.084     CA 54
323   32.504  0.011     CB 54
324   24.193  0.021    CG1 54
325   23.174  0.000    CG2 54
326    8.789  0.005      H 54
327    3.720  0.003     HA 54
328    2.106  0.007     HB 54
329    1.001  0.008    QG1 54
330    0.787  0.005    QG2 54
331  120.336  0.010      N 54
332  178.677  0.000      C 55
333   58.322  0.088     CA 55
334   39.679  0.085     CB 55
335    8.691  0.002      H 55
336    4.343  0.004     HA 55
337    2.888  0.006    HB2 55
338    2.756  0.003    HB3 55
339  118.616  0.014      N 55
340  176.564  0.000      C 56
341   62.591  0.160     CA 56
342   63.682  0.077     CB 56
343    8.744  0.009      H 56
344    4.159  0.008     HA 56
```

```
345    3.919  0.008    HB2 56
346    4.024  0.008    HB3 56
347  115.419  0.016      N 56
348  180.741  0.000      C 57
349   56.307  0.086     CA 57
350   19.296  0.010     CB 57
351    7.638  0.003      H 57
352    4.107  0.005     HA 57
353    1.341  0.002     QB 57
354  124.318  0.013      N 57
355  178.656  0.000      C 58
356   59.108  0.131     CA 58
357   42.838  0.070     CB 58
358   25.021  0.069    CD1 58
359   24.363  0.009    CD2 58
360   27.687  0.091     CG 58
361    8.392  0.007      H 58
362    3.974  0.010     HA 58
363    2.002  0.006    HB2 58
364    1.336  0.005    HB3 58
365    1.794  0.005     HG 58
366    0.851  0.012    QD1 58
367    0.700  0.010    QD2 58
368  118.543  0.015      N 58
369  174.223  0.000      C 59
370   61.658  0.161     CA 59
371   64.420  0.004     CB 59
372    7.832  0.004      H 59
373    4.122  0.009     HA 59
374    3.995  0.009    HB2 59
375    3.995  0.009    HB3 59
376  110.851  0.007      N 59
377  176.340  0.000      C 60
378   57.565  0.125     CA 60
379   33.415  0.095     CB 60
380   29.793  0.066     CD 60
381   42.926  0.100     CE 60
382   26.052  0.051     CG 60
383    7.073  0.002      H 60
384    4.157  0.009     HA 60
385    2.009  0.009    HB2 60
386    1.861  0.007    HB3 60
387    1.673  0.007    HD2 60
388    1.678  0.009    HD3 60
389    2.959  0.007    HE2 60
390    2.966  0.008    HE3 60
391    1.681  0.005    HG2 60
392    1.529  0.007    HG3 60
393  118.073  0.016      N 60
394  174.671  0.000      C 61
395   54.534  0.123     CA 61
396   43.767  0.053     CB 61
397    7.517  0.004      H 61
398    4.697  0.006     HA 61
399    2.232  0.005    HB2 61
400    3.208  0.005    HB3 61
401  121.471  0.008      N 61
402  178.138  0.000      C 62
403   60.448  0.108     CA 62
404   30.782  0.115     CB 62
405   37.173  0.008     CG 62
406    8.682  0.008      H 62
407    3.910  0.006     HA 62
408    2.027  0.004    HB2 62
409    2.026  0.002    HB3 62
410    2.366  0.003    HG2 62
411    2.295  0.004    HG3 62
412  125.751  0.011      N 62
413  178.380  0.000      C 63
414   60.271  0.091     CA 63
415   29.815  0.063     CB 63
416   35.255  0.093     CG 63
417    8.064  0.003      H 63
418    4.121  0.006     HA 63
419    1.914  0.003    HB2 63
420    2.011  0.008    HB3 63
421    2.269  0.005    HG2 63
```

```
422    2.480  0.009    HG3 63
423  117.838  0.015      N 63
424  177.601  0.000      C 64
425   67.211  0.096     CA 64
426   38.348  0.083     CB 64
427   14.677  0.016    CD1 64
428   31.960  0.008    CG1 64
429   18.901  0.004    CG2 64
430    8.421  0.015      H 64
431    3.428  0.004     HA 64
432    2.085  0.009     HB 64
433    1.744  0.006   HG12 64
434    0.787  0.009    QD1 64
435    0.790  0.006    QG2 64
436  121.112  0.045      N 64
437  178.269  0.000      C 65
438   59.040  0.199     CA 65
439   33.345  0.036     CB 65
440   29.845  0.057     CD 65
441   42.844  0.000     CE 65
442   25.453  0.031     CG 65
443    8.016  0.009      H 65
444    4.088  0.009     HA 65
445    1.862  0.005    HB2 65
446    1.865  0.003    HB3 65
447    1.718  0.004    HD2 65
448    1.718  0.002    HD3 65
449    2.998  0.010    HE2 65
450    3.006  0.000    HE3 65
451    1.511  0.005    HG2 65
452    1.508  0.001    HG3 65
453  116.289  0.042      N 65
454  178.543  0.000      C 66
455   59.184  0.112     CA 66
456   35.252  0.076     CB 66
457   29.904  0.034     CD 66
458   42.841  0.009     CE 66
459   26.420  0.015     CG 66
460    7.764  0.009      H 66
461    4.324  0.004     HA 66
462    1.901  0.004    HB2 66
463    2.004  0.004    HB3 66
464    1.716  0.008    HD2 66
465    1.708  0.005    HD3 66
466    3.001  0.002    HE2 66
467    3.003  0.003    HE3 66
468    1.467  0.008    HG2 66
469    1.610  0.003    HG3 66
470  115.520  0.024      N 66
471  175.334  0.000      C 67
472   58.002  0.096     CA 67
473   33.301  0.176     CB 67
474   37.732  0.016     CG 67
475    8.331  0.010      H 67
476    4.459  0.008     HA 67
477    2.067  0.010    HB2 67
478    2.066  0.005    HB3 67
479    2.213  0.008    HG2 67
480    2.365  0.002    HG3 67
481  115.048  0.091      N 67
482  174.999  0.000      C 68
483   50.847  0.098     CA 68
484   23.074  0.009     CB 68
485    7.870  0.010      H 68
486    5.003  0.004     HA 68
487    1.527  0.001     QB 68
488  119.016  0.014      N 68
489  173.247  0.000      C 69
490   56.591  0.019     CA 69
491   33.540  0.004     CB 69
492    8.562  0.003      H 69
493    4.382  0.005     HA 69
494    1.864  0.005    HB2 69
495    1.768  0.005    HB3 69
496  122.196  0.019      N 69
497  174.775  0.000      C 70
498   60.378  0.099     CA 70
```

```
499   41.422  0.035     CB 70
500   15.372  0.013    CD1 70
501   27.761  0.081    CG1 70
502   18.822  0.029    CG2 70
503    8.211  0.012      H 70
504    4.855  0.008     HA 70
505    1.516  0.006     HB 70
506    1.330  0.009   HG12 70
507    0.680  0.005   HG13 70
508    0.595  0.004    QD1 70
509    0.671  0.006    QG2 70
510  125.388  0.023      N 70
511  172.962  0.000      C 71
512   52.569  0.114     CA 71
513   42.743  0.118     CB 71
514    9.211  0.008      H 71
515    5.169  0.003     HA 71
516    2.595  0.006    HB2 71
517    2.195  0.004    HB3 71
518  126.356  0.013      N 71
519  175.226  0.000      C 72
520   61.602  0.119     CA 72
521   36.098  0.075     CB 72
522   22.792  0.061    CG1 72
523   22.718  0.018    CG2 72
524    8.300  0.012      H 72
525    4.887  0.007     HA 72
526    1.823  0.008     HB 72
527    0.780  0.005    QG1 72
528    0.696  0.002    QG2 72
529  119.007  0.054      N 72
530  171.621  0.000      C 73
531   62.043  0.093     CA 73
532   72.668  0.108     CB 73
533   22.884  0.003    CG2 73
534    8.601  0.005      H 73
535    4.684  0.006     HA 73
536    3.681  0.001     HB 73
537    1.054  0.003    QG2 73
538  123.404  0.041      N 73
539  176.399  0.000      C 74
540   50.702  0.095     CA 74
541   24.470  0.058     CB 74
542   10.017  0.010      H 74
543    5.725  0.003     HA 74
544    1.347  0.002     QB 74
545  129.797  0.013      N 74
546  174.901  0.000      C 75
547   59.532  0.142     CA 75
548   41.521  0.136     CB 75
549  134.016  0.022    CD1 75
550  134.018  0.000    CD2 75
551  119.350  0.067    CE1 75
552  119.313  0.048    CE2 75
553    8.929  0.003      H 75
554    4.612  0.008     HA 75
555    2.770  0.008    HB2 75
556    3.044  0.006    HB3 75
557    7.078  0.002    HD1 75
558    7.078  0.005    HD2 75
559    6.783  0.004    HE1 75
560    6.785  0.003    HE2 75
561  119.098  0.009      N 75
562  176.768  0.000      C 76
563   57.026  0.121     CA 76
564   27.602  0.041     CB 76
565   34.404  0.069     CG 76
566   10.082  0.008      H 76
567    3.727  0.003     HA 76
568    1.700  0.003    HB2 76
569    2.012  0.007    HB3 76
570    1.834  0.003    HG2 76
571    1.690  0.006    HG3 76
572  130.882  0.022      N 76
573  173.173  0.000      C 77
574   46.664  0.125     CA 77
575    9.218  0.006      H 77
```

```
576    4.275  0.004    HA2 77
577    3.591  0.002    HA3 77
578  106.396  0.014      N 77
579  175.585  0.000      C 78
580   56.458  0.079     CA 78
581   33.459  0.014     CB 78
582   30.667  0.043     CD 78
583   42.979  0.029     CE 78
584    7.964  0.011      H 78
585    4.891  0.006     HA 78
586    1.618  0.010    HD2 78
587    1.764  0.008    HD3 78
588    3.070  0.010    HE2 78
589    3.121  0.005    HE3 78
590  121.088  0.020      N 78
591  172.662  0.000      C 79
592   62.667  0.028     CA 79
593   35.267  0.031     CB 79
594   22.989  0.068    CG1 79
595   22.956  0.002    CG2 79
596    8.898  0.009      H 79
597    4.309  0.011     HA 79
598    2.038  0.003     HB 79
599    0.789  0.009    QG1 79
600    0.789  0.005    QG2 79
601  122.912  0.008      N 79
602  174.406  0.000      C 80
603   54.962  0.095     CA 80
604   45.283  0.094     CB 80
605   26.701  0.000    CD1 80
606   28.525  0.019     CG 80
607    9.508  0.004      H 80
608    4.937  0.009     HA 80
609    1.832  0.006    HB2 80
610    1.192  0.009    HB3 80
611    1.073  0.010     HG 80
612    0.356  0.006    QD1 80
613  131.036  0.010      N 80
614  174.480  0.000      C 81
615   53.850  0.057     CA 81
616   44.127  0.050     CB 81
617   25.056  0.018    CD1 81
618   26.442  0.007    CD2 81
619   27.359  0.000     CG 81
620    8.364  0.005      H 81
621    4.888  0.011     HA 81
622    1.009  0.004    HB2 81
623    1.894  0.007    HB3 81
624    1.506  0.007     HG 81
625    0.591  0.006    QD1 81
626    0.573  0.009    QD2 81
627  124.583  0.004      N 81
628  173.952  0.000      C 82
629   58.973  0.102     CA 82
630   35.546  0.036     CB 82
631   21.501  0.005    CG1 82
632   19.759  0.002    CG2 82
633    8.180  0.006      H 82
634    4.131  0.007     HA 82
635    1.991  0.004     HB 82
636   -0.448  0.002    QG1 82
637    0.402  0.003    QG2 82
638  113.407  0.022      N 82
639  174.886  0.000      C 83
640   44.504  0.117     CA 83
641    9.662  0.006      H 83
642    5.675  0.002    HA2 83
643    3.340  0.004    HA3 83
644  107.677  0.024      N 83
645  174.350  0.000      C 84
646   56.850  0.008     CA 84
647   33.853  0.006     CB 84
648    9.057  0.004      H 84
649  123.584  0.023      N 84
650   62.782  0.006     CA 85
651   67.222  0.000     CB 85
652    8.366  0.005      H 85
```

```
653  114.430  0.088      N 85
654  175.810  0.000      C 86
655   65.003  0.055     CA 86
656   33.354  0.079     CB 86
657   52.480  0.035     CD 86
658   28.126  0.009     CG 86
659    4.354  0.004     HA 86
660    2.222  0.003    HB2 86
661    2.092  0.011    HB3 86
662    3.828  0.003    HD2 86
663    3.824  0.003    HD3 86
664    2.145  0.004    HG2 86
665    2.146  0.006    HG3 86
666  175.699  0.000      C 87
667   53.244  0.208     CA 87
668   42.290  0.103     CB 87
669    7.400  0.012      H 87
670    4.898  0.002     HA 87
671    2.798  0.003    HB2 87
672    2.953  0.002    HB3 87
673  112.562  0.087      N 87
674  179.540  0.000      C 88
675   55.607  0.082     CA 88
676   19.528  0.068     CB 88
677    9.111  0.009      H 88
678    4.042  0.013     HA 88
679    1.456  0.008     QB 88
680  127.024  0.038      N 88
681  179.667  0.000      C 89
682   60.707  0.127     CA 89
683   29.721  0.016     CB 89
684   37.558  0.003     CG 89
685    8.527  0.003      H 89
686    4.060  0.005     HA 89
687    2.077  0.002    HB2 89
688    1.945  0.008    HB3 89
689    2.259  0.003    HG2 89
690    2.258  0.005    HG3 89
691  120.709  0.005      N 89
692  178.643  0.000      C 90
693   58.650  0.089     CA 90
694   41.765  0.024     CB 90
695   24.352  0.060    CD1 90
696   24.339  0.012    CD2 90
697   27.946  0.156     CG 90
698    7.949  0.004      H 90
699    3.940  0.004     HA 90
700    1.883  0.007    HB2 90
701    1.301  0.009    HB3 90
702    1.692  0.007     HG 90
703    0.830  0.006    QD1 90
704    0.710  0.003    QD2 90
705  119.223  0.047      N 90
706  175.678  0.000      C 91
707   61.372  0.118     CA 91
708   63.564  0.053     CB 91
709    6.922  0.013      H 91
710    3.942  0.007     HA 91
711    3.897  0.003    HB2 91
712    3.826  0.003    HB3 91
713  112.668  0.133      N 91
714  180.831  0.000      C 92
715   56.045  0.119     CA 92
716   18.894  0.204     CB 92
717    7.787  0.003      H 92
718    3.862  0.009     HA 92
719    1.460  0.009     QB 92
720  122.203  0.013      N 92
721  177.924  0.000      C 93
722   60.429  0.197     CA 93
723   31.583  0.005     CB 93
724   44.965  0.018     CD 93
725   28.908  0.033     CG 93
726    8.112  0.004      H 93
727    4.026  0.008     HA 93
728    1.801  0.006    HB2 93
729    1.898  0.005    HB3 93
```

```
 730    3.183  0.007    HD2 93
 731    3.261  0.006    HD3 93
 732    1.779  0.001    HG2 93
 733    1.555  0.006    HG3 93
 734  120.613  0.049      N 93
 735  178.440  0.000      C 94
 736   56.029  0.062     CA 94
 737   19.176  0.041     CB 94
 738    8.101  0.005      H 94
 739    3.754  0.004     HA 94
 740    1.335  0.005     QB 94
 741  121.200  0.047      N 94
 742  177.153  0.000      C 95
 743   61.094  0.051     CA 95
 744   33.135  0.013     CB 95
 745   30.872  0.048     CD 95
 746   43.023  0.060     CE 95
 747    7.599  0.005      H 95
 748    3.641  0.010     HA 95
 749    1.677  0.012    HB2 95
 750    2.082  0.005    HB3 95
 751    1.679  0.006    HD2 95
 752    1.752  0.012    HD3 95
 753    2.894  0.002    HE2 95
 754    3.075  0.005    HE3 95
 755  116.685  0.012      N 95
 756  179.334  0.000      C 96
 757   59.940  0.106     CA 96
 758   29.294  0.021     CB 96
 759   34.478  0.015     CG 96
 760    7.754  0.005      H 96
 761    3.946  0.006     HA 96
 762    2.206  0.008    HB2 96
 763    2.209  0.007    HB3 96
 764    2.366  0.007    HG2 96
 765    2.408  0.009    HG3 96
 766  118.337  0.010      N 96
 767  179.622  0.000      C 97
 768   66.266  0.047     CA 97
 769   39.275  0.023     CB 97
# 770   15.677  0.029    CD1 97
# 771   24.388  0.026    CG1 97
# 772   18.777  0.012    CG2 97
 773    8.566  0.006      H 97
# 774    3.552  0.012     HA 97
# 775    1.809  0.007     HB 97
# 776    1.348  0.007   HG12 97
# 777    0.840  0.002    QD1 97
# 778    0.687  0.006    QG2 97
 779  120.633  0.012      N 97
 780  177.663  0.000      C 98
 781   56.537  0.130     CA 98
 782   20.412  0.222     CB 98
 783    8.268  0.006      H 98
 784    3.883  0.003     HA 98
 785    1.366  0.006     QB 98
 786  122.084  0.114      N 98
 787  177.095  0.000      C 99
 788   58.750  0.061     CA 99
 789   34.591  0.006     CB 99
 790   33.130  0.017     CG 99
 791    8.094  0.005      H 99
 792    3.886  0.012     HA 99
 793    2.139  0.004    HB2 99
 794    2.090  0.011    HB3 99
 795    2.391  0.005    HG2 99
 796    2.668  0.003    HG3 99
 797  112.053  0.011      N 99
 798  174.134  0.000      C 100
 799   45.739  0.099     CA 100
 800    7.509  0.006      H 100
 801    4.264  0.008    HA2 100
 802    3.716  0.010    HA3 100
 803  104.372  0.009      N 100
 804  175.927  0.000      C 101
 805   63.507  0.085     CA 101
 806   32.522  0.080     CB 101
```

```
807   24.111  0.020    CG1 101
808   23.247  0.000    CG2 101
809    7.289  0.004      H 101
810    3.795  0.007     HA 101
811    2.107  0.007     HB 101
812    0.966  0.007    QG1 101
813    0.787  0.004    QG2 101
814  122.787  0.007      N 101
815  175.482  0.000      C 102
816   57.152  0.131     CA 102
817   41.565  0.033     CB 102
818    8.546  0.008      H 102
819    4.265  0.004     HA 102
820    2.538  0.010    HB2 102
821    2.590  0.002    HB3 102
822  129.235  0.008      N 102
823  173.605  0.000      C 103
824   45.623  0.067     CA 103
825    8.500  0.006      H 103
826    4.255  0.007    HA2 103
827    3.496  0.006    HA3 103
828  113.812  0.008      N 103
829  175.845  0.000      C 104
830   52.204  0.079     CA 104
831   19.199  0.004     CB 104
832    7.714  0.002      H 104
833    4.270  0.002     HA 104
834    1.159  0.003     QB 104
835  122.115  0.010      N 104
836  176.087  0.000      C 105
837   55.426  0.115     CA 105
838   40.684  0.007     CB 105
839    9.144  0.007      H 105
840    4.791  0.014     HA 105
841    2.798  0.005    HB2 105
842    2.793  0.005    HB3 105
843  123.360  0.025      N 105
844  174.076  0.000      C 106
845   56.775  0.006     CA 106
846   34.853  0.028     CB 106
847   37.127  0.010     CG 106
848    7.725  0.005      H 106
849    4.877  0.004     HA 106
850    1.914  0.006    HB2 106
851    1.783  0.007    HB3 106
852    1.924  0.005    HG2 106
853    1.956  0.008    HG3 106
854  118.819  0.055      N 106
855  174.360  0.000      C 107
856   61.740  0.009     CA 107
857   35.769  0.005     CB 107
858    8.681  0.005      H 107
859  122.777  0.013      N 107
860  174.442  0.000      C 108
861   57.327  0.002     CA 108
862   40.611  0.005     CB 108
863  133.696  0.006    CD1 108
864  133.700  0.000    CD2 108
865  119.067  0.005    CE1 108
866    9.594  0.006      H 108
867    6.969  0.002    HD1 108
868    6.967  0.002    HD2 108
869    6.646  0.001    HE1 108
870    6.641  0.000    HE2 108
871  128.545  0.010      N 108
872  174.608  0.000      C 109
873   52.984  0.166     CA 109
874   39.208  0.043     CB 109
875    8.899  0.013      H 109
876    4.892  0.006     HA 109
877    3.235  0.006    HB2 109
878    2.116  0.005    HB3 109
879  121.787  0.008      N 109
880  174.859  0.000      C 110
881   54.985  0.008     CA 110
882   30.059  0.006     CB 110
883    8.586  0.005      H 110
```

```
884  126.212  0.017      N 110
885  177.395  0.000      C 111
886   64.041  0.095     CA 111
887   38.786  0.039     CB 111
888   16.008  0.003    CD1 111
889   19.661  0.006    CG2 111
890    7.510  0.004      H 111
891    3.385  0.010     HA 111
892    1.670  0.003     HB 111
893    0.685  0.001    QD1 111
894    0.499  0.004    QG2 111
895  120.638  0.008      N 111
896  175.255  0.000      C 112
897   53.742  0.004     CA 112
898   31.351  0.003     CB 112
899    8.382  0.006      H 112
900  128.288  0.019      N 112
901  176.722  0.000      C 113
902   56.227  0.124     CA 113
903   29.898  0.070     CB 113
904   35.034  0.019     CG 113
905    8.565  0.008      H 113
906    4.960  0.003     HA 113
907    1.983  0.006    HB2 113
908    2.025  0.006    HB3 113
909    2.274  0.001    HG2 113
910    2.271  0.001    HG3 113
911  121.245  0.022      N 113
912  171.061  0.000      C 114
913   45.317  0.054     CA 114
914    8.151  0.008      H 114
915    3.864  0.008    HA2 114
916    4.224  0.008    HA3 114
917  115.185  0.009      N 114
918   54.033  0.080     CA 115
919   29.558  0.011     CB 115
920   34.413  0.023     CG 115
921    7.903  0.005      H 115
922    4.484  0.005     HA 115
923    2.043  0.006    HB2 115
924    1.805  0.002    HB3 115
925    2.416  0.008    HG2 115
926    2.447  0.007    HG3 115
927  115.852  0.004      N 115
928  177.337  0.000      C 116
929   64.336  0.003     CA 116
930   33.714  0.000     CB 116
931  175.481  0.000      C 117
932   62.714  0.067     CA 117
933   39.312  0.053     CB 117
934   15.015  0.012    CD1 117
935   17.880  0.023    CG2 117
936    7.235  0.003      H 117
937    4.245  0.011     HA 117
938    2.028  0.004     HB 117
939    0.825  0.004    QD1 117
940    0.823  0.004    QG2 117
941  118.035  0.013      N 117
942  174.367  0.000      C 118
943   44.869  0.117     CA 118
944    8.117  0.004      H 118
945    4.372  0.007    HA2 118
946    3.886  0.006    HA3 118
947  106.683  0.010      N 118
948  180.666  0.000      C 119
949   58.816  0.084     CA 119
950   43.027  0.028     CB 119
951   25.661  0.004    CD1 119
952   24.504  0.006    CD2 119
953   28.025  0.015     CG 119
954    8.388  0.008      H 119
955    4.072  0.004     HA 119
956    1.579  0.003    HB2 119
957    1.725  0.007    HB3 119
958    1.742  0.005     HG 119
959    0.955  0.004    QD1 119
960    0.917  0.006    QD2 119
```

961	118.761	0.030	N	119
962	177.039	0.000	C	120
963	47.986	0.109	CA	120
964	8.921	0.004	H	120
965	3.850	0.008	HA2	120
966	3.856	0.013	HA3	120
967	109.088	0.008	N	120
968	178.795	0.000	C	121
969	60.744	0.003	CA	121
970	30.620	0.064	CB	121
971	38.024	0.015	CG	121
972	8.173	0.003	H	121
973	4.127	0.005	HA	121
974	2.246	0.007	HB2	121
975	2.028	0.005	HB3	121
976	2.486	0.003	HG2	121
977	2.267	0.003	HG3	121
978	123.826	0.006	N	121
979	180.721	0.000	C	122
980	56.059	0.127	CA	122
981	19.080	0.060	CB	122
982	8.443	0.004	H	122
983	4.205	0.005	HA	122
984	1.475	0.002	QB	122
985	122.626	0.010	N	122
986	177.353	0.000	C	123
987	62.656	0.001	CA	123
988	63.438	0.093	CB	123
989	8.349	0.006	H	123
990	4.381	0.009	HA	123
991	4.068	0.006	HB2	123
992	4.010	0.005	HB3	123
993	114.755	0.008	N	123
994	176.973	0.000	C	124
995	57.570	0.083	CA	124
996	39.520	0.065	CB	124
997	7.759	0.006	H	124
998	4.669	0.005	HA	124
999	2.860	0.003	HB2	124
1000	3.100	0.002	HB3	124
1001	122.791	0.031	N	124
1002	178.314	0.000	C	125
1003	58.905	0.079	CA	125
1004	40.895	0.087	CB	125
1005	8.176	0.005	H	125
1006	4.448	0.008	HA	125
1007	3.118	0.006	HB2	125
1008	2.611	0.005	HB3	125
1009	121.109	0.021	N	125
1010	176.741	0.000	C	126
1011	67.650	0.107	CA	126
1012	69.683	0.128	CB	126
1013	23.028	0.000	CG2	126
1014	7.896	0.007	H	126
1015	3.987	0.007	HA	126
1016	4.377	0.007	HB	126
1017	1.320	0.002	QG2	126
1018	115.915	0.007	N	126
1019	178.246	0.000	C	127
1020	61.962	0.136	CA	127
1021	29.875	0.045	CB	127
1022	128.229	0.031	CD1	127
1023	125.281	0.009	CH2	127
1024	115.247	0.015	CZ2	127
1025	120.661	0.000	CZ3	127
1026	7.885	0.004	H	127
1027	4.431	0.008	HA	127
1028	3.377	0.005	HB2	127
1029	3.725	0.010	HB3	127
1030	7.296	0.009	HD1	127
1031	7.180	0.031	HH2	127
1032	7.458	0.001	HZ2	127
1033	7.117	0.000	HZ3	127
1034	124.991	0.007	N	127
1035	177.992	0.000	C	128
1036	67.452	0.060	CA	128
1037	38.461	0.134	CB	128

1038	15.045	0.042	CD1	128
1039	20.538	0.000	CG2	128
1040	8.767	0.007	H	128
1041	3.294	0.008	HA	128
1042	2.144	0.006	HB	128
1043	0.841	0.011	QD1	128
1044	0.719	0.007	QG2	128
1045	121.564	0.005	N	128
1046	175.846	0.000	C	129
1047	68.366	0.066	CA	129
1048	69.176	0.059	CB	129
1049	23.703	0.015	CG2	129
1050	8.748	0.003	H	129
1051	3.450	0.003	HA	129
1052	4.417	0.003	HB	129
1053	1.105	0.002	QG2	129
1054	118.077	0.018	N	129
1055	176.583	0.000	C	130
1056	67.805	0.111	CA	130
1057	69.648	0.109	CB	130
1058	22.377	0.041	CG2	130
1059	8.059	0.003	H	130
1060	3.783	0.002	HA	130
1061	4.354	0.001	HB	130
1062	1.224	0.004	QG2	130
1063	116.905	0.059	N	130
1064	179.483	0.000	C	131
1065	59.847	0.084	CA	131
1066	32.891	0.080	CB	131
1067	29.455	0.058	CD	131
1068	42.456	0.028	CE	131
1069	25.261	0.059	CG	131
1070	8.146	0.003	H	131
1071	3.904	0.007	HA	131
1072	1.518	0.008	HB2	131
1073	1.400	0.013	HB3	131
1074	1.340	0.011	HD2	131
1075	1.395	0.005	HD3	131
1076	2.602	0.002	HE2	131
1077	2.342	0.009	HE3	131
1078	1.013	0.010	HG2	131
1079	0.812	0.005	HG3	131
1080	124.221	0.059	N	131
1081	178.101	0.000	C	132
1082	68.299	0.048	CA	132
1083	32.609	0.009	CB	132
1084	24.227	0.001	CG1	132
1085	23.283	0.041	CG2	132
1086	9.176	0.005	H	132
1087	3.349	0.011	HA	132
1088	1.006	0.011	QG1	132
1089	0.786	0.003	QG2	132
1090	120.235	0.010	N	132
1091	178.925	0.000	C	133
1092	61.552	0.053	CA	133
1093	31.128	0.037	CB	133
1094	44.199	0.033	CD	133
1095	8.690	0.003	H	133
1096	3.797	0.009	HA	133
1097	1.901	0.003	HB2	133
1098	1.946	0.005	HB3	133
1099	3.188	0.007	HD2	133
1100	3.186	0.005	HD3	133
1101	117.544	0.007	N	133
1102	176.870	0.000	C	134
1103	62.751	0.087	CA	134
1104	63.623	0.053	CB	134
1105	8.074	0.005	H	134
1106	4.210	0.006	HA	134
1107	4.016	0.004	HB2	134
1108	4.015	0.003	HB3	134
1109	113.254	0.010	N	134
1110	179.464	0.000	C	135
1111	58.853	0.093	CA	135
1112	29.922	0.106	CB	135
1113	35.230	0.003	CG	135
1114	7.944	0.004	H	135

```
1115    4.130  0.007     HA 135
1116    1.987  0.008    HB2 135
1117    1.986  0.009    HB3 135
1118    2.373  0.006    HG2 135
1119    2.402  0.006    HG3 135
1120  120.359  0.011      N 135
1121  178.802  0.000      C 136
1122   59.153  0.007     CA 136
1123   42.362  0.006     CB 136
1124    8.456  0.004      H 136
1125  120.435  0.009      N 136
1126  178.647  0.000      C 137
1127   58.269  0.059     CA 137
1128   43.184  0.082     CB 137
1129   25.767  0.012    CD1 137
1130   24.975  0.064    CD2 137
1131   28.029  0.011     CG 137
1132    7.714  0.004      H 137
1133    4.189  0.005     HA 137
1134    1.856  0.004    HB2 137
1135    1.756  0.003    HB3 137
1136    1.757  0.003     HG 137
1137    0.929  0.006    QD1 137
1138    0.938  0.003    QD2 137
1139  116.989  0.011      N 137
1140  174.668  0.000      C 138
1141   63.230  0.004     CA 138
1142   70.329  0.125     CB 138
1143   22.539  0.012    CG2 138
1144    7.450  0.005      H 138
1145    4.276  0.005     HA 138
1146    4.395  0.002     HB 138
1147    1.274  0.002    QG2 138
1148  107.380  0.031      N 138
1149  175.546  0.000      C 139
1150   58.670  0.062     CA 139
1151   65.182  0.137     CB 139
1152    7.221  0.003      H 139
1153    4.486  0.004     HA 139
1154    3.976  0.004    HB2 139
1155    3.878  0.010    HB3 139
1156  115.703  0.006      N 139
1157  176.755  0.000      C 140
1158   56.569  0.140     CA 140
1159   41.422  0.010     CB 140
1160    8.824  0.007      H 140
1161    4.566  0.004     HA 140
1162    2.722  0.011    HB2 140
1163    2.694  0.015    HB3 140
1164  126.239  0.018      N 140
1165  176.879  0.000      C 141
1166   56.843  0.130     CA 141
1167   44.608  0.041     CB 141
1168   24.028  0.020    CD1 141
1169   26.148  0.020    CD2 141
1170   28.038  0.014     CG 141
1171    8.134  0.005      H 141
1172    4.225  0.007     HA 141
1173    1.560  0.007    HB2 141
1174    1.371  0.004    HB3 141
1175    1.593  0.011     HG 141
1176    0.771  0.006    QD1 141
1177    0.868  0.007    QD2 141
1178  117.979  0.023      N 141
1179  175.487  0.000      C 142
1180   61.586  0.119     CA 142
1181   34.917  0.125     CB 142
1182   23.069  0.027    CG1 142
1183   22.165  0.008    CG2 142
1184    7.050  0.005      H 142
1185    4.456  0.002     HA 142
1186    2.098  0.003     HB 142
1187    0.817  0.005    QG1 142
1188    0.837  0.005    QG2 142
1189  112.212  0.336      N 142
1190  177.272  0.000      C 143
1191   57.832  0.002     CA 143
```

1192	33.036	0.008	CB	143
1193	8.486	0.004	H	143
1194	125.343	0.103	N	143
1195	175.271	0.000	C	144
1196	61.452	0.009	CA	144
1197	63.947	0.012	CB	144
1198	8.516	0.019	H	144
1199	119.014	0.110	N	144
1200	174.636	0.000	C	145
1201	60.795	0.085	CA	145
1202	64.536	0.092	CB	145
1203	8.111	0.009	H	145
1204	4.371	0.008	HA	145
1205	3.885	0.005	HB2	145
1206	3.886	0.007	HB3	145
1207	115.671	0.094	N	145
1208	173.188	0.000	C	146
1209	53.982	0.112	CA	146
1210	39.738	0.111	CB	146
1211	8.301	0.004	H	146
1212	4.924	0.004	HA	146
1213	2.856	0.002	HB2	146
1214	3.363	0.004	HB3	146
1215	120.680	0.019	N	146
1216	174.671	0.000	C	147
1217	62.228	0.123	CA	147
1218	35.257	0.033	CB	147
1219	22.627	0.066	CG1	147
1220	23.275	0.067	CG2	147
1221	7.519	0.011	H	147
1222	4.583	0.005	HA	147
1223	1.867	0.006	HB	147
1224	0.704	0.005	QG1	147
1225	0.837	0.008	QG2	147
1226	117.770	0.013	N	147
1227	174.586	0.000	C	148
1228	55.919	0.128	CA	148
1229	35.958	0.023	CB	148
1230	30.758	0.008	CD	148
1231	42.896	0.010	CE	148
1232	22.875	0.009	CG	148
1233	8.852	0.004	H	148
1234	4.608	0.007	HA	148
1235	1.318	0.010	HB2	148
1236	1.488	0.007	HB3	148
1237	1.777	0.000	HD2	148
1238	1.683	0.004	HD3	148
1239	3.074	0.007	HE2	148
1240	3.077	0.005	HE3	148
1241	1.055	0.002	HG2	148
1242	1.062	0.000	HG3	148
1243	128.488	0.026	N	148
1244	175.194	0.000	C	149
1245	61.916	0.004	CA	149
1246	35.831	0.023	CB	149
1247	22.787	0.009	CG1	149
1248	22.724	0.041	CG2	149
1249	8.398	0.006	H	149
1250	4.738	0.005	HA	149
1251	1.831	0.006	HB	149
1252	0.887	0.003	QG1	149
1253	0.757	0.009	QG2	149
1254	124.716	0.058	N	149
1255	171.249	0.000	C	150
1256	62.736	0.032	CA	150
1257	72.174	0.099	CB	150
1258	21.071	0.015	CG2	150
1259	8.568	0.004	H	150
1260	4.481	0.005	HA	150
1261	3.846	0.003	HB	150
1262	1.002	0.006	QG2	150
1263	122.209	0.017	N	150
1264	173.987	0.000	C	151
1265	63.147	0.049	CA	151
1266	70.199	0.126	CB	151
1267	21.914	0.049	CG2	151
1268	7.918	0.012	H	151

```
1269    5.444  0.003     HA 151
1270    3.768  0.005     HB 151
1271    0.947  0.004    QG2 151
1272  123.096  0.007      N 151
1273   55.349  0.005     CA 152
1274   34.084  0.000     CB 152
1275    9.248  0.004      H 152
1276  126.410  0.017      N 152
1277  176.914  0.000      C 153
1278   53.304  0.137     CA 153
1279   38.376  0.105     CB 153
1280    4.427  0.002     HA 153
1281    3.451  0.003    HB2 153
1282    2.240  0.003    HB3 153
1283  170.905  0.000      C 154
1284   46.791  0.048     CA 154
1285    9.426  0.012      H 154
1286    3.263  0.003    HA2 154
1287    4.046  0.000    HA3 154
1288  105.279  0.037      N 154
1289  173.770  0.000      C 155
1290   55.292  0.004     CA 155
1291   31.208  0.004     CB 155
1292    7.742  0.011      H 155
1293  122.830  0.042      N 155
1294  173.759  0.000      C 156
1295   62.607  0.039     CA 156
1296   34.441  0.035     CB 156
1297   25.451  0.045    CG1 156
1298   22.995  0.044    CG2 156
1299    8.452  0.011      H 156
1300    4.357  0.005     HA 156
1301    1.744  0.005     HB 156
1302    1.010  0.007    QG1 156
1303    0.761  0.008    QG2 156
1304  127.366  0.004      N 156
1305  175.407  0.000      C 157
1306   56.953  0.005     CA 157
1307   42.634  0.003     CB 157
1308    9.672  0.009      H 157
1309  125.804  0.010      N 157
1310  174.102  0.000      C 158
1311   53.165  0.003     CA 158
1312   46.561  0.008     CB 158
1313    7.509  0.004      H 158
1314  119.140  0.007      N 158
1315  173.915  0.000      C 159
1316   54.937  0.090     CA 159
1317   39.588  0.053     CB 159
1318   32.449  0.005     CG 159
1319    8.410  0.009      H 159
1320    4.143  0.005     HA 159
1321    1.408  0.005    HB2 159
1322    2.031  0.005    HB3 159
1323    2.162  0.010    HG2 159
1324    2.160  0.008    HG3 159
1325  119.471  0.049      N 159
1326  171.109  0.000      C 160
1327   45.663  0.125     CA 160
1328    8.259  0.007      H 160
1329    4.858  0.002    HA2 160
1330    2.629  0.001    HA3 160
1331  108.581  0.008      N 160
1332  175.915  0.000      C 161
1333   55.199  0.120     CA 161
1334   44.401  0.090     CB 161
1335   26.701  0.000    CD1 161
1336   25.097  0.030    CD2 161
1337   28.305  0.053     CG 161
1338    7.557  0.005      H 161
1339    5.339  0.005     HA 161
1340    1.872  0.007    HB2 161
1341    1.553  0.005    HB3 161
1342    1.377  0.006     HG 161
1343    0.942  0.009    QD1 161
1344    0.848  0.007    QD2 161
1345  124.573  0.031      N 161
```

1346	176.196	0.000	C	162
1347	59.465	0.122	CA	162
1348	38.054	0.067	CB	162
1349	23.438	0.003	CG1	162
1350	19.207	0.004	CG2	162
1351	8.074	0.005	H	162
1352	4.872	0.002	HA	162
1353	2.164	0.006	HB	162
1354	0.334	0.002	QG1	162
1355	-0.289	0.002	QG2	162
1356	116.068	0.011	N	162
1357	176.154	0.000	C	163
1358	61.435	0.072	CA	163
1359	72.267	0.161	CB	163
1360	23.013	0.030	CG2	163
1361	8.404	0.005	H	163
1362	4.904	0.004	HA	163
1363	4.734	0.006	HB	163
1364	1.287	0.005	QG2	163
1365	110.506	0.025	N	163
1366	179.011	0.000	C	164
1367	61.681	0.112	CA	164
1368	30.191	0.066	CB	164
1369	37.555	0.006	CG	164
1370	9.096	0.006	H	164
1371	3.834	0.009	HA	164
1372	1.991	0.003	HB2	164
1373	1.886	0.009	HB3	164
1374	2.048	0.003	HG2	164
1375	1.874	0.009	HG3	164
1376	120.861	0.012	N	164
1377	179.511	0.000	C	165
1378	60.433	0.093	CA	165
1379	31.087	0.027	CB	165
1380	44.207	0.029	CD	165
1381	26.006	0.000	CG	165
1382	8.291	0.011	H	165
1383	4.021	0.007	HA	165
1384	1.851	0.004	HB2	165
1385	1.763	0.007	HB3	165
1386	3.189	0.004	HD2	165
1387	3.191	0.005	HD3	165
1388	1.557	0.004	HG2	165
1389	1.437	0.005	HG3	165
1390	118.341	0.043	N	165
1391	177.706	0.000	C	166
1392	59.949	0.064	CA	166
1393	31.872	0.028	CB	166
1394	38.186	0.024	CG	166
1395	7.553	0.003	H	166
1396	3.788	0.003	HA	166
1397	2.229	0.004	HB2	166
1398	1.809	0.004	HB3	166
1399	2.369	0.011	HG2	166
1400	2.326	0.003	HG3	166
1401	121.042	0.003	N	166
1402	180.024	0.000	C	167
1403	55.641	0.083	CA	167
1404	21.064	0.080	CB	167
1405	8.259	0.011	H	167
1406	3.272	0.009	HA	167
1407	1.321	0.003	QB	167
1408	120.640	0.044	N	167
1409	178.704	0.000	C	168
1410	59.996	0.001	CA	168
1411	33.356	0.014	CB	168
1412	26.092	0.033	CG	168
1413	7.535	0.005	H	168
1414	3.973	0.003	HA	168
1415	1.875	0.005	HB2	168
1416	1.869	0.007	HB3	168
1417	1.558	0.007	HG2	168
1418	1.436	0.007	HG3	168
1419	117.065	0.006	N	168
1420	179.694	0.000	C	169
1421	55.795	0.068	CA	169
1422	19.795	0.127	CB	169

```
1423    7.580  0.003      H 169
1424    4.188  0.004     HA 169
1425    1.362  0.006     QB 169
1426  120.746  0.004      N 169
1427  178.507  0.000      C 170
1428   56.023  0.060     CA 170
1429   19.204  0.097     CB 170
1430    8.769  0.004      H 170
1431    3.865  0.003     HA 170
1432    1.448  0.003     QB 170
1433  118.980  0.010      N 170
1434  179.203  0.000      C 171
1435   56.351  0.093     CA 171
1436   19.188  0.111     CB 171
1437    7.925  0.002      H 171
1438    3.869  0.004     HA 171
1439    1.452  0.007     QB 171
1440  118.298  0.006      N 171
1441  179.525  0.000      C 172
1442   58.445  0.105     CA 172
1443   42.059  0.064     CB 172
1444    8.117  0.006      H 172
1445    4.239  0.009     HA 172
1446    2.818  0.010    HB2 172
1447    2.678  0.002    HB3 172
1448  118.848  0.037      N 172
1449  179.201  0.000      C 173
1450   66.398  0.082     CA 173
1451   39.670  0.061     CB 173
1452   15.094  0.012    CD1 173
1453   25.761  0.012    CG1 173
1454   18.727  0.005    CG2 173
1455    8.240  0.009      H 173
1456    3.509  0.006     HA 173
1457    1.773  0.004     HB 173
1458    1.852  0.004   HG12 173
1459    1.938  0.003   HG13 173
1460    0.830  0.003    QD1 173
1461    0.745  0.004    QG2 173
1462  120.033  0.033      N 173
1463  177.672  0.000      C 174
1464   56.380  0.035     CA 174
1465   20.197  0.038     CB 174
1466    8.197  0.005      H 174
1467    3.886  0.004     HA 174
1468    1.373  0.002     QB 174
1469  120.173  0.026      N 174
1470  174.965  0.000      C 175
1471   61.762  0.004     CA 175
1472   64.079  0.001     CB 175
1473    7.974  0.006      H 175
1474  108.010  0.009      N 175
1475  176.356  0.000      C 176
1476   57.232  0.037     CA 176
1477   31.991  0.060     CB 176
1478   44.301  0.039     CD 176
1479   29.246  0.083     CG 176
1480    7.106  0.004      H 176
1481    4.356  0.008     HA 176
1482    1.648  0.004    HB2 176
1483    2.105  0.005    HB3 176
1484    3.181  0.005    HD2 176
1485    3.043  0.001    HD3 176
1486    1.812  0.006    HG2 176
1487    1.581  0.008    HG3 176
1488  119.407  0.006      N 176
1489  176.045  0.000      C 177
1490   64.456  0.060     CA 177
1491   32.353  0.051     CB 177
1492   24.086  0.019    CG1 177
1493   22.107  0.027    CG2 177
1494    7.164  0.005      H 177
1495    3.813  0.004     HA 177
1496    2.109  0.002     HB 177
1497    0.960  0.005    QG1 177
1498    0.936  0.003    QG2 177
1499  123.370  0.015      N 177
```

1500	174.839	0.000	C	178
1501	61.111	0.068	CA	178
1502	63.956	0.104	CB	178
1503	8.614	0.006	H	178
1504	4.134	0.004	HA	178
1505	3.909	0.003	HB2	178
1506	3.908	0.002	HB3	178
1507	124.502	0.011	N	178
1508	174.218	0.000	C	179
1509	45.624	0.116	CA	179
1510	8.641	0.003	H	179
1511	3.501	0.001	HA2	179
1512	4.353	0.007	HA3	179
1513	113.033	0.027	N	179
1514	177.308	0.000	C	180
1515	65.359	0.003	CA	180
1516	33.044	0.006	CB	180
1517	7.420	0.010	H	180
1518	120.308	0.046	N	180
1519	58.138	0.044	CA	181
1520	35.204	0.037	CB	181
1521	26.124	0.035	CG	181
1522	9.219	0.019	H	181
1523	4.248	0.006	HA	181
1524	1.524	0.008	HB2	181
1525	1.535	0.012	HB3	181
1526	2.981	0.006	HE2	181
1527	1.431	0.004	HG2	181
1528	1.557	0.004	HG3	181
1529	128.378	0.074	N	181
1530	172.453	0.000	C	182
1531	56.369	0.002	CA	182
1532	34.045	0.004	CB	182
1533	44.308	0.000	CD	182
1534	7.368	0.004	H	182
1535	2.952	0.000	HD2	182
1536	116.903	0.028	N	182
1537	174.415	0.000	C	183
1538	61.776	0.004	CA	183
1539	35.232	0.004	CB	183
1540	22.595	0.014	CG1	183
1541	21.799	0.021	CG2	183
1542	8.585	0.004	H	183
1543	4.788	0.006	HA	183
1544	1.872	0.013	HB	183
1545	0.663	0.004	QG1	183
1546	0.706	0.007	QG2	183
1547	124.241	0.005	N	183
1548	173.659	0.000	C	184
1549	62.080	0.012	CA	184
1550	70.757	0.009	CB	184
1551	9.265	0.015	H	184
1552	127.394	0.045	N	184
1553	172.793	0.000	C	185
1554	60.672	0.087	CA	185
1555	70.517	0.065	CB	185
1556	22.563	0.029	CG2	185
1557	8.778	0.011	H	185
1558	4.295	0.008	HA	185
1559	3.680	0.002	HB	185
1560	0.978	0.003	QG2	185
1561	116.034	0.010	N	185
1562	176.369	0.000	C	186
1563	51.345	0.143	CA	186
1564	18.315	0.042	CB	186
1565	8.417	0.009	H	186
1566	4.511	0.004	HA	186
1567	0.575	0.006	QB	186
1568	128.899	0.012	N	186
1569	175.942	0.000	C	187
1570	58.675	0.006	CA	187
1571	41.500	0.085	CB	187
1572	132.588	0.031	CD1	187
1573	132.614	0.015	CD2	187
1574	8.001	0.005	H	187
1575	4.752	0.003	HA	187
1576	2.295	0.002	HB2	187

1577	2.346	0.006	HB3	187
1578	7.093	0.001	HD1	187
1579	7.089	0.007	HD2	187
1580	119.217	0.006	N	187
1581	172.658	0.000	C	188
1582	62.719	0.068	CA	188
1583	71.311	0.227	CB	188
1584	20.320	0.000	CG2	188
1585	8.827	0.009	H	188
1586	4.321	0.006	HA	188
1587	4.199	0.003	HB	188
1588	1.311	0.003	QG2	188
1589	118.642	0.009	N	188
1590	177.333	0.000	C	189
1591	57.927	0.116	CA	189
1592	40.986	0.102	CB	189
1593	132.119	0.037	CD1	189
1594	132.112	0.043	CD2	189
1595	130.296	0.009	CE1	189
1596	130.299	0.001	CE2	189
1597	8.584	0.008	H	189
1598	5.586	0.008	HA	189
1599	3.023	0.007	HB2	189
1600	2.799	0.008	HB3	189
1601	7.137	0.002	HD1	189
1602	7.136	0.001	HD2	189
1603	7.316	0.003	HE1	189
1604	7.315	0.005	HE2	189
1605	123.641	0.012	N	189
1606	175.978	0.000	C	190
1607	61.931	0.015	CA	190
1608	40.896	0.007	CB	190
1609	14.645	0.015	CD1	190
1610	29.123	0.016	CG1	190
1611	18.294	0.009	CG2	190
1612	8.158	0.005	H	190
1613	4.072	0.003	HA	190
1614	1.552	0.002	HB	190
1615	1.371	0.004	HG12	190
1616	0.896	0.002	HG13	190
1617	0.797	0.002	QD1	190
1618	0.855	0.002	QG2	190
1619	123.187	0.023	N	190
1620	176.830	0.000	C	191
1621	57.262	0.073	CA	191
1622	34.168	0.047	CB	191
1623	29.924	0.016	CD	191
1624	42.853	0.018	CE	191
1625	25.756	0.024	CG	191
1626	8.556	0.008	H	191
1627	4.395	0.006	HA	191
1628	1.865	0.009	HB2	191
1629	1.864	0.003	HB3	191
1630	1.702	0.006	HD2	191
1631	1.713	0.011	HD3	191
1632	2.999	0.006	HE2	191
1633	2.999	0.008	HE3	191
1634	1.465	0.007	HG2	191
1635	1.475	0.011	HG3	191
1636	126.625	0.015	N	191
1637	174.485	0.000	C	192
1638	46.216	0.064	CA	192
1639	8.496	0.008	H	192
1640	3.987	0.012	HA2	192
1641	3.973	0.003	HA3	192
1642	110.452	0.002	N	192
1643	174.313	0.000	C	193
1644	46.237	0.066	CA	193
1645	8.408	0.004	H	193
1646	3.980	0.003	HA2	193
1647	3.979	0.000	HA3	193
1648	109.018	0.019	N	193
1649	177.477	0.000	C	194
1650	56.150	0.076	CA	194
1651	43.316	0.073	CB	194
1652	24.361	0.017	CD1	194
1653	25.856	0.049	CD2	194

```
1654    8.191  0.007      H 194
1655    4.305  0.003     HA 194
1656    1.581  0.009    HB2 194
1657    1.530  0.009    HB3 194
1658    0.830  0.002    QD1 194
1659    0.890  0.002    QD2 194
1660  121.462  0.012      N 194
1661  176.233  0.000      C 195
1662   57.534  0.072     CA 195
1663   31.060  0.068     CB 195
1664   37.054  0.015     CG 195
1665    8.495  0.011      H 195
1666    4.178  0.003     HA 195
1667    1.889  0.006    HB2 195
1668    1.853  0.009    HB3 195
1669    2.199  0.006    HG2 195
1670    2.122  0.004    HG3 195
1671  121.232  0.007      N 195
1672  174.673  0.000      C 196
1673   56.648  0.099     CA 196
1674   30.775  0.158     CB 196
1675    8.345  0.003      H 196
1676    4.617  0.006     HA 196
1677    3.024  0.008    HB2 196
1678    3.120  0.005    HB3 196
1679  119.704  0.019      N 196
1680    8.423  0.006      H 197
1681  120.143  0.011      N 197
```

Appendix III NOESY assignments used for YraP structure determination

```
128 ILE  HA   131 LYS  HE2    4.32       #peak   18 #SUP  0.97  #QF  0.97
131 LYS  HG2  131 LYS  HE2    3.87       #peak   21 #SUP  0.99  #QF  0.99
190 ILE  QG2  192 GLY  HA3    3.28       #peak   23 #SUP  0.30  #QF  0.30
 45 VAL  HB    49 THR  QG2    4.10       #peak   25 #SUP  0.44  #QF  0.44
 62 GLU  HG2   66 LYS  HE2    3.79       #peak   38 #SUP  0.97  #QF  0.97
 62 GLU  HG3   66 LYS  HE2    3.38       #peak   39 #SUP  0.98  #QF  0.98
 62 GLU  HG3   66 LYS  HD3    3.31       #peak   43 #SUP  0.91  #QF  0.91
 47 ASP  HB2   74 ALA  QB     4.47       #peak   53 #SUP  0.88  #QF  0.88
 89 GLU  HG3   90 LEU  QD2    3.43       #peak   54 #SUP  0.78  #QF  0.28
 90 LEU  HA    93 ARG  HD3    3.98       #peak   60 #SUP  0.82  #QF  0.24
 99 MET  HA    99 MET  HG3    3.71       #peak   61 #SUP  0.84  #QF  0.84
 99 MET  HA    99 MET  HG2    3.71       #peak   62 #SUP  1.00
 98 ALA  QB    99 MET  HG2    4.65       #peak   75 #SUP  0.66  #QF  0.66
 99 MET  H     99 MET  HG2    4.09       #peak   77 #SUP  0.99  #QF  0.99
 53 ARG  HG2  102 ASP  HB3    3.87       #peak   78 #SUP  0.93  #QF  0.93
 53 ARG  HG2  102 ASP  HB2    3.90       #peak   80 #SUP  0.94  #QF  0.94
 77 GLY  HA3  105 ASN  HB2    5.50       #peak   83 #SUP  0.92  #QF  0.88
 77 GLY  HA3  105 ASN  HB3    5.50       #peak   83 #SUP  0.92  #QF  0.88
 62 GLU  HG3   66 LYS  HG2    4.12       #peak   85 #SUP  0.94  #QF  0.94
 37 ASP  HA    38 PRO  HA     4.55       #peak   86 #SUP  0.72  #QF  0.72
 92 ALA  QB    96 GLN  HG3    4.27       #peak  101 #SUP  0.80  #QF  0.80
 96 GLN  HG3   97 ILE  H      3.66       #peak  103 #SUP  0.98  #QF  0.98
 96 GLN  H     96 GLN  HG3    3.42       #peak  104 #SUP  0.99  #QF  0.99
 93 ARG  H     96 GLN  HG3    4.51       #peak  105 #SUP  1.00  #QF  0.86
 94 ALA  H     96 GLN  HG3    4.85       #peak  105 #SUP  1.00  #QF  0.86
131 LYS  HD2  135 GLN  HG3    5.50       #peak  126 #SUP  0.80  #QF  0.36
135 GLN  HG3  138 THR  QG2    5.08       #peak  127 #SUP  0.85  #QF  0.85
134 SER  HB3  135 GLN  HG3    4.09       #peak  130 #SUP  0.77  #QF  0.77
135 GLN  HA   135 GLN  HG3    3.56       #peak  131 #SUP  0.99  #QF  0.99
141 LEU  HB3  141 LEU  QD1    3.34       #peak  133 #SUP  0.98  #QF  0.98
129 THR  QG2  149 VAL  HB     4.14       #peak  145 #SUP  0.84  #QF  0.84
151 THR  QG2  156 VAL  HB     4.33       #peak  148 #SUP  0.89  #QF  0.89
```

```
 75 TYR  HE1   161 LEU  QD1    4.64       #peak   150 #SUP  0.88  #QF  0.88
 87 ASN  HB2    90 LEU  QD1    4.25       #peak   173 #SUP  0.75  #QF  0.75
163 THR  HA    190 ILE  HB     3.50       #peak   183 #SUP  1.00  #QF  0.42
163 THR  HB    190 ILE  HB     3.96       #peak   184 #SUP  0.99  #QF  0.99
190 ILE  HA    190 ILE  HG12   3.38       #peak   188 #SUP  0.92  #QF  0.92
195 GLU  HA    196 HIS  HA     4.57       #peak   195 #SUP  0.85  #QF  0.85
195 GLU  HA    195 GLU  HG3    3.84       #peak   197 #SUP  0.96  #QF  0.96
 20 ILE  QG2    22 ALA  QB     3.32       #peak   211 #SUP  0.22  #QF  0.22
 20 ILE  HA     20 ILE  QG2    3.13       #peak   213 #SUP  0.99  #QF  0.99
 20 ILE  HB     20 ILE  QD1    3.28       #peak   217 #SUP  0.96  #QF  0.96
 20 ILE  HA     20 ILE  QD1    3.85       #peak   218 #SUP  0.97  #QF  0.97
 36 THR  HA     37 ASP  HA     4.46       #peak   228 #SUP  0.95  #QF  0.95
 52 VAL  H      52 VAL  QG1    3.98       #peak   244 #SUP  0.94  #QF  0.94
 52 VAL  QG1    53 ARG  H      3.73       #peak   245 #SUP  0.99  #QF  0.99
 46 ASP  HB3    49 THR  QG2    4.38       #peak   262 #SUP  0.96  #QF  0.96
130 THR  HA    133 ARG  H      3.86       #peak   267 #SUP  0.90  #QF  0.90
130 THR  HA    132 VAL  QG2    4.70       #peak   273 #SUP  0.72  #QF  0.72
126 THR  HA    129 THR  H      4.00       #peak   287 #SUP  0.94  #QF  0.94
126 THR  HA    129 THR  QG2    3.65       #peak   289 #SUP  0.91  #QF  0.91
 61 ASP  HB2    64 ILE  QD1    3.21       #peak   292 #SUP  0.93  #QF  0.93
 67 GLU  HA     67 GLU  HG2    3.90       #peak   294 #SUP  0.98  #QF  0.98
 68 ALA  QB     90 LEU  HB2    3.44       #peak   300 #SUP  0.71  #QF  0.71
 64 ILE  HB     68 ALA  QB     4.67       #peak   301 #SUP  0.99  #QF  0.97
 67 GLU  HB2    68 ALA  QB     5.07       #peak   301 #SUP  0.99  #QF  0.97
 64 ILE  HA     68 ALA  QB     3.20       #peak   303 #SUP  0.95  #QF  0.95
 68 ALA  QB     69 ARG  H      3.51       #peak   306 #SUP  0.94  #QF  0.94
 68 ALA  QB     85 SER  H      5.13       #peak   307 #SUP  0.90  #QF  0.90
 68 ALA  H      68 ALA  QB     3.55       #peak   308 #SUP  0.85  #QF  0.85
 70 ILE  HA     70 ILE  QG2    3.22       #peak   309 #SUP  1.00
 70 ILE  QG2    71 ASN  HA     4.28       #peak   310 #SUP  0.98  #QF  0.98
156 VAL  QG2   157 PHE  H      4.26       #peak   314 #SUP  0.92  #QF  0.92
 73 THR  QG2    74 ALA  HA     4.32       #peak   334 #SUP  0.97  #QF  0.97
 73 THR  QG2    80 LEU  H      3.98       #peak   337 #SUP  0.92  #QF  0.92
 73 THR  QG2    75 TYR  H      4.75       #peak   338 #SUP  0.90  #QF  0.90
 73 THR  H      73 THR  QG2    3.94       #peak   339 #SUP  0.99  #QF  0.99
148 LYS  HG2   149 VAL  H      5.50       #peak   340 #SUP  0.93  #QF  0.44
148 LYS  HG3   149 VAL  H      5.50       #peak   340 #SUP  0.93  #QF  0.44
 75 TYR  HD2    79 VAL  HA     5.03       #peak   345 #SUP  0.46  #QF  0.46
 87 ASN  HB3    89 GLU  HG2    4.11       #peak   356 #SUP  0.99  #QF  0.97
 95 LYS  HA     98 ALA  QB     3.79       #peak   360 #SUP  0.95  #QF  0.57
174 ALA  HA    177 VAL  HB     4.70       #peak   363 #SUP  0.90  #QF  0.56
 98 ALA  QB     99 MET  H      3.38       #peak   367 #SUP  0.97  #QF  0.44
 58 LEU  H      98 ALA  QB     4.41       #peak   368 #SUP  0.29  #QF  0.29
101 VAL  HA    102 ASP  HB3    4.64       #peak   370 #SUP  0.91  #QF  0.91
102 ASP  HB3   104 ALA  QB     4.46       #peak   373 #SUP  0.80  #QF  0.80
 53 ARG  HG3   102 ASP  HB2    4.46       #peak   374 #SUP  0.82  #QF  0.82
 50 LEU  HG    104 ALA  QB     4.84       #peak   381 #SUP  0.68  #QF  0.68
 99 MET  HA    104 ALA  QB     3.80       #peak   383 #SUP  0.99  #QF  0.99
104 ALA  QB    106 GLU  HA     5.14       #peak   384 #SUP  0.52  #QF  0.52
104 ALA  QB    105 ASN  HA     4.48       #peak   385 #SUP  0.84  #QF  0.84
162 VAL  QG2   163 THR  QG2    4.71       #peak   411 #SUP  0.95  #QF  0.95
 37 ASP  HA     38 PRO  HB3    4.89       #peak   423 #SUP  0.97  #QF  0.97
135 GLN  HG2   138 THR  QG2    4.08       #peak   442 #SUP  0.79  #QF  0.79
 59 SER  HA     65 LYS  HG2    3.41       #peak   448 #SUP  0.43  #QF  0.43
147 VAL  HA    147 VAL  QG2    3.33       #peak   458 #SUP  0.97  #QF  0.97
147 VAL  HA    147 VAL  QG1    3.33       #peak   459 #SUP  0.98  #QF  0.98
147 VAL  HA    160 GLY  HA2    4.34       #peak   461 #SUP  0.97  #QF  0.97
147 VAL  HA    160 GLY  HA3    3.87       #peak   462 #SUP  0.93  #QF  0.93
147 VAL  QG2   148 LYS  H      4.31       #peak   471 #SUP  0.80  #QF  0.80
156 VAL  H     156 VAL  QG1    3.96       #peak   486 #SUP  0.77  #QF  0.77
117 ILE  HB    153 ASN  HA     4.55       #peak   490 #SUP  0.56  #QF  0.30
117 ILE  QD1   153 ASN  HA     4.67       #peak   491 #SUP  0.98  #QF  0.94
128 ILE  HB    151 THR  HB     5.02       #peak   493 #SUP  0.92  #QF  0.68
162 VAL  HB    166 GLU  HA     5.50       #peak   493 #SUP  0.92  #QF  0.68
169 ALA  QB    172 ASP  HB2    4.17       #peak   507 #SUP  0.96  #QF  0.96
183 VAL  QG2   184 THR  H      4.27       #peak   512 #SUP  0.72  #QF  0.72
189 PHE  HA    189 PHE  HD1    3.98       #peak   527 #SUP  0.84  #QF  0.84
162 VAL  QG2   189 PHE  HA     4.08       #peak   529 #SUP  0.89  #QF  0.89
163 THR  HA    190 ILE  QG2    3.41       #peak   531 #SUP  0.98  #QF  0.98
163 THR  HB    190 ILE  QG2    3.90       #peak   532 #SUP  0.99  #QF  0.99
190 ILE  QG2   191 LYS  HA     4.24       #peak   534 #SUP  0.67  #QF  0.67
```

161 LEU HA 190 ILE QD1 4.76 #peak 541 #SUP 1.00
163 THR HB 190 ILE QD1 4.44 #peak 543 #SUP 0.98 #QF 0.98
162 VAL QG1 190 ILE QD1 4.36 #peak 547 #SUP 0.93 #QF 0.93
162 VAL QG2 190 ILE QD1 4.96 #peak 548 #SUP 0.98 #QF 0.98
52 VAL HA 55 ASN HB2 4.08 #peak 568 #SUP 0.99 #QF 0.99
128 ILE H 128 ILE QD1 3.16 #peak 593 #SUP 0.91 #QF 0.91
117 ILE QD1 121 GLU HG2 4.26 #peak 595 #SUP 0.95 #QF 0.95
78 LYS HD2 79 VAL H 4.85 #peak 597 #SUP 0.81 #QF 0.81
61 ASP HB3 64 ILE HB 3.99 #peak 606 #SUP 0.99 #QF 0.99
61 ASP HA 64 ILE HB 4.39 #peak 608 #SUP 0.98 #QF 0.98
96 GLN HA 96 GLN HG3 3.86 #peak 612 #SUP 0.94 #QF 0.94
162 VAL HA 162 VAL QG2 3.59 #peak 620 #SUP 0.97 #QF 0.97
81 LEU H 81 LEU QD2 4.42 #peak 650 #SUP 0.92 #QF 0.92
125 ASP HA 129 THR H 4.59 #peak 657 #SUP 0.86 #QF 0.78
81 LEU H 81 LEU QD1 4.42 #peak 664 #SUP 0.86 #QF 0.86
81 LEU QD1 82 VAL H 5.09 #peak 665 #SUP 0.98 #QF 0.98
52 VAL H 52 VAL QG2 3.37 #peak 691 #SUP 0.95 #QF 0.95
90 LEU H 90 LEU QD1 4.63 #peak 693 #SUP 0.98 #QF 0.98
129 THR QG2 132 VAL H 5.01 #peak 707 #SUP 0.96 #QF 0.96
129 THR QG2 130 THR H 4.41 #peak 708 #SUP 0.97 #QF 0.97
129 THR QG2 149 VAL HA 5.02 #peak 710 #SUP 0.99 #QF 0.99
78 LYS HA 78 LYS HD3 4.00 #peak 717 #SUP 0.96 #QF 0.96
162 VAL QG1 167 ALA H 4.82 #peak 719 #SUP 0.99 #QF 0.99
162 VAL QG1 166 GLU HB3 4.23 #peak 725 #SUP 0.75 #QF 0.75
149 VAL HA 149 VAL QG1 3.24 #peak 742 #SUP 1.00
159 MET HA 187 PHE HA 3.94 #peak 751 #SUP 0.70 #QF 0.70
82 VAL HA 82 VAL QG1 3.75 #peak 754 #SUP 1.00
81 LEU HA 82 VAL QG1 4.43 #peak 763 #SUP 0.95 #QF 0.91
82 VAL QG1 109 ASN HA 4.87 #peak 763 #SUP 0.95 #QF 0.91
82 VAL QG1 83 GLY H 4.24 #peak 766 #SUP 1.00
82 VAL QG1 150 THR QG2 3.56 #peak 767 #SUP 0.59 #QF 0.59
82 VAL H 82 VAL QG2 3.39 #peak 784 #SUP 0.97 #QF 0.97
82 VAL QG2 83 GLY H 4.25 #peak 785 #SUP 1.00
81 LEU HA 82 VAL QG2 3.75 #peak 789 #SUP 0.76 #QF 0.76
82 VAL HA 82 VAL QG2 3.84 #peak 792 #SUP 1.00
71 ASN HB2 82 VAL QG2 3.43 #peak 793 #SUP 0.99 #QF 0.99
71 ASN HB3 82 VAL QG2 3.43 #peak 794 #SUP 0.95 #QF 0.95
111 ILE H 111 ILE QG2 3.88 #peak 799 #SUP 0.83 #QF 0.83
69 ARG H 111 ILE QG2 4.52 #peak 800 #SUP 0.95 #QF 0.95
85 SER H 111 ILE QG2 3.35 #peak 801 #SUP 0.96 #QF 0.81
70 ILE HA 111 ILE QG2 4.40 #peak 802 #SUP 0.36 #QF 0.36
82 VAL QG2 83 GLY HA3 4.87 #peak 805 #SUP 0.97 #QF 0.97
82 VAL HA 111 ILE QG2 4.62 #peak 806 #SUP 0.99 #QF 0.99
111 ILE HA 111 ILE QG2 3.12 #peak 808 #SUP 0.74 #QF 0.74
162 VAL QG2 189 PHE HD1 4.22 #peak 816 #SUP 0.90 #QF 0.90
162 VAL QG2 187 PHE HA 4.90 #peak 820 #SUP 0.97 #QF 0.97
109 ASN HA 111 ILE QD1 4.27 #peak 866 #SUP 0.73 #QF 0.73
111 ILE H 111 ILE QD1 3.96 #peak 867 #SUP 0.87 #QF 0.87
85 SER H 111 ILE QD1 5.02 #peak 868 #SUP 0.99 #QF 0.71
111 ILE QD1 112 ARG H 5.37 #peak 868 #SUP 0.99 #QF 0.71
82 VAL H 111 ILE QD1 5.50 #peak 870 #SUP 0.94 #QF 0.94
111 ILE HA 111 ILE QD1 3.87 #peak 873 #SUP 0.98 #QF 0.98
111 ILE HB 111 ILE QD1 3.36 #peak 874 #SUP 1.00
77 GLY HA2 105 ASN HB3 5.50 #peak 899 #SUP 0.96 #QF 0.72
53 ARG HB2 53 ARG HD3 3.88 #peak 908 #SUP 1.00
53 ARG HB3 53 ARG HD3 3.88 #peak 909 #SUP 1.00
49 THR HA 53 ARG HD3 4.54 #peak 917 #SUP 0.97 #QF 0.97
78 LYS HE3 108 TYR HE1 4.03 #peak 938 #SUP 0.83 #QF 0.83
148 LYS HE2 149 VAL H 5.50 #peak 939 #SUP 0.71 #QF 0.32
148 LYS HE3 149 VAL H 5.50 #peak 939 #SUP 0.71 #QF 0.32
75 TYR HE2 78 LYS HE2 4.89 #peak 941 #SUP 0.96 #QF 0.96
63 GLN HA 66 LYS HE2 3.46 #peak 943 #SUP 0.82 #QF 0.57
63 GLN HA 66 LYS HE3 3.67 #peak 943 #SUP 0.82 #QF 0.57
127 TRP HZ2 131 LYS HE3 4.50 #peak 945 #SUP 0.99 #QF 0.99
127 TRP HD1 131 LYS HE3 5.50 #peak 946 #SUP 0.97 #QF 0.97
128 ILE HA 131 LYS HE3 4.32 #peak 948 #SUP 0.98 #QF 0.98
131 LYS HG2 131 LYS HE3 3.87 #peak 949 #SUP 1.00
127 TRP HZ2 131 LYS HE2 4.50 #peak 951 #SUP 0.98 #QF 0.98
127 TRP HD1 131 LYS HE2 5.50 #peak 952 #SUP 0.90 #QF 0.90
47 ASP HB2 48 GLY HA2 4.23 #peak 957 #SUP 0.78 #QF 0.78
125 ASP HB3 151 THR QG2 4.26 #peak 977 #SUP 0.99 #QF 0.99

146 ASN HB3 147 VAL HB 4.86 #peak 984 #SUP 0.81 #QF 0.81
146 ASN HB3 162 VAL QG1 4.32 #peak 985 #SUP 0.90 #QF 0.90
146 ASN HB3 147 VAL QG1 5.50 #peak 986 #SUP 0.97 #QF 0.97
173 ILE H 173 ILE HB 3.87 #peak 988 #SUP 0.40 #QF 0.40
125 ASP HA 128 ILE HB 4.07 #peak 1012 #SUP 0.33 #QF 0.33
128 ILE HB 151 THR QG2 3.96 #peak 1014 #SUP 0.73 #QF 0.73
128 ILE HB 128 ILE QD1 3.46 #peak 1015 #SUP 0.63 #QF 0.63
67 GLU HG2 90 LEU QD2 5.28 #peak 1016 #SUP 0.84 #QF 0.84
125 ASP HB2 151 THR HB 4.31 #peak 1033 #SUP 0.97 #QF 0.97
125 ASP HB3 151 THR HB 4.31 #peak 1034 #SUP 0.95 #QF 0.95
151 THR HB 152 GLU H 4.56 #peak 1036 #SUP 0.96 #QF 0.96
78 LYS HA 106 GLU HG2 5.09 #peak 1038 #SUP 1.00 #QF 0.76
130 THR HB 131 LYS H 3.93 #peak 1044 #SUP 0.97 #QF 0.97
126 THR HB 127 TRP HB3 4.89 #peak 1046 #SUP 0.51 #QF 0.51
125 ASP HA 151 THR HB 3.78 #peak 1050 #SUP 0.84 #QF 0.84
129 THR H 129 THR HB 3.51 #peak 1051 #SUP 0.82 #QF 0.82
148 LYS HB3 148 LYS HD2 4.22 #peak 1055 #SUP 1.00
129 THR HA 129 THR QG2 3.50 #peak 1071 #SUP 0.99 #QF 0.99
135 GLN HG2 136 LEU H 5.50 #peak 1074 #SUP 0.98 #QF 0.96
96 GLN HG3 98 ALA H 4.95 #peak 1075 #SUP 0.98 #QF 0.98
95 LYS H 96 GLN HG3 4.52 #peak 1081 #SUP 0.72 #QF 0.72
49 THR HA 52 VAL HB 3.30 #peak 1083 #SUP 0.96 #QF 0.96
49 THR HA 52 VAL QG1 4.02 #peak 1085 #SUP 0.87 #QF 0.87
173 ILE HA 174 ALA HA 4.94 #peak 1090 #SUP 0.72 #QF 0.72
173 ILE HA 173 ILE QG2 3.20 #peak 1098 #SUP 0.99 #QF 0.99
131 LYS H 131 LYS HB3 4.15 #peak 1104 #SUP 0.93 #QF 0.93
131 LYS HB2 132 VAL H 4.38 #peak 1112 #SUP 1.00
162 VAL QG1 166 GLU HB2 4.23 #peak 1123 #SUP 1.00
162 VAL QG2 166 GLU HB2 4.71 #peak 1124 #SUP 0.87 #QF 0.87
83 GLY H 111 ILE HA 4.71 #peak 1128 #SUP 0.91 #QF 0.91
151 THR HA 157 PHE H 4.47 #peak 1144 #SUP 0.93 #QF 0.93
151 THR HA 156 VAL HA 4.27 #peak 1146 #SUP 0.99 #QF 0.99
127 TRP HZ2 131 LYS HD2 4.30 #peak 1160 #SUP 0.93 #QF 0.93
127 TRP HZ2 131 LYS HD3 4.20 #peak 1161 #SUP 0.94 #QF 0.94
96 GLN H 96 GLN HB2 3.87 #peak 1165 #SUP 1.00 #QF 0.98
96 GLN H 96 GLN HB3 4.18 #peak 1165 #SUP 1.00 #QF 0.98
67 GLU HG3 68 ALA QB 3.92 #peak 1196 #SUP 0.38 #QF 0.38
167 ALA H 167 ALA QB 3.13 #peak 1202 #SUP 0.99 #QF 0.99
150 THR H 150 THR QG2 3.10 #peak 1208 #SUP 0.91 #QF 0.91
122 ALA QB 123 SER HA 3.96 #peak 1224 #SUP 0.72 #QF 0.72
129 THR H 130 THR QG2 4.53 #peak 1225 #SUP 0.28 #QF 0.28
130 THR H 130 THR QG2 3.79 #peak 1227 #SUP 0.94 #QF 0.94
53 ARG H 53 ARG HG3 4.17 #peak 1280 #SUP 0.97 #QF 0.97
151 THR QG2 157 PHE H 5.27 #peak 1295 #SUP 0.89 #QF 0.89
151 THR QG2 152 GLU H 3.37 #peak 1296 #SUP 0.73 #QF 0.73
151 THR HA 151 THR QG2 3.49 #peak 1299 #SUP 0.99 #QF 0.99
125 ASP HB2 151 THR QG2 4.26 #peak 1302 #SUP 0.98 #QF 0.98
46 ASP HB2 49 THR HB 3.95 #peak 1307 #SUP 0.99 #QF 0.99
46 ASP HB3 49 THR HB 3.95 #peak 1312 #SUP 0.96 #QF 0.96
172 ASP HB3 173 ILE QD1 3.72 #peak 1314 #SUP 0.96 #QF 0.96
172 ASP HB2 173 ILE QD1 3.68 #peak 1315 #SUP 0.96 #QF 0.96
47 ASP HB3 74 ALA QB 3.60 #peak 1316 #SUP 0.46 #QF 0.46
172 ASP HB2 173 ILE HG12 4.27 #peak 1317 #SUP 0.98 #QF 0.98
172 ASP HB3 173 ILE HG12 4.39 #peak 1318 #SUP 0.90 #QF 0.90
172 ASP HB2 173 ILE H 3.50 #peak 1325 #SUP 0.97 #QF 0.97
172 ASP HB3 173 ILE H 3.45 #peak 1326 #SUP 0.97 #QF 0.97
172 ASP H 172 ASP HB3 3.91 #peak 1327 #SUP 0.99 #QF 0.99
51 GLU HA 51 GLU HG2 3.98 #peak 1328 #SUP 1.00
51 GLU HA 54 VAL QG1 4.49 #peak 1332 #SUP 0.96 #QF 0.96
51 GLU HA 55 ASN H 4.73 #peak 1336 #SUP 0.91 #QF 0.91
51 GLU HA 54 VAL H 4.60 #peak 1337 #SUP 0.86 #QF 0.86
51 GLU HG2 74 ALA QB 3.81 #peak 1343 #SUP 0.99 #QF 0.99
55 ASN HB3 56 SER HA 4.29 #peak 1355 #SUP 0.95 #QF 0.95
52 VAL HA 55 ASN HB3 3.48 #peak 1357 #SUP 0.99 #QF 0.99
55 ASN HB3 58 LEU H 5.50 #peak 1362 #SUP 0.93 #QF 0.93
55 ASN H 55 ASN HB2 3.53 #peak 1363 #SUP 0.97 #QF 0.97
55 ASN H 55 ASN HB3 3.47 #peak 1364 #SUP 0.97 #QF 0.97
52 VAL QG2 55 ASN HB2 5.41 #peak 1365 #SUP 0.99 #QF 0.99
52 VAL QG2 55 ASN HB3 5.20 #peak 1372 #SUP 1.00
56 SER HB2 57 ALA H 3.93 #peak 1374 #SUP 0.93 #QF 0.93
91 SER HB2 111 ILE QD1 3.27 #peak 1377 #SUP 0.35 #QF 0.35

```
 57 ALA  HA    60 LYS  HE2    5.46       #peak  1383 #SUP  0.92  #QF  0.85
 57 ALA  HA    60 LYS  HE3    5.46       #peak  1383 #SUP  0.92  #QF  0.85
 57 ALA  HA    60 LYS  HB2    4.04       #peak  1384 #SUP  0.97  #QF  0.97
 57 ALA  HA    60 LYS  HB3    4.04       #peak  1385 #SUP  0.67  #QF  0.67
 58 LEU  HB2   58 LEU  QD1    4.04       #peak  1392 #SUP  1.00
 58 LEU  HB2   58 LEU  QD2    4.04       #peak  1393 #SUP  1.00  #QF  0.99
 58 LEU  HB3   58 LEU  QD2    4.04       #peak  1397 #SUP  0.99  #QF  0.97
 58 LEU  HB3   58 LEU  QD1    4.04       #peak  1398 #SUP  0.62  #QF  0.62
 58 LEU  H     58 LEU  HG     4.83       #peak  1406 #SUP  0.99  #QF  0.99
161 LEU  QD2  162 VAL  HA     5.14       #peak  1409 #SUP  0.97  #QF  0.97
 20 ILE  HA    21 ALA  HA     4.59       #peak  1418 #SUP  0.69  #QF  0.69
 20 ILE  HA    20 ILE  HG13   3.71       #peak  1420 #SUP  0.98  #QF  0.98
 57 ALA  QB    60 LYS  HB2    4.34       #peak  1436 #SUP  0.85  #QF  0.85
 61 ASP  HA    64 ILE  QD1    4.19       #peak  1457 #SUP  0.72  #QF  0.72
 62 GLU  HA    62 GLU  HG3    3.99       #peak  1463 #SUP  0.89  #QF  0.89
 62 GLU  HB2   65 LYS  HG3    4.31       #peak  1468 #SUP  0.81  #QF  0.81
 61 ASP  HA    62 GLU  HB3    4.85       #peak  1471 #SUP  0.84  #QF  0.84
 63 GLN  HG3   64 ILE  HA     4.86       #peak  1472 #SUP  0.92  #QF  0.92
131 LYS  HA   135 GLN  HG3    4.11       #peak  1473 #SUP  0.75  #QF  0.75
 63 GLN  HB2   90 LEU  QD1    3.27       #peak  1485 #SUP  0.84  #QF  0.84
 61 ASP  HB3   64 ILE  QD1    3.39       #peak  1495 #SUP  0.97  #QF  0.95
 64 ILE  HA    64 ILE  QD1    3.86       #peak  1496 #SUP  0.96  #QF  0.96
 64 ILE  QD1   94 ALA  HA     3.48       #peak  1497 #SUP  0.81  #QF  0.81
 64 ILE  QD1   94 ALA  H      3.68       #peak  1500 #SUP  0.87  #QF  0.87
 64 ILE  H     64 ILE  QD1    4.86       #peak  1501 #SUP  1.00  #QF  0.95
163 THR  H    190 ILE  QD1    5.16       #peak  1501 #SUP  1.00  #QF  0.95
190 ILE  QD1  191 LYS  H      4.74       #peak  1502 #SUP  0.33  #QF  0.33
 62 GLU  HA    65 LYS  HG3    3.16       #peak  1505 #SUP  0.98  #QF  0.60
 65 LYS  HE2   66 LYS  HA     5.50       #peak  1508 #SUP  0.83  #QF  0.43
 65 LYS  HE3   66 LYS  HA     5.50       #peak  1508 #SUP  0.83  #QF  0.43
 66 LYS  HA    66 LYS  HG2    3.87       #peak  1532 #SUP  1.00
 66 LYS  HG2   67 GLU  HG2    3.54       #peak  1534 #SUP  0.99  #QF  0.99
 66 LYS  HG2   67 GLU  HG3    3.71       #peak  1535 #SUP  0.99  #QF  0.99
 62 GLU  HA    65 LYS  HD3    3.30       #peak  1538 #SUP  0.83  #QF  0.83
 64 ILE  H     66 LYS  HD3    5.50       #peak  1542 #SUP  0.45  #QF  0.45
 66 LYS  HA    67 GLU  HG2    5.14       #peak  1546 #SUP  0.98  #QF  0.98
 63 GLN  HA    67 GLU  HG2    3.73       #peak  1547 #SUP  0.37  #QF  0.37
 66 LYS  HD3   67 GLU  HG2    4.43       #peak  1550 #SUP  0.97  #QF  0.84
 66 LYS  HD2   67 GLU  HG2    4.71       #peak  1550 #SUP  0.97  #QF  0.84
 67 GLU  HA    68 ALA  HA     4.92       #peak  1555 #SUP  0.65  #QF  0.65
 68 ALA  HA    86 PRO  HD2    4.62       #peak  1556 #SUP  0.98  #QF  0.88
 68 ALA  HA    86 PRO  HD3    4.62       #peak  1556 #SUP  0.98  #QF  0.88
 67 GLU  HB2   68 ALA  HA     5.06       #peak  1560 #SUP  0.86  #QF  0.86
 68 ALA  HA    90 LEU  HB2    5.50       #peak  1561 #SUP  0.82  #QF  0.75
 68 ALA  HA    90 LEU  QD1    4.42       #peak  1564 #SUP  0.68  #QF  0.68
168 LYS  HB3  170 ALA  H      5.50       #peak  1572 #SUP  0.78  #QF  0.78
 57 ALA  QB    60 LYS  HB3    4.34       #peak  1578 #SUP  0.71  #QF  0.47
 70 ILE  HA    83 GLY  HA3    4.48       #peak  1594 #SUP  0.88  #QF  0.88
 70 ILE  HB    70 ILE  QD1    3.32       #peak  1596 #SUP  1.00
 70 ILE  QG2   83 GLY  HA3    4.86       #peak  1611 #SUP  0.92  #QF  0.92
 70 ILE  QD1   90 LEU  HB2    4.50       #peak  1620 #SUP  0.72  #QF  0.72
 64 ILE  HB    70 ILE  QD1    4.19       #peak  1621 #SUP  0.63  #QF  0.63
 64 ILE  HA    70 ILE  QD1    3.81       #peak  1622 #SUP  0.97  #QF  0.97
 70 ILE  QD1   91 SER  HA     3.57       #peak  1623 #SUP  0.91  #QF  0.91
 70 ILE  QD1   94 ALA  HA     5.15       #peak  1624 #SUP  0.76  #QF  0.76
 65 LYS  HA    70 ILE  QD1    4.76       #peak  1625 #SUP  0.43  #QF  0.43
 68 ALA  HA    70 ILE  QD1    4.23       #peak  1627 #SUP  0.66  #QF  0.66
 70 ILE  QD1   71 ASN  HA     5.50       #peak  1628 #SUP  0.64  #QF  0.64
 70 ILE  QD1   83 GLY  HA3    5.07       #peak  1629 #SUP  0.88  #QF  0.88
 70 ILE  QD1   95 LYS  H      5.49       #peak  1631 #SUP  0.96  #QF  0.96
 70 ILE  QD1   94 ALA  H      4.96       #peak  1633 #SUP  0.94  #QF  0.94
 70 ILE  H     70 ILE  QD1    4.68       #peak  1634 #SUP  0.91  #QF  0.91
 70 ILE  QD1   71 ASN  H      5.13       #peak  1637 #SUP  0.99  #QF  0.99
 71 ASN  HB3   82 VAL  H      5.01       #peak  1648 #SUP  0.95  #QF  0.95
 51 GLU  HA    72 VAL  HB     5.14       #peak  1660 #SUP  0.87  #QF  0.87
 72 VAL  HB    73 THR  H      4.42       #peak  1662 #SUP  0.96  #QF  0.96
133 ARG  H    149 VAL  HB     5.50       #peak  1663 #SUP  0.44  #QF  0.25
 72 VAL  HA    73 THR  HA     4.63       #peak  1666 #SUP  0.69  #QF  0.69
 51 GLU  HG2   73 THR  HA     4.31       #peak  1668 #SUP  0.99  #QF  0.99
 51 GLU  HG3   73 THR  HA     4.31       #peak  1669 #SUP  0.98  #QF  0.98
 72 VAL  HA    73 THR  HB     4.84       #peak  1674 #SUP  0.74  #QF  0.54
```

73 THR HB 81 LEU HA 5.10 #peak 1674 #SUP 0.74 #QF 0.54
73 THR H 73 THR HB 3.70 #peak 1675 #SUP 0.95 #QF 0.95
73 THR HB 80 LEU H 4.14 #peak 1676 #SUP 0.96 #QF 0.96
73 THR HB 80 LEU HB3 3.43 #peak 1677 #SUP 0.84 #QF 0.84
73 THR QG2 75 TYR HE2 5.50 #peak 1683 #SUP 0.69 #QF 0.69
73 THR QG2 75 TYR HB3 4.41 #peak 1685 #SUP 0.63 #QF 0.63
74 ALA HA 79 VAL HA 3.96 #peak 1689 #SUP 0.94 #QF 0.94
74 ALA HA 79 VAL QG1 4.46 #peak 1693 #SUP 0.94 #QF 0.43
74 ALA HA 79 VAL QG2 4.46 #peak 1693 #SUP 0.94 #QF 0.43
74 ALA HA 80 LEU H 4.32 #peak 1696 #SUP 0.92 #QF 0.92
73 THR HA 74 ALA QB 4.16 #peak 1698 #SUP 0.69 #QF 0.69
51 GLU HA 74 ALA QB 3.74 #peak 1700 #SUP 0.86 #QF 0.86
51 GLU HG3 74 ALA QB 3.81 #peak 1703 #SUP 0.97 #QF 0.97
74 ALA QB 78 LYS H 5.50 #peak 1706 #SUP 0.58 #QF 0.58
74 ALA QB 75 TYR H 3.36 #peak 1707 #SUP 0.90 #QF 0.90
74 ALA QB 80 LEU H 4.68 #peak 1708 #SUP 0.82 #QF 0.82
75 TYR HE1 76 GLN HB2 5.05 #peak 1718 #SUP 0.94 #QF 0.94
75 TYR HE1 76 GLN HG2 4.07 #peak 1724 #SUP 0.87 #QF 0.87
75 TYR HE1 76 GLN HG3 4.07 #peak 1735 #SUP 0.82 #QF 0.82
78 LYS HA 79 VAL HA 4.87 #peak 1745 #SUP 0.72 #QF 0.49
79 VAL HA 80 LEU HA 5.22 #peak 1745 #SUP 0.72 #QF 0.49
72 VAL QG1 73 THR H 4.16 #peak 1752 #SUP 0.81 #QF 0.81
51 GLU HA 54 VAL QG2 4.49 #peak 1755 #SUP 0.90 #QF 0.51
66 LYS HG3 67 GLU HG2 3.48 #peak 1758 #SUP 0.97 #QF 0.97
86 PRO HA 87 ASN HA 4.58 #peak 1796 #SUP 0.60 #QF 0.60
87 ASN HA 88 ALA QB 4.02 #peak 1814 #SUP 0.87 #QF 0.87
171 ALA QB 173 ILE H 5.48 #peak 1820 #SUP 0.91 #QF 0.58
170 ALA QB 174 ALA H 5.50 #peak 1820 #SUP 0.91 #QF 0.58
171 ALA QB 174 ALA H 5.50 #peak 1820 #SUP 0.91 #QF 0.58
87 ASN HB2 88 ALA QB 5.18 #peak 1822 #SUP 0.84 #QF 0.84
171 ALA QB 172 ASP HB2 4.43 #peak 1823 #SUP 0.44 #QF 0.44
87 ASN HB2 89 GLU HB2 5.12 #peak 1836 #SUP 0.65 #QF 0.65
87 ASN HB2 89 GLU HG2 4.37 #peak 1842 #SUP 1.00 #QF 0.98
87 ASN HB2 89 GLU HG3 4.75 #peak 1842 #SUP 1.00 #QF 0.98
87 ASN HA 89 GLU HG2 5.04 #peak 1843 #SUP 0.96 #QF 0.88
89 GLU HG3 93 ARG H 5.50 #peak 1845 #SUP 0.28 #QF 0.28
90 LEU HA 90 LEU HG 3.56 #peak 1849 #SUP 0.92 #QF 0.92
64 ILE H 90 LEU QD1 4.75 #peak 1865 #SUP 0.54 #QF 0.54
91 SER HA 94 ALA H 4.28 #peak 1880 #SUP 0.93 #QF 0.93
64 ILE HA 91 SER HA 5.50 #peak 1881 #SUP 0.36 #QF 0.36
90 LEU HB2 91 SER HA 4.08 #peak 1882 #SUP 0.90 #QF 0.90
91 SER HA 92 ALA QB 5.11 #peak 1883 #SUP 0.71 #QF 0.71
91 SER HA 94 ALA QB 3.51 #peak 1884 #SUP 0.77 #QF 0.77
70 ILE HG12 91 SER HA 3.32 #peak 1885 #SUP 0.74 #QF 0.74
91 SER HA 111 ILE QG2 4.23 #peak 1886 #SUP 0.79 #QF 0.79
91 SER HB2 94 ALA H 5.50 #peak 1895 #SUP 0.42 #QF 0.42
170 ALA HA 173 ILE HG12 4.07 #peak 1897 #SUP 0.41 #QF 0.41
88 ALA QB 91 SER H 4.83 #peak 1907 #SUP 0.57 #QF 0.57
94 ALA HA 97 ILE H 4.69 #peak 1918 #SUP 0.94 #QF 0.94
57 ALA H 57 ALA QB 3.20 #peak 1929 #SUP 0.97 #QF 0.93
94 ALA H 94 ALA QB 3.16 #peak 1930 #SUP 0.99 #QF 0.99
57 ALA QB 98 ALA H 4.02 #peak 1931 #SUP 0.43 #QF 0.43
57 ALA QB 58 LEU H 3.25 #peak 1932 #SUP 0.87 #QF 0.87
57 ALA QB 97 ILE H 5.50 #peak 1933 #SUP 0.99 #QF 0.59
94 ALA QB 97 ILE H 5.50 #peak 1933 #SUP 0.99 #QF 0.59
162 VAL QG2 166 GLU HB3 4.71 #peak 1937 #SUP 1.00
90 LEU HA 93 ARG HD2 3.98 #peak 1939 #SUP 0.71 #QF 0.71
93 ARG HD2 94 ALA H 5.50 #peak 1943 #SUP 0.99 #QF 0.64
64 ILE QD1 93 ARG HD2 4.06 #peak 1949 #SUP 0.47 #QF 0.47
98 ALA QB 99 MET HG3 4.65 #peak 1951 #SUP 0.91 #QF 0.91
95 LYS H 98 ALA QB 5.26 #peak 1960 #SUP 0.88 #QF 0.66
28 ALA QB 29 ALA H 5.28 #peak 1961 #SUP 0.93 #QF 0.40
99 MET HG2 104 ALA QB 4.63 #peak 1981 #SUP 0.98 #QF 0.98
99 MET HG3 104 ALA QB 4.63 #peak 1989 #SUP 0.99 #QF 0.99
54 VAL HB 98 ALA QB 3.56 #peak 1995 #SUP 0.44 #QF 0.44
54 VAL H 54 VAL QG1 3.92 #peak 2003 #SUP 0.50 #QF 0.50
91 SER HB3 111 ILE HB 4.64 #peak 2012 #SUP 0.93 #QF 0.93
91 SER HB3 111 ILE QG2 3.47 #peak 2015 #SUP 0.95 #QF 0.95
91 SER HB3 111 ILE H 5.50 #peak 2016 #SUP 0.93 #QF 0.93
101 VAL HA 102 ASP HB2 5.03 #peak 2022 #SUP 0.89 #QF 0.89
45 VAL HB 103 GLY HA2 5.18 #peak 2026 #SUP 0.96 #QF 0.96

49 THR QG2 103 GLY HA3 5.07 #peak 2029 #SUP 0.94 #QF 0.34
103 GLY HA3 104 ALA QB 5.50 #peak 2029 #SUP 0.94 #QF 0.34
99 MET H 104 ALA QB 5.38 #peak 2035 #SUP 1.00
139 SER HB3 142 VAL HB 4.21 #peak 2044 #SUP 0.91 #QF 0.62
61 ASP HA 62 GLU HG2 4.77 #peak 2046 #SUP 0.74 #QF 0.74
62 GLU HG2 63 GLN HA 4.62 #peak 2048 #SUP 0.82 #QF 0.82
62 GLU HG2 66 LYS HD3 3.84 #peak 2049 #SUP 0.98 #QF 0.98
82 VAL HA 109 ASN HA 4.61 #peak 2052 #SUP 0.61 #QF 0.61
63 GLN HG3 90 LEU QD1 3.80 #peak 2062 #SUP 0.99 #QF 0.99
63 GLN HG3 90 LEU QD2 3.16 #peak 2063 #SUP 0.93 #QF 0.93
115 GLN HA 115 GLN HG3 3.90 #peak 2077 #SUP 0.83 #QF 0.83
115 GLN HG3 117 ILE HA 4.64 #peak 2078 #SUP 0.23 #QF 0.23
95 LYS H 96 GLN HG2 4.41 #peak 2082 #SUP 0.75 #QF 0.75
96 GLN H 96 GLN HG2 3.41 #peak 2083 #SUP 0.95 #QF 0.95
93 ARG H 96 GLN HG2 4.90 #peak 2084 #SUP 0.71 #QF 0.71
96 GLN HG2 97 ILE H 4.04 #peak 2086 #SUP 0.96 #QF 0.96
115 GLN HG3 117 ILE QG2 4.25 #peak 2090 #SUP 0.37 #QF 0.37
115 GLN HA 115 GLN HG2 3.81 #peak 2095 #SUP 0.84 #QF 0.84
117 ILE QG2 153 ASN HA 4.15 #peak 2106 #SUP 0.93 #QF 0.93
117 ILE QG2 121 GLU HG2 3.96 #peak 2108 #SUP 0.43 #QF 0.43
173 ILE HA 173 ILE QD1 3.09 #peak 2113 #SUP 0.97 #QF 0.97
124 ASN HA 128 ILE QD1 4.18 #peak 2114 #SUP 0.81 #QF 0.81
173 ILE QD1 177 VAL H 5.50 #peak 2116 #SUP 0.69 #QF 0.69
173 ILE H 173 ILE QD1 3.10 #peak 2118 #SUP 0.69 #QF 0.69
156 VAL HA 156 VAL QG2 3.55 #peak 2137 #SUP 0.99 #QF 0.99
123 SER HB3 124 ASN HA 4.91 #peak 2138 #SUP 0.91 #QF 0.91
56 SER HB3 57 ALA H 3.93 #peak 2139 #SUP 0.91 #QF 0.91
124 ASN HA 127 TRP HB2 3.95 #peak 2142 #SUP 0.98 #QF 0.98
124 ASN HA 127 TRP HB3 3.62 #peak 2143 #SUP 0.94 #QF 0.94
124 ASN HA 128 ILE H 4.37 #peak 2147 #SUP 0.96 #QF 0.96
146 ASN HB2 147 VAL HB 5.50 #peak 2152 #SUP 0.82 #QF 0.82
146 ASN HB2 162 VAL QG1 4.29 #peak 2153 #SUP 0.92 #QF 0.92
127 TRP HA 128 ILE HA 4.91 #peak 2154 #SUP 0.74 #QF 0.74
127 TRP HH2 128 ILE HA 5.50 #peak 2155 #SUP 0.83 #QF 0.83
128 ILE HA 131 LYS H 4.79 #peak 2157 #SUP 0.94 #QF 0.94
128 ILE HA 131 LYS HB2 4.54 #peak 2161 #SUP 0.94 #QF 0.94
128 ILE HA 131 LYS HB3 4.54 #peak 2162 #SUP 0.98 #QF 0.97
128 ILE HA 132 VAL QG1 3.85 #peak 2164 #SUP 0.50 #QF 0.50
128 ILE HA 128 ILE QD1 4.03 #peak 2165 #SUP 0.98 #QF 0.98
128 ILE HA 128 ILE QG2 3.55 #peak 2171 #SUP 0.94 #QF 0.94
128 ILE QG2 129 THR HA 4.26 #peak 2172 #SUP 0.96 #QF 0.96
128 ILE QG2 129 THR H 3.59 #peak 2178 #SUP 0.90 #QF 0.90
128 ILE QG2 132 VAL H 4.35 #peak 2179 #SUP 0.62 #QF 0.62
131 LYS HA 131 LYS HG3 3.34 #peak 2182 #SUP 0.94 #QF 0.94
46 ASP HA 49 THR HB 5.50 #peak 2202 #SUP 0.98 #QF 0.98
131 LYS H 132 VAL QG1 4.18 #peak 2211 #SUP 0.60 #QF 0.60
132 VAL H 132 VAL QG1 3.30 #peak 2212 #SUP 0.92 #QF 0.92
132 VAL HA 132 VAL QG2 3.39 #peak 2213 #SUP 0.72 #QF 0.72
173 ILE HA 176 ARG HD2 3.37 #peak 2228 #SUP 0.46 #QF 0.46
82 VAL HA 111 ILE HA 4.31 #peak 2233 #SUP 0.93 #QF 0.93
138 THR QG2 139 SER H 4.35 #peak 2271 #SUP 0.86 #QF 0.86
141 LEU QD1 165 ARG HD2 3.88 #peak 2290 #SUP 0.85 #QF 0.58
141 LEU QD1 165 ARG HD3 3.88 #peak 2290 #SUP 0.85 #QF 0.58
127 TRP HA 130 THR QG2 3.75 #peak 2305 #SUP 0.34 #QF 0.34
127 TRP HA 127 TRP HD1 4.95 #peak 2308 #SUP 0.97 #QF 0.97
91 SER HB3 111 ILE QD1 3.27 #peak 2318 #SUP 0.96 #QF 0.24
146 ASN HA 162 VAL QG1 4.83 #peak 2319 #SUP 0.91 #QF 0.91
55 ASN H 72 VAL HB 4.65 #peak 2326 #SUP 0.44 #QF 0.44
147 VAL QG1 148 LYS H 4.31 #peak 2335 #SUP 0.96 #QF 0.96
147 VAL QG1 162 VAL QG2 4.33 #peak 2337 #SUP 0.95 #QF 0.95
149 VAL HA 150 THR HA 4.43 #peak 2340 #SUP 0.84 #QF 0.84
149 VAL HA 150 THR HB 4.49 #peak 2342 #SUP 0.92 #QF 0.92
150 THR H 150 THR HB 3.37 #peak 2345 #SUP 0.96 #QF 0.96
150 THR QG2 151 THR HA 4.15 #peak 2352 #SUP 0.91 #QF 0.91
149 VAL HA 150 THR QG2 4.22 #peak 2353 #SUP 0.90 #QF 0.90
150 THR QG2 152 GLU H 4.47 #peak 2357 #SUP 0.79 #QF 0.79
150 THR QG2 157 PHE H 3.45 #peak 2358 #SUP 0.86 #QF 0.86
82 VAL HB 150 THR QG2 4.56 #peak 2363 #SUP 0.94 #QF 0.94
152 GLU H 156 VAL HB 5.50 #peak 2373 #SUP 0.72 #QF 0.55
156 VAL HB 184 THR H 5.50 #peak 2373 #SUP 0.72 #QF 0.55
151 THR HA 156 VAL QG1 4.54 #peak 2381 #SUP 0.99 #QF 0.99

131 LYS H 131 LYS HG3 4.68 #peak 2382 #SUP 0.95 #QF 0.95
127 TRP HH2 131 LYS HG3 4.32 #peak 2383 #SUP 0.80 #QF 0.80
131 LYS HG3 132 VAL H 4.05 #peak 2385 #SUP 0.80 #QF 0.80
156 VAL QG1 157 PHE H 4.26 #peak 2386 #SUP 0.99 #QF 0.99
160 GLY HA3 161 LEU HA 5.20 #peak 2398 #SUP 0.99 #QF 0.60
161 LEU HA 162 VAL HA 5.50 #peak 2398 #SUP 0.99 #QF 0.60
161 LEU HA 188 THR QG2 4.09 #peak 2401 #SUP 0.62 #QF 0.62
75 TYR HE1 161 LEU HA 4.86 #peak 2402 #SUP 0.90 #QF 0.90
75 TYR HE1 161 LEU HG 4.88 #peak 2415 #SUP 0.95 #QF 0.95
75 TYR HE1 161 LEU QD2 4.64 #peak 2417 #SUP 0.83 #QF 0.83
162 VAL QG2 164 GLU HA 5.46 #peak 2420 #SUP 0.95 #QF 0.95
163 THR HB 164 GLU HG2 5.50 #peak 2431 #SUP 0.99 #QF 0.99
167 ALA HA 189 PHE HD1 5.12 #peak 2453 #SUP 0.93 #QF 0.93
167 ALA HA 170 ALA H 4.64 #peak 2455 #SUP 0.90 #QF 0.90
162 VAL QG1 167 ALA HA 4.67 #peak 2457 #SUP 0.95 #QF 0.95
162 VAL QG2 167 ALA HA 3.92 #peak 2458 #SUP 0.93 #QF 0.93
162 VAL QG1 167 ALA QB 4.66 #peak 2460 #SUP 0.98 #QF 0.98
167 ALA QB 168 LYS HB3 4.03 #peak 2461 #SUP 0.26 #QF 0.26
162 VAL HB 167 ALA QB 3.81 #peak 2462 #SUP 0.91 #QF 0.91
164 GLU HA 167 ALA QB 3.15 #peak 2465 #SUP 0.91 #QF 0.91
162 VAL HA 167 ALA QB 5.16 #peak 2468 #SUP 0.93 #QF 0.93
169 ALA HA 172 ASP HB2 3.34 #peak 2474 #SUP 0.87 #QF 0.87
169 ALA HA 172 ASP HB3 3.75 #peak 2475 #SUP 0.96 #QF 0.96
169 ALA HA 173 ILE H 4.26 #peak 2477 #SUP 0.96 #QF 0.96
174 ALA QB 175 SER H 3.91 #peak 2483 #SUP 0.97 #QF 0.97
167 ALA HA 170 ALA QB 3.69 #peak 2487 #SUP 0.82 #QF 0.82
91 SER H 92 ALA QB 5.32 #peak 2489 #SUP 0.67 #QF 0.36
174 ALA QB 177 VAL H 4.89 #peak 2493 #SUP 0.65 #QF 0.65
57 ALA H 98 ALA QB 4.59 #peak 2494 #SUP 0.33 #QF 0.33
176 ARG HA 176 ARG HD2 4.63 #peak 2495 #SUP 0.92 #QF 0.92
176 ARG HB3 177 VAL H 4.43 #peak 2505 #SUP 0.98 #QF 0.46
183 VAL HA 183 VAL QG2 3.55 #peak 2519 #SUP 1.00
189 PHE HA 190 ILE HA 4.40 #peak 2527 #SUP 0.97 #QF 0.97
163 THR HA 190 ILE HA 4.65 #peak 2528 #SUP 1.00
190 ILE HA 191 LYS HA 4.48 #peak 2529 #SUP 0.91 #QF 0.91
162 VAL QG2 190 ILE QG2 5.50 #peak 2531 #SUP 0.69 #QF 0.69
189 PHE HA 190 ILE QG2 5.48 #peak 2532 #SUP 0.94 #QF 0.94
161 LEU HA 190 ILE HG12 5.00 #peak 2535 #SUP 0.96 #QF 0.96
64 ILE QD1 98 ALA H 5.50 #peak 2539 #SUP 0.62 #QF 0.62
191 LYS HB2 192 GLY H 5.50 #peak 2545 #SUP 0.79 #QF 0.55
191 LYS HB3 192 GLY H 5.50 #peak 2545 #SUP 0.79 #QF 0.55
163 THR HB 192 GLY HA3 4.19 #peak 2546 #SUP 0.67 #QF 0.67
163 THR HA 192 GLY HA3 5.50 #peak 2547 #SUP 0.86 #QF 0.81
195 GLU HA 195 GLU HG2 3.84 #peak 2568 #SUP 0.95 #QF 0.95
195 GLU HG2 196 HIS HA 5.01 #peak 2570 #SUP 0.88 #QF 0.88
195 GLU HG3 196 HIS HA 5.01 #peak 2571 #SUP 0.95 #QF 0.95
52 VAL HA 55 ASN H 4.16 #peak 2600 #SUP 0.98 #QF 0.98
53 ARG HB2 53 ARG HD2 3.88 #peak 2602 #SUP 0.89 #QF 0.89
53 ARG HB3 53 ARG HD2 3.88 #peak 2603 #SUP 0.91 #QF 0.91
66 LYS HA 66 LYS HG3 3.73 #peak 2607 #SUP 0.66 #QF 0.66
20 ILE HA 20 ILE HG12 3.71 #peak 2611 #SUP 0.75 #QF 0.75
20 ILE HA 21 ALA QB 3.86 #peak 2612 #SUP 0.90 #QF 0.90
51 GLU HA 51 GLU HG3 3.98 #peak 2735 #SUP 0.96 #QF 0.96
65 LYS HA 65 LYS HE2 5.50 #peak 2769 #SUP 0.88 #QF 0.28
65 LYS HA 65 LYS HE3 5.50 #peak 2769 #SUP 0.88 #QF 0.28
70 ILE HA 70 ILE HG13 4.19 #peak 2790 #SUP 0.97 #QF 0.97
70 ILE HA 70 ILE HG12 4.21 #peak 2793 #SUP 0.77 #QF 0.77
72 VAL HA 72 VAL QG1 3.34 #peak 2802 #SUP 0.97 #QF 0.97
72 VAL HA 72 VAL QG2 3.34 #peak 2803 #SUP 0.95 #QF 0.95
92 ALA QB 96 GLN HG2 3.97 #peak 2848 #SUP 0.60 #QF 0.60
92 ALA QB 95 LYS H 4.77 #peak 2850 #SUP 0.76 #QF 0.76
92 ALA QB 93 ARG H 3.87 #peak 2851 #SUP 0.98 #QF 0.96
171 ALA QB 172 ASP H 4.33 #peak 2851 #SUP 0.98 #QF 0.96
92 ALA QB 96 GLN H 5.23 #peak 2852 #SUP 0.95 #QF 0.87
122 ALA QB 124 ASN H 5.50 #peak 2852 #SUP 0.95 #QF 0.87
117 ILE HB 117 ILE QD1 4.10 #peak 2899 #SUP 1.00 #QF 0.99
141 LEU HA 141 LEU QD2 3.30 #peak 2979 #SUP 0.80 #QF 0.80
146 ASN HB3 147 VAL QG2 5.50 #peak 2987 #SUP 0.89 #QF 0.70
161 LEU HA 188 THR HA 5.07 #peak 3009 #SUP 0.99 #QF 0.99
161 LEU HA 161 LEU HG 3.59 #peak 3011 #SUP 0.99 #QF 0.91
183 VAL HA 183 VAL QG1 3.55 #peak 3067 #SUP 0.99 #QF 0.99

156 VAL H 183 VAL QG1 5.38 #peak 3068 #SUP 0.78 #QF 0.78
183 VAL QG1 184 THR H 4.27 #peak 3069 #SUP 0.77 #QF 0.77
185 THR HA 185 THR QG2 3.29 #peak 3081 #SUP 0.99 #QF 0.99
190 ILE HA 190 ILE HG13 3.38 #peak 3108 #SUP 0.93 #QF 0.93
79 VAL HA 79 VAL QG1 3.62 #peak 3202 #SUP 1.00 #QF 0.97
79 VAL HA 79 VAL QG2 3.62 #peak 3202 #SUP 1.00 #QF 0.97
90 LEU HB2 90 LEU QD1 3.14 #peak 3210 #SUP 0.98 #QF 0.98
90 LEU HB2 90 LEU QD2 3.84 #peak 3211 #SUP 1.00
90 LEU HB3 90 LEU QD1 3.12 #peak 3216 #SUP 0.98 #QF 0.98
90 LEU HB3 90 LEU QD2 3.70 #peak 3217 #SUP 1.00
52 VAL QG1 55 ASN HB3 4.45 #peak 3247 #SUP 0.97 #QF 0.97
70 ILE QG2 91 SER HA 3.98 #peak 3250 #SUP 0.65 #QF 0.65
20 ILE QG2 21 ALA HA 4.14 #peak 3258 #SUP 0.55 #QF 0.55
25 VAL H 25 VAL QG1 5.26 #peak 3265 #SUP 0.77 #QF 0.38
25 VAL H 25 VAL QG2 5.26 #peak 3265 #SUP 0.77 #QF 0.38
131 LYS HB2 131 LYS HE2 4.82 #peak 3274 #SUP 0.98 #QF 0.98
40 SER HA 41 VAL HA 4.63 #peak 3278 #SUP 0.49 #QF 0.49
45 VAL HB 103 GLY HA3 3.97 #peak 3283 #SUP 0.95 #QF 0.95
52 VAL QG2 53 ARG H 3.93 #peak 3285 #SUP 1.00
51 GLU HB3 52 VAL QG2 4.28 #peak 3287 #SUP 0.88 #QF 0.88
78 LYS HE2 79 VAL H 4.63 #peak 3290 #SUP 0.99 #QF 0.99
70 ILE QG2 95 LYS H 5.17 #peak 3294 #SUP 0.51 #QF 0.51
78 LYS HE2 108 TYR HE1 4.03 #peak 3312 #SUP 0.80 #QF 0.80
129 THR QG2 150 THR HA 3.85 #peak 3328 #SUP 0.54 #QF 0.54
74 ALA HA 75 TYR HA 4.69 #peak 3340 #SUP 0.93 #QF 0.93
50 LEU HA 52 VAL H 4.86 #peak 3343 #SUP 0.62 #QF 0.62
74 ALA QB 75 TYR HA 4.31 #peak 3360 #SUP 0.86 #QF 0.86
140 ASP HB2 141 LEU QD2 5.50 #peak 3363 #SUP 0.89 #QF 0.30
140 ASP HB3 141 LEU QD2 5.50 #peak 3363 #SUP 0.89 #QF 0.30
141 LEU QD2 165 ARG HD2 4.30 #peak 3364 #SUP 0.83 #QF 0.73
141 LEU QD2 165 ARG HD3 4.30 #peak 3364 #SUP 0.83 #QF 0.73
90 LEU HA 93 ARG H 4.40 #peak 3372 #SUP 0.77 #QF 0.62
90 LEU HA 94 ALA H 4.80 #peak 3372 #SUP 0.77 #QF 0.62
122 ALA HA 125 ASP HB2 4.16 #peak 3399 #SUP 0.90 #QF 0.90
161 LEU HA 162 VAL QG2 4.66 #peak 3406 #SUP 0.96 #QF 0.96
161 LEU HA 162 VAL QG1 4.86 #peak 3407 #SUP 0.96 #QF 0.96
159 MET HA 186 ALA QB 4.18 #peak 3412 #SUP 0.94 #QF 0.94
83 GLY H 111 ILE QG2 4.62 #peak 3427 #SUP 0.92 #QF 0.92
84 GLN H 111 ILE QG2 4.98 #peak 3428 #SUP 0.87 #QF 0.87
61 ASP HB2 64 ILE QG2 3.55 #peak 3434 #SUP 0.84 #QF 0.84
61 ASP HB3 64 ILE QG2 4.06 #peak 3436 #SUP 1.00
64 ILE HA 64 ILE QG2 3.13 #peak 3437 #SUP 0.98 #QF 0.98
58 LEU HA 64 ILE QG2 3.47 #peak 3439 #SUP 0.49 #QF 0.49
64 ILE QG2 94 ALA HA 4.58 #peak 3440 #SUP 0.75 #QF 0.75
64 ILE H 64 ILE QG2 4.26 #peak 3441 #SUP 0.80 #QF 0.80
64 ILE QG2 94 ALA H 5.24 #peak 3442 #SUP 0.92 #QF 0.92
173 ILE QG2 174 ALA HA 3.82 #peak 3448 #SUP 0.88 #QF 0.88
173 ILE H 173 ILE QG2 4.07 #peak 3450 #SUP 0.95 #QF 0.75
173 ILE QG2 174 ALA H 4.30 #peak 3450 #SUP 0.95 #QF 0.75
135 GLN H 173 ILE QG2 5.50 #peak 3451 #SUP 0.76 #QF 0.41
173 ILE QG2 175 SER H 5.50 #peak 3451 #SUP 0.76 #QF 0.41
75 TYR HE1 160 GLY HA2 5.10 #peak 3467 #SUP 0.74 #QF 0.57
75 TYR HD1 160 GLY HA2 5.09 #peak 3468 #SUP 0.85 #QF 0.74
93 ARG H 93 ARG HD3 4.99 #peak 3472 #SUP 0.99 #QF 0.95
93 ARG HD3 94 ALA H 5.50 #peak 3472 #SUP 0.99 #QF 0.95
73 THR HB 80 LEU HB2 5.31 #peak 3478 #SUP 0.98 #QF 0.98
80 LEU HB2 82 VAL QG1 5.32 #peak 3481 #SUP 0.99 #QF 0.99
80 LEU HB3 82 VAL QG1 5.02 #peak 3482 #SUP 0.99 #QF 0.99
53 ARG H 53 ARG HD3 4.88 #peak 3484 #SUP 0.94 #QF 0.94
53 ARG HD3 54 VAL H 5.09 #peak 3485 #SUP 0.80 #QF 0.80
131 LYS HB2 131 LYS HE3 4.82 #peak 3491 #SUP 0.93 #QF 0.93
132 VAL HA 132 VAL QG1 3.21 #peak 3507 #SUP 0.83 #QF 0.83
131 LYS HG2 132 VAL HA 3.59 #peak 3508 #SUP 1.00 #QF 0.51
63 GLN HG2 90 LEU QD1 4.24 #peak 3511 #SUP 0.99 #QF 0.99
63 GLN HG2 90 LEU QD2 3.51 #peak 3512 #SUP 0.99 #QF 0.99
64 ILE HA 90 LEU HB3 4.26 #peak 3520 #SUP 0.66 #QF 0.66
82 VAL QG1 111 ILE HA 5.50 #peak 3528 #SUP 0.97 #QF 0.97
78 LYS HA 78 LYS HD2 5.16 #peak 3529 #SUP 0.89 #QF 0.89
95 LYS HA 95 LYS HD3 4.96 #peak 3531 #SUP 0.99 #QF 0.99
78 LYS HD3 108 TYR HD1 4.94 #peak 3534 #SUP 0.32 #QF 0.32
151 THR HA 155 GLU H 4.93 #peak 3537 #SUP 0.79 #QF 0.79

131 LYS HD2 132 VAL H 5.16 #peak 3538 #SUP 0.89 #QF 0.89
131 LYS HD3 132 VAL H 5.34 #peak 3539 #SUP 0.94 #QF 0.94
64 ILE H 66 LYS HE2 5.50 #peak 3551 #SUP 0.52 #QF 0.31
64 ILE H 66 LYS HE3 5.50 #peak 3551 #SUP 0.52 #QF 0.31
66 LYS HE3 67 GLU HG2 3.33 #peak 3555 #SUP 0.82 #QF 0.82
62 GLU HA 65 LYS HE2 4.03 #peak 3557 #SUP 0.83 #QF 0.68
62 GLU HA 65 LYS HE3 4.03 #peak 3557 #SUP 0.83 #QF 0.68
148 LYS HB2 148 LYS HD3 4.22 #peak 3648 #SUP 0.84 #QF 0.84
148 LYS HB2 148 LYS HD2 4.22 #peak 3649 #SUP 0.93 #QF 0.93
148 LYS HB3 148 LYS HD3 4.22 #peak 3657 #SUP 0.99 #QF 0.99
161 LEU HA 190 ILE HG13 5.00 #peak 3676 #SUP 0.90 #QF 0.90
172 ASP HB2 173 ILE HA 5.14 #peak 3685 #SUP 0.98 #QF 0.98
172 ASP HB3 173 ILE HA 4.62 #peak 3687 #SUP 0.98 #QF 0.98
176 ARG HA 176 ARG HD3 4.06 #peak 3693 #SUP 0.93 #QF 0.93
176 ARG HD2 177 VAL H 4.56 #peak 3694 #SUP 0.97 #QF 0.97
173 ILE H 176 ARG HD2 5.50 #peak 3695 #SUP 0.92 #QF 0.92
176 ARG HD3 177 VAL H 4.53 #peak 3698 #SUP 0.95 #QF 0.95
173 ILE HA 176 ARG HD3 4.03 #peak 3699 #SUP 0.97 #QF 0.97
173 ILE QG2 176 ARG HD3 4.25 #peak 3706 #SUP 0.91 #QF 0.91
176 ARG HB2 176 ARG HD2 4.15 #peak 3711 #SUP 0.91 #QF 0.91
176 ARG HB3 176 ARG HD2 4.15 #peak 3714 #SUP 0.99 #QF 0.99
132 VAL HA 135 GLN HG3 4.31 #peak 3747 #SUP 0.93 #QF 0.93
131 LYS HD2 132 VAL HA 5.50 #peak 3750 #SUP 0.89 #QF 0.57
131 LYS HA 132 VAL HA 4.86 #peak 3751 #SUP 0.92 #QF 0.92
51 GLU HB2 52 VAL QG2 4.28 #peak 3757 #SUP 0.92 #QF 0.92
53 ARG HA 53 ARG HD2 4.06 #peak 3760 #SUP 1.00
53 ARG HA 53 ARG HD3 4.06 #peak 3761 #SUP 0.83 #QF 0.83
52 VAL QG1 53 ARG HA 3.57 #peak 3763 #SUP 0.88 #QF 0.88
50 LEU HA 50 LEU HG 3.72 #peak 3772 #SUP 0.93 #QF 0.50
137 LEU HA 137 LEU HG 3.61 #peak 3783 #SUP 0.84 #QF 0.84
63 GLN HA 63 GLN HG2 3.68 #peak 3785 #SUP 0.86 #QF 0.86
117 ILE QD1 121 GLU HB2 4.41 #peak 3790 #SUP 0.92 #QF 0.27
48 GLY HA2 51 GLU HB2 4.70 #peak 3804 #SUP 0.84 #QF 0.84
54 VAL HA 98 ALA QB 3.61 #peak 3817 #SUP 0.75 #QF 0.50
61 ASP HA 64 ILE QG2 5.50 #peak 3822 #SUP 1.00
59 SER H 64 ILE QG2 5.50 #peak 3823 #SUP 0.92 #QF 0.29
64 ILE QG2 68 ALA H 5.50 #peak 3823 #SUP 0.92 #QF 0.29
79 VAL H 106 GLU HB3 5.00 #peak 3839 #SUP 0.78 #QF 0.78
79 VAL H 106 GLU HB2 5.00 #peak 3840 #SUP 0.54 #QF 0.54
78 LYS HA 106 GLU HG3 5.09 #peak 3848 #SUP 1.00 #QF 0.64
36 THR HB 37 ASP HA 4.80 #peak 3861 #SUP 0.71 #QF 0.71
52 VAL HB 53 ARG HA 4.49 #peak 3866 #SUP 0.95 #QF 0.95
54 VAL HA 55 ASN HA 5.01 #peak 3873 #SUP 0.63 #QF 0.63
55 ASN HA 72 VAL HB 3.65 #peak 3874 #SUP 0.20 #QF 0.20
55 ASN HA 58 LEU H 4.19 #peak 3877 #SUP 0.93 #QF 0.93
63 GLN HG2 66 LYS HE2 5.50 #peak 3881 #SUP 0.80 #QF 0.63
63 GLN HG2 66 LYS HE3 5.50 #peak 3881 #SUP 0.80 #QF 0.63
63 GLN HG2 64 ILE HA 4.81 #peak 3882 #SUP 0.89 #QF 0.89
63 GLN HG2 90 LEU HA 4.89 #peak 3883 #SUP 0.76 #QF 0.76
64 ILE HA 90 LEU HB2 3.90 #peak 3891 #SUP 0.96 #QF 0.96
67 GLU HA 67 GLU HG3 4.01 #peak 3895 #SUP 0.99 #QF 0.99
66 LYS HG3 67 GLU HG3 3.95 #peak 3898 #SUP 0.99 #QF 0.99
89 GLU HG3 90 LEU HA 4.08 #peak 3920 #SUP 0.50 #QF 0.50
70 ILE QG2 71 ASN H 3.84 #peak 3935 #SUP 0.99 #QF 0.99
77 GLY HA2 105 ASN HB2 5.50 #peak 3942 #SUP 0.93 #QF 0.65
169 ALA HA 173 ILE QD1 3.68 #peak 3944 #SUP 0.48 #QF 0.48
129 THR HA 132 VAL H 4.50 #peak 3948 #SUP 0.92 #QF 0.92
129 THR HA 132 VAL QG2 3.83 #peak 3949 #SUP 0.75 #QF 0.68
127 TRP HH2 131 LYS HD2 4.75 #peak 3967 #SUP 0.89 #QF 0.89
127 TRP HH2 131 LYS HD3 4.55 #peak 3968 #SUP 0.96 #QF 0.96
131 LYS HB3 131 LYS HE2 4.82 #peak 3971 #SUP 0.88 #QF 0.88
131 LYS HB3 131 LYS HE3 4.82 #peak 3974 #SUP 0.27 #QF 0.27
131 LYS HG3 135 GLN HG3 4.34 #peak 3981 #SUP 0.97 #QF 0.94
135 GLN H 135 GLN HG2 4.26 #peak 3996 #SUP 0.98 #QF 0.98
141 LEU HA 141 LEU HG 3.25 #peak 4003 #SUP 0.84 #QF 0.84
124 ASN HB3 128 ILE QD1 3.93 #peak 4006 #SUP 0.82 #QF 0.82
147 VAL HB 162 VAL QG1 4.57 #peak 4016 #SUP 0.97 #QF 0.97
148 LYS HA 148 LYS HE2 4.68 #peak 4017 #SUP 0.98 #QF 0.93
148 LYS HA 148 LYS HE3 4.68 #peak 4017 #SUP 0.98 #QF 0.93
78 LYS HE3 79 VAL H 4.63 #peak 4060 #SUP 0.95 #QF 0.95
75 TYR H 78 LYS HD3 5.39 #peak 4085 #SUP 0.89 #QF 0.29

78 LYS HD3 79 VAL H 5.50 #peak 4085 #SUP 0.89 #QF 0.29
80 LEU HG 82 VAL QG1 4.02 #peak 4090 #SUP 0.96 #QF 0.96
80 LEU H 80 LEU HG 5.10 #peak 4092 #SUP 0.96 #QF 0.96
70 ILE HA 83 GLY HA2 4.91 #peak 4110 #SUP 0.87 #QF 0.87
128 ILE HB 129 THR HA 4.52 #peak 4112 #SUP 0.96 #QF 0.96
72 VAL HA 81 LEU HG 4.89 #peak 4125 #SUP 0.94 #QF 0.42
81 LEU HG 109 ASN HA 5.39 #peak 4125 #SUP 0.94 #QF 0.42
156 VAL HA 156 VAL QG1 3.55 #peak 4137 #SUP 0.93 #QF 0.93
93 ARG HA 93 ARG HG2 3.78 #peak 4147 #SUP 0.96 #QF 0.96
64 ILE QD1 93 ARG HD3 4.06 #peak 4169 #SUP 0.81 #QF 0.81
95 LYS HA 95 LYS HD2 4.96 #peak 4176 #SUP 0.94 #QF 0.94
131 LYS HA 131 LYS HG2 3.74 #peak 4192 #SUP 0.95 #QF 0.95
134 SER HB3 135 GLN H 4.69 #peak 4207 #SUP 0.82 #QF 0.56
134 SER HB3 135 GLN HG2 4.48 #peak 4208 #SUP 0.83 #QF 0.83
135 GLN HA 135 GLN HG2 3.12 #peak 4209 #SUP 0.84 #QF 0.84
142 VAL HA 166 GLU HG2 3.90 #peak 4212 #SUP 0.95 #QF 0.95
141 LEU HB2 141 LEU QD1 3.34 #peak 4213 #SUP 1.00
54 VAL H 54 VAL QG2 3.92 #peak 4217 #SUP 0.51 #QF 0.51
148 LYS HA 148 LYS HD2 4.09 #peak 4223 #SUP 0.96 #QF 0.96
148 LYS HA 148 LYS HD3 4.09 #peak 4224 #SUP 0.99 #QF 0.99
149 VAL HA 149 VAL QG2 3.24 #peak 4238 #SUP 0.97 #QF 0.97
156 VAL H 156 VAL QG2 3.96 #peak 4242 #SUP 0.98 #QF 0.98
166 GLU HA 166 GLU HG2 3.70 #peak 4255 #SUP 0.90 #QF 0.90
166 GLU HA 169 ALA QB 3.39 #peak 4257 #SUP 0.65 #QF 0.65
166 GLU HB2 167 ALA QB 5.50 #peak 4265 #SUP 0.95 #QF 0.91
181 LYS HA 181 LYS HG2 4.06 #peak 4275 #SUP 0.93 #QF 0.93
181 LYS HA 181 LYS HG3 4.06 #peak 4280 #SUP 0.94 #QF 0.94
53 ARG H 53 ARG HD2 4.88 #peak 4285 #SUP 0.98 #QF 0.98
170 ALA HA 173 ILE HB 4.32 #peak 4288 #SUP 0.52 #QF 0.32
173 ILE HB 174 ALA HA 4.91 #peak 4288 #SUP 0.52 #QF 0.32
63 GLN HB2 90 LEU QD2 3.17 #peak 4289 #SUP 0.93 #QF 0.93
151 THR HA 156 VAL QG2 4.54 #peak 4298 #SUP 0.98 #QF 0.98
82 VAL QG1 150 THR HB 5.50 #peak 4325 #SUP 0.96 #QF 0.96
85 SER H 111 ILE HB 5.26 #peak 4338 #SUP 0.98 #QF 0.89
111 ILE HB 112 ARG H 5.50 #peak 4338 #SUP 0.98 #QF 0.89
81 LEU H 109 ASN HA 5.50 #peak 4356 #SUP 0.99 #QF 0.98
49 THR HA 49 THR QG2 3.21 #peak 4383 #SUP 0.69 #QF 0.69
27 THR H 27 THR QG2 4.81 #peak 4385 #SUP 0.94 #QF 0.94
46 ASP HB2 49 THR QG2 4.38 #peak 4389 #SUP 1.00
47 ASP HB2 48 GLY HA3 5.31 #peak 4393 #SUP 0.98 #QF 0.80
48 GLY HA2 51 GLU HB3 4.70 #peak 4397 #SUP 0.79 #QF 0.79
48 GLY HA2 52 VAL QG2 5.10 #peak 4399 #SUP 0.96 #QF 0.77
49 THR HB 52 VAL HB 5.43 #peak 4404 #SUP 0.98 #QF 0.98
49 THR QG2 52 VAL H 4.94 #peak 4407 #SUP 0.66 #QF 0.66
68 ALA QB 90 LEU HB3 4.48 #peak 4409 #SUP 0.97 #QF 0.97
51 GLU HA 52 VAL QG2 5.11 #peak 4410 #SUP 0.53 #QF 0.53
51 GLU HA 73 THR HA 5.50 #peak 4411 #SUP 0.94 #QF 0.94
49 THR HA 53 ARG HD2 4.54 #peak 4426 #SUP 0.79 #QF 0.79
53 ARG HD2 54 VAL H 5.09 #peak 4427 #SUP 1.00
132 VAL H 132 VAL QG2 3.78 #peak 4434 #SUP 0.97 #QF 0.97
58 LEU HA 58 LEU HG 3.94 #peak 4441 #SUP 0.96 #QF 0.96
62 GLU HG3 63 GLN HA 3.83 #peak 4451 #SUP 0.86 #QF 0.86
63 GLN HA 66 LYS HD3 3.58 #peak 4453 #SUP 0.93 #QF 0.89
63 GLN HG3 66 LYS HE2 5.50 #peak 4456 #SUP 0.77 #QF 0.49
63 GLN HG3 66 LYS HE3 5.50 #peak 4456 #SUP 0.77 #QF 0.49
64 ILE HA 68 ALA HA 5.12 #peak 4460 #SUP 0.78 #QF 0.78
64 ILE QG2 68 ALA HA 5.28 #peak 4463 #SUP 0.99 #QF 0.99
58 LEU HA 64 ILE QD1 3.53 #peak 4464 #SUP 0.49 #QF 0.49
191 LYS H 191 LYS HE2 5.50 #peak 4467 #SUP 0.82 #QF 0.32
191 LYS H 191 LYS HE3 5.50 #peak 4467 #SUP 0.82 #QF 0.32
67 GLU HG3 90 LEU QD1 3.76 #peak 4471 #SUP 0.76 #QF 0.76
67 GLU HG3 90 LEU QD2 4.41 #peak 4472 #SUP 0.78 #QF 0.78
70 ILE HA 70 ILE QD1 3.64 #peak 4475 #SUP 0.96 #QF 0.96
71 ASN HB2 82 VAL H 5.01 #peak 4479 #SUP 0.97 #QF 0.97
73 THR HB 79 VAL HA 5.50 #peak 4482 #SUP 0.82 #QF 0.82
73 THR HB 82 VAL QG1 5.38 #peak 4484 #SUP 0.97 #QF 0.97
74 ALA QB 75 TYR HB3 4.74 #peak 4488 #SUP 0.58 #QF 0.58
74 ALA QB 75 TYR HB2 5.50 #peak 4489 #SUP 0.76 #QF 0.76
162 VAL QG2 188 THR HA 5.50 #peak 4511 #SUP 0.98 #QF 0.98
188 THR HA 189 PHE HA 4.70 #peak 4512 #SUP 0.92 #QF 0.92
135 GLN HA 138 THR QG2 3.49 #peak 4537 #SUP 0.79 #QF 0.79

```
 82 VAL  HB    83 GLY  H      5.04        #peak  4541 #SUP  0.98  #QF  0.98
 81 LEU  H     82 VAL  QG2    5.25        #peak  4543 #SUP  0.86  #QF  0.86
 82 VAL  QG2   83 GLY  HA2    4.44        #peak  4544 #SUP  0.69  #QF  0.69
 86 PRO  HB2   87 ASN  HB2    5.49        #peak  4565 #SUP  0.88  #QF  0.22
 87 ASN  HB2   90 LEU  HB2    4.47        #peak  4567 #SUP  0.56  #QF  0.56
 86 PRO  HB2   87 ASN  HB3    5.50        #peak  4569 #SUP  0.95  #QF  0.84
 87 ASN  HB3   89 GLU  HB2    5.50        #peak  4569 #SUP  0.95  #QF  0.84
 87 ASN  HB3   88 ALA  QB     5.12        #peak  4572 #SUP  0.87  #QF  0.87
 88 ALA  QB    89 GLU  HB3    4.63        #peak  4579 #SUP  0.22  #QF  0.22
 91 SER  HB2  111 ILE  H      5.50        #peak  4592 #SUP  0.66  #QF  0.66
 91 SER  HB3   94 ALA  H      5.50        #peak  4594 #SUP  0.93  #QF  0.41
162 VAL  QG2  170 ALA  QB     4.59        #peak  4596 #SUP  0.93  #QF  0.93
162 VAL  QG1  170 ALA  QB     4.45        #peak  4597 #SUP  0.85  #QF  0.85
 96 GLN  HA    96 GLN  HG2    3.46        #peak  4609 #SUP  0.95  #QF  0.95
 93 ARG  HA    96 GLN  HB2    4.11        #peak  4613 #SUP  0.98  #QF  0.98
 93 ARG  HA    96 GLN  HG2    3.54        #peak  4614 #SUP  0.97  #QF  0.97
 93 ARG  HA    96 GLN  HG3    3.13        #peak  4615 #SUP  0.94  #QF  0.94
 94 ALA  HA    96 GLN  HG3    5.06        #peak  4616 #SUP  0.96  #QF  0.96
102 ASP  HB3  103 GLY  HA3    5.35        #peak  4641 #SUP  0.89  #QF  0.89
 99 MET  HA   102 ASP  HB2    5.30        #peak  4644 #SUP  0.85  #QF  0.85
 99 MET  HA   102 ASP  HB3    4.79        #peak  4645 #SUP  0.77  #QF  0.77
 91 SER  HB2  111 ILE  HB     4.64        #peak  4661 #SUP  1.00
 91 SER  HB2  111 ILE  QG2    3.47        #peak  4662 #SUP  0.99  #QF  0.99
 82 VAL  QG1  111 ILE  QG2    5.50        #peak  4666 #SUP  0.84  #QF  0.84
 82 VAL  HA   111 ILE  QD1    4.35        #peak  4670 #SUP  0.94  #QF  0.94
 70 ILE  HG13 111 ILE  QG2    3.57        #peak  4672 #SUP  0.68  #QF  0.68
113 GLN  HA   114 GLY  HA2    4.91        #peak  4674 #SUP  0.94  #QF  0.94
113 GLN  HA   114 GLY  HA3    4.91        #peak  4682 #SUP  0.94  #QF  0.94
117 ILE  QD1  121 GLU  HB3    4.41        #peak  4724 #SUP  0.96  #QF  0.48
123 SER  HB2  124 ASN  HA     4.91        #peak  4736 #SUP  0.82  #QF  0.82
122 ALA  HA   125 ASP  HB3    4.16        #peak  4739 #SUP  0.89  #QF  0.89
129 THR  QG2  131 LYS  H      4.98        #peak  4766 #SUP  0.96  #QF  0.96
129 THR  QG2  130 THR  HA     3.26        #peak  4772 #SUP  0.81  #QF  0.81
130 THR  QG2  132 VAL  H      4.34        #peak  4778 #SUP  0.28  #QF  0.28
129 THR  HA   132 VAL  QG1    3.79        #peak  4782 #SUP  0.79  #QF  0.79
131 LYS  HA   135 GLN  HG2    5.04        #peak  4792 #SUP  0.54  #QF  0.54
135 GLN  H    135 GLN  HG3    4.20        #peak  4797 #SUP  1.00
140 ASP  HA   141 LEU  HG     5.17        #peak  4834 #SUP  0.93  #QF  0.93
139 SER  HB2  142 VAL  HB     4.21        #peak  4838 #SUP  0.76  #QF  0.76
147 VAL  QG2  162 VAL  QG2    4.33        #peak  4840 #SUP  0.84  #QF  0.84
146 ASN  HA   147 VAL  HA     4.70        #peak  4853 #SUP  0.86  #QF  0.86
124 ASN  HB2  128 ILE  QD1    3.93        #peak  4856 #SUP  0.76  #QF  0.76
149 VAL  H    149 VAL  QG2    3.79        #peak  4875 #SUP  0.91  #QF  0.91
150 THR  HA   157 PHE  H      5.14        #peak  4877 #SUP  0.92  #QF  0.92
 45 VAL  H     45 VAL  QG1    5.07        #peak  4880 #SUP  1.00  #QF  0.67
 45 VAL  H     45 VAL  QG2    5.07        #peak  4880 #SUP  1.00  #QF  0.67
117 ILE  QG2  121 GLU  HG3    3.96        #peak  4887 #SUP  0.70  #QF  0.27
117 ILE  QD1  121 GLU  HG3    4.26        #peak  4887 #SUP  0.70  #QF  0.27
152 GLU  H    156 VAL  HA     4.69        #peak  4893 #SUP  0.65  #QF  0.65
131 LYS  HG3  132 VAL  HA     4.67        #peak  4897 #SUP  0.78  #QF  0.78
147 VAL  HB   159 MET  HA     4.75        #peak  4898 #SUP  0.21  #QF  0.21
 75 TYR  HD1  160 GLY  HA3    5.50        #peak  4908 #SUP  0.67  #QF  0.67
148 LYS  H    160 GLY  HA3    5.10        #peak  4909 #SUP  0.94  #QF  0.91
160 GLY  HA2  162 VAL  QG1    5.50        #peak  4910 #SUP  0.98  #QF  0.98
147 VAL  QG1  160 GLY  HA2    5.26        #peak  4911 #SUP  0.93  #QF  0.93
147 VAL  QG2  160 GLY  HA2    5.26        #peak  4912 #SUP  0.98  #QF  0.98
160 GLY  HA3  162 VAL  QG1    4.79        #peak  4913 #SUP  0.91  #QF  0.91
160 GLY  HA2  161 LEU  HG     5.50        #peak  4918 #SUP  0.42  #QF  0.42
147 VAL  HB   160 GLY  HA2    5.50        #peak  4919 #SUP  0.97  #QF  0.34
161 LEU  HG   190 ILE  HA     4.79        #peak  4924 #SUP  0.88  #QF  0.88
161 LEU  QD1  162 VAL  HA     5.14        #peak  4927 #SUP  0.99  #QF  0.99
147 VAL  HB   162 VAL  QG2    5.19        #peak  4931 #SUP  0.67  #QF  0.67
163 THR  HA   189 PHE  HA     5.00        #peak  4935 #SUP  0.81  #QF  0.81
142 VAL  QG1  166 GLU  HB2    5.50        #peak  4942 #SUP  0.91  #QF  0.55
142 VAL  QG2  166 GLU  HB2    5.50        #peak  4942 #SUP  0.91  #QF  0.55
142 VAL  QG1  166 GLU  HB3    5.50        #peak  4943 #SUP  0.91  #QF  0.39
142 VAL  QG2  166 GLU  HB3    5.50        #peak  4943 #SUP  0.91  #QF  0.39
166 GLU  HA   166 GLU  HG3    3.70        #peak  4944 #SUP  0.98  #QF  0.98
166 GLU  HG2  167 ALA  H      5.33        #peak  4956 #SUP  0.93  #QF  0.93
166 GLU  HG3  167 ALA  H      5.33        #peak  4957 #SUP  1.00
142 VAL  HA   166 GLU  HG3    3.90        #peak  4959 #SUP  0.99  #QF  0.99
```

162 VAL QG2 166 GLU HG2 5.50 #peak 4960 #SUP 0.96 #QF 0.96
162 VAL QG2 166 GLU HG3 5.50 #peak 4961 #SUP 1.00
162 VAL QG1 166 GLU HG2 4.79 #peak 4962 #SUP 0.98 #QF 0.98
162 VAL QG1 166 GLU HG3 4.79 #peak 4963 #SUP 1.00
164 GLU HA 167 ALA HA 5.50 #peak 4964 #SUP 0.93 #QF 0.85
167 ALA HA 170 ALA HA 5.50 #peak 4964 #SUP 0.93 #QF 0.85
47 ASP HB3 48 GLY HA2 5.09 #peak 4969 #SUP 0.56 #QF 0.56
173 ILE HA 173 ILE HG12 3.72 #peak 4970 #SUP 0.98 #QF 0.98
173 ILE HA 173 ILE HG13 3.99 #peak 4974 #SUP 0.97 #QF 0.97
173 ILE HA 174 ALA QB 5.17 #peak 4975 #SUP 0.94 #QF 0.94
136 LEU H 173 ILE QG2 5.50 #peak 4980 #SUP 0.46 #QF 0.46
173 ILE HB 173 ILE QD1 3.77 #peak 4990 #SUP 0.99 #QF 0.99
161 LEU HA 190 ILE HA 5.32 #peak 5020 #SUP 0.94 #QF 0.94
75 TYR HE2 78 LYS HE3 4.89 #peak 5073 #SUP 0.99 #QF 0.99
81 LEU H 81 LEU HG 5.08 #peak 5089 #SUP 0.86 #QF 0.86
81 LEU QD2 82 VAL H 5.09 #peak 5091 #SUP 0.94 #QF 0.94
130 THR HA 133 ARG HD2 3.67 #peak 5106 #SUP 0.47 #QF 0.21
130 THR HA 133 ARG HD3 3.67 #peak 5106 #SUP 0.47 #QF 0.21
159 MET HA 159 MET HG2 4.25 #peak 5145 #SUP 0.88 #QF 0.74
159 MET HA 159 MET HG3 4.25 #peak 5145 #SUP 0.88 #QF 0.74
93 ARG HA 93 ARG HG3 3.78 #peak 5148 #SUP 0.51 #QF 0.24
165 ARG HA 165 ARG HG2 4.03 #peak 5148 #SUP 0.51 #QF 0.24
165 ARG HA 165 ARG HG3 4.03 #peak 5149 #SUP 0.84 #QF 0.84
168 LYS HG2 170 ALA H 5.07 #peak 5171 #SUP 0.96 #QF 0.96
168 LYS HG3 170 ALA H 5.07 #peak 5172 #SUP 0.91 #QF 0.91
168 LYS HA 168 LYS HG2 4.03 #peak 5191 #SUP 0.99 #QF 0.99
168 LYS HA 168 LYS HG3 4.03 #peak 5192 #SUP 0.85 #QF 0.76
168 LYS HA 171 ALA QB 4.03 #peak 5192 #SUP 0.85 #QF 0.76
162 VAL HB 163 THR QG2 3.97 #peak 5195 #SUP 0.47 #QF 0.47
162 VAL QG1 163 THR QG2 4.84 #peak 5198 #SUP 0.98 #QF 0.98
163 THR HB 164 GLU HG3 5.50 #peak 5203 #SUP 0.90 #QF 0.90
142 VAL HA 166 GLU HA 4.64 #peak 5208 #SUP 0.96 #QF 0.96
162 VAL QG1 166 GLU HA 5.41 #peak 5209 #SUP 1.00
166 GLU HB3 167 ALA QB 5.50 #peak 5211 #SUP 0.90 #QF 0.90
169 ALA HA 173 ILE HG12 4.15 #peak 5214 #SUP 0.75 #QF 0.75
169 ALA QB 171 ALA H 4.76 #peak 5216 #SUP 0.82 #QF 0.82
87 ASN HB3 90 LEU H 5.12 #peak 5219 #SUP 0.94 #QF 0.52
171 ALA H 172 ASP HB2 5.50 #peak 5219 #SUP 0.94 #QF 0.52
47 ASP HB2 49 THR H 5.50 #peak 5220 #SUP 0.94 #QF 0.94
64 ILE HA 90 LEU HA 4.68 #peak 5268 #SUP 0.76 #QF 0.76
59 SER HA 65 LYS HD2 3.83 #peak 5273 #SUP 0.89 #QF 0.81
89 GLU HG2 90 LEU HG 4.26 #peak 5275 #SUP 0.78 #QF 0.43
89 GLU HG3 90 LEU HG 4.40 #peak 5275 #SUP 0.78 #QF 0.43
20 ILE HB 21 ALA H 3.98 #peak 5 #SUP 0.76 #QF 0.76
21 ALA H 22 ALA QB 3.96 #peak 7 #SUP 0.29 #QF 0.29
20 ILE QG2 21 ALA H 3.54 #peak 8 #SUP 0.69 #QF 0.69
20 ILE HG13 21 ALA H 4.93 #peak 9 #SUP 0.56 #QF 0.56
33 LYS HB2 34 ALA H 4.74 #peak 34 #SUP 0.97 #QF 0.91
69 ARG HB2 70 ILE H 5.36 #peak 34 #SUP 0.97 #QF 0.91
38 PRO HB2 39 ARG H 4.07 #peak 56 #SUP 0.84 #QF 0.84
39 ARG HB2 40 SER H 4.49 #peak 65 #SUP 0.75 #QF 0.75
48 GLY H 49 THR QG2 4.58 #peak 109 #SUP 0.75 #QF 0.75
135 GLN H 138 THR QG2 4.48 #peak 112 #SUP 0.68 #QF 0.68
51 GLU H 51 GLU HB2 3.90 #peak 118 #SUP 0.91 #QF 0.91
49 THR QG2 51 GLU H 4.76 #peak 120 #SUP 0.85 #QF 0.36
51 GLU H 52 VAL QG2 5.12 #peak 120 #SUP 0.85 #QF 0.36
52 VAL H 53 ARG H 3.65 #peak 122 #SUP 0.98 #QF 0.98
51 GLU HB3 52 VAL H 3.72 #peak 125 #SUP 0.75 #QF 0.75
54 VAL H 54 VAL HB 3.66 #peak 133 #SUP 0.90 #QF 0.90
55 ASN H 56 SER HB3 5.36 #peak 138 #SUP 0.82 #QF 0.82
55 ASN HB2 56 SER H 3.88 #peak 145 #SUP 0.85 #QF 0.85
55 ASN HB3 56 SER H 3.53 #peak 146 #SUP 0.78 #QF 0.78
56 SER H 57 ALA QB 4.46 #peak 147 #SUP 0.72 #QF 0.72
52 VAL QG1 56 SER H 4.18 #peak 148 #SUP 0.71 #QF 0.71
57 ALA H 58 LEU H 3.86 #peak 151 #SUP 0.95 #QF 0.95
55 ASN HA 59 SER H 4.24 #peak 159 #SUP 0.85 #QF 0.85
57 ALA QB 60 LYS H 4.68 #peak 178 #SUP 0.97 #QF 0.95
63 GLN H 63 GLN HG2 3.59 #peak 204 #SUP 0.64 #QF 0.64
63 GLN H 66 LYS HD3 4.16 #peak 207 #SUP 0.53 #QF 0.53
62 GLU HG2 63 GLN H 3.98 #peak 208 #SUP 0.77 #QF 0.77
65 LYS H 65 LYS HD2 4.85 #peak 215 #SUP 0.73 #QF 0.30

65 LYS H 65 LYS HD3 5.11 #peak 215 #SUP 0.73 #QF 0.30
66 LYS H 67 GLU HG2 4.72 #peak 220 #SUP 0.62 #QF 0.62
68 ALA H 69 ARG H 4.55 #peak 228 #SUP 0.91 #QF 0.91
150 THR H 151 THR H 4.42 #peak 237 #SUP 0.90 #QF 0.90
69 ARG H 70 ILE H 4.13 #peak 238 #SUP 0.83 #QF 0.83
70 ILE H 71 ASN H 4.45 #peak 252 #SUP 0.97 #QF 0.97
71 ASN H 82 VAL QG2 3.80 #peak 255 #SUP 0.72 #QF 0.72
74 ALA H 74 ALA QB 3.83 #peak 259 #SUP 0.87 #QF 0.87
76 GLN H 77 GLY H 4.15 #peak 269 #SUP 0.76 #QF 0.76
76 GLN HB3 77 GLY H 4.84 #peak 274 #SUP 0.96 #QF 0.95
87 ASN H 88 ALA H 4.48 #peak 290 #SUP 0.67 #QF 0.67
86 PRO HB2 87 ASN H 3.62 #peak 297 #SUP 0.61 #QF 0.61
87 ASN H 90 LEU QD1 3.83 #peak 298 #SUP 0.80 #QF 0.80
87 ASN H 90 LEU QD2 4.97 #peak 299 #SUP 0.68 #QF 0.68
87 ASN HA 89 GLU H 3.83 #peak 310 #SUP 0.97 #QF 0.97
87 ASN HB3 89 GLU H 3.74 #peak 311 #SUP 0.90 #QF 0.90
89 GLU H 90 LEU HG 4.25 #peak 315 #SUP 0.63 #QF 0.63
90 LEU H 92 ALA H 4.45 #peak 318 #SUP 0.61 #QF 0.61
88 ALA HA 91 SER H 3.83 #peak 323 #SUP 0.26 #QF 0.26
91 SER H 93 ARG H 4.63 #peak 329 #SUP 0.90 #QF 0.90
94 ALA H 95 LYS H 3.79 #peak 333 #SUP 0.61 #QF 0.61
95 LYS H 96 GLN H 3.86 #peak 340 #SUP 0.96 #QF 0.96
93 ARG HA 95 LYS H 4.57 #peak 341 #SUP 0.84 #QF 0.84
91 SER HA 95 LYS H 4.46 #peak 342 #SUP 0.91 #QF 0.91
95 LYS H 95 LYS HB2 4.14 #peak 343 #SUP 0.94 #QF 0.94
94 ALA QB 95 LYS H 3.54 #peak 344 #SUP 0.92 #QF 0.92
64 ILE QD1 95 LYS H 4.54 #peak 345 #SUP 0.95 #QF 0.95
93 ARG HA 96 GLN H 3.87 #peak 347 #SUP 0.94 #QF 0.94
94 ALA QB 96 GLN H 5.02 #peak 351 #SUP 0.91 #QF 0.44
97 ILE H 98 ALA H 3.59 #peak 356 #SUP 0.88 #QF 0.88
98 ALA H 99 MET H 3.66 #peak 357 #SUP 0.87 #QF 0.87
98 ALA H 100 GLY H 4.17 #peak 358 #SUP 0.97 #QF 0.97
94 ALA HA 98 ALA H 4.10 #peak 359 #SUP 0.85 #QF 0.85
96 GLN HB2 98 ALA H 5.03 #peak 362 #SUP 0.96 #QF 0.96
101 VAL H 102 ASP H 4.39 #peak 384 #SUP 0.77 #QF 0.77
101 VAL HB 102 ASP H 4.60 #peak 388 #SUP 0.84 #QF 0.84
103 GLY H 104 ALA QB 4.25 #peak 398 #SUP 0.50 #QF 0.50
105 ASN HB2 106 GLU H 4.19 #peak 408 #SUP 0.92 #QF 0.66
105 ASN HB3 106 GLU H 4.19 #peak 408 #SUP 0.92 #QF 0.66
104 ALA HA 106 GLU H 4.01 #peak 409 #SUP 0.68 #QF 0.68
112 ARG H 113 GLN H 4.41 #peak 410 #SUP 0.89 #QF 0.89
111 ILE QG2 112 ARG H 3.69 #peak 414 #SUP 0.82 #QF 0.82
117 ILE H 117 ILE QG2 3.53 #peak 433 #SUP 0.66 #QF 0.66
118 GLY H 119 LEU H 4.22 #peak 437 #SUP 0.75 #QF 0.75
155 GLU H 156 VAL H 5.34 #peak 475 #SUP 1.00 #QF 0.98
126 THR H 128 ILE H 4.64 #peak 484 #SUP 0.96 #QF 0.96
127 TRP H 128 ILE H 3.53 #peak 485 #SUP 0.99 #QF 0.99
124 ASN HA 127 TRP H 3.84 #peak 486 #SUP 0.93 #QF 0.93
127 TRP H 128 ILE QD1 4.38 #peak 491 #SUP 0.86 #QF 0.86
125 ASP HA 128 ILE H 4.11 #peak 493 #SUP 0.98 #QF 0.74
127 TRP HB3 128 ILE H 4.03 #peak 494 #SUP 0.93 #QF 0.93
127 TRP HB2 128 ILE H 4.17 #peak 496 #SUP 0.96 #QF 0.96
128 ILE H 128 ILE HB 3.75 #peak 497 #SUP 0.90 #QF 0.90
128 ILE H 151 THR QG2 4.64 #peak 499 #SUP 0.89 #QF 0.89
128 ILE H 128 ILE QG2 3.81 #peak 501 #SUP 0.93 #QF 0.93
129 THR H 132 VAL QG1 4.37 #peak 503 #SUP 0.90 #QF 0.90
129 THR H 151 THR QG2 4.11 #peak 504 #SUP 0.93 #QF 0.93
128 ILE HB 129 THR H 4.03 #peak 507 #SUP 0.89 #QF 0.89
129 THR H 151 THR HB 4.21 #peak 510 #SUP 0.93 #QF 0.93
127 TRP H 129 THR H 4.49 #peak 514 #SUP 0.85 #QF 0.85
129 THR H 130 THR H 3.72 #peak 515 #SUP 0.93 #QF 0.93
131 LYS H 132 VAL H 4.18 #peak 525 #SUP 0.90 #QF 0.90
131 LYS HB3 132 VAL H 4.38 #peak 529 #SUP 0.97 #QF 0.97
132 VAL H 133 ARG H 4.28 #peak 530 #SUP 0.75 #QF 0.75
133 ARG H 134 SER H 3.89 #peak 531 #SUP 0.99 #QF 0.99
132 VAL QG1 133 ARG H 3.84 #peak 533 #SUP 0.83 #QF 0.83
133 ARG H 149 VAL QG1 4.48 #peak 534 #SUP 0.53 #QF 0.53
132 VAL QG2 134 SER H 4.28 #peak 535 #SUP 0.72 #QF 0.72
133 ARG HA 136 LEU H 4.00 #peak 548 #SUP 0.49 #QF 0.49
136 LEU H 137 LEU H 3.77 #peak 549 #SUP 0.88 #QF 0.88
135 GLN H 136 LEU H 3.53 #peak 550 #SUP 0.69 #QF 0.69

137 LEU H 138 THR H 3.65 #peak 559 #SUP 0.95 #QF 0.95
137 LEU HA 138 THR H 3.56 #peak 562 #SUP 0.97 #QF 0.97
137 LEU HA 139 SER H 3.72 #peak 578 #SUP 0.62 #QF 0.62
140 ASP H 141 LEU HG 4.36 #peak 589 #SUP 0.73 #QF 0.73
139 SER H 140 ASP H 4.68 #peak 591 #SUP 0.93 #QF 0.93
140 ASP H 141 LEU H 3.77 #peak 592 #SUP 0.86 #QF 0.86
141 LEU HG 142 VAL H 4.53 #peak 603 #SUP 0.91 #QF 0.91
141 LEU H 142 VAL H 3.73 #peak 606 #SUP 0.74 #QF 0.74
146 ASN HB3 147 VAL H 4.16 #peak 617 #SUP 0.74 #QF 0.74
146 ASN HB2 147 VAL H 3.99 #peak 618 #SUP 0.78 #QF 0.78
150 THR HB 151 THR H 4.57 #peak 632 #SUP 0.76 #QF 0.76
151 THR H 151 THR HB 3.77 #peak 633 #SUP 0.64 #QF 0.64
71 ASN H 72 VAL H 4.70 #peak 639 #SUP 0.92 #QF 0.92
151 THR QG2 154 GLY H 4.13 #peak 644 #SUP 0.31 #QF 0.31
161 LEU H 162 VAL QG1 4.21 #peak 653 #SUP 0.91 #QF 0.91
161 LEU H 161 LEU HB2 3.88 #peak 656 #SUP 0.97 #QF 0.97
161 LEU H 161 LEU HB3 3.88 #peak 657 #SUP 0.95 #QF 0.95
164 GLU H 165 ARG H 3.66 #peak 660 #SUP 0.88 #QF 0.88
163 THR HA 164 GLU H 3.54 #peak 661 #SUP 0.96 #QF 0.96
163 THR HA 165 ARG H 4.80 #peak 671 #SUP 0.82 #QF 0.82
167 ALA H 168 LYS H 3.71 #peak 684 #SUP 0.60 #QF 0.60
170 ALA H 171 ALA H 3.54 #peak 693 #SUP 0.91 #QF 0.91
173 ILE QG2 177 VAL H 4.12 #peak 706 #SUP 0.71 #QF 0.71
174 ALA HA 177 VAL H 3.80 #peak 707 #SUP 0.65 #QF 0.65
177 VAL H 178 SER H 4.58 #peak 710 #SUP 0.72 #QF 0.72
177 VAL HB 178 SER H 4.50 #peak 714 #SUP 0.69 #QF 0.69
179 GLY H 180 VAL H 3.61 #peak 716 #SUP 0.66 #QF 0.66
178 SER HA 180 VAL H 3.84 #peak 727 #SUP 0.37 #QF 0.37
189 PHE H 189 PHE HB3 3.77 #peak 745 #SUP 0.91 #QF 0.91
189 PHE H 190 ILE H 4.65 #peak 752 #SUP 0.96 #QF 0.96
189 PHE HB2 190 ILE H 4.42 #peak 755 #SUP 0.76 #QF 0.76
189 PHE HB3 190 ILE H 4.42 #peak 756 #SUP 0.88 #QF 0.88
189 PHE H 189 PHE HB2 3.77 #peak 792 #SUP 0.84 #QF 0.84
144 SER H 145 SER H 3.60 #peak 804 #SUP 0.93 #QF 0.93
70 ILE HB 71 ASN H 4.31 #peak 815 #SUP 0.95 #QF 0.95
70 ILE HG13 71 ASN H 4.83 #peak 816 #SUP 0.95 #QF 0.95
102 ASP H 104 ALA H 4.38 #peak 818 #SUP 0.76 #QF 0.76
147 VAL HB 148 LYS H 4.37 #peak 821 #SUP 0.95 #QF 0.95
62 GLU H 64 ILE H 4.27 #peak 830 #SUP 0.56 #QF 0.56
61 ASP HB3 62 GLU H 4.20 #peak 831 #SUP 0.80 #QF 0.80
62 GLU H 65 LYS HD3 4.59 #peak 832 #SUP 0.79 #QF 0.79
62 GLU H 65 LYS HG2 4.86 #peak 833 #SUP 0.89 #QF 0.41
62 GLU H 65 LYS HG3 5.16 #peak 833 #SUP 0.89 #QF 0.41
33 LYS HB3 34 ALA H 4.74 #peak 834 #SUP 0.77 #QF 0.66
69 ARG HB3 70 ILE H 5.36 #peak 834 #SUP 0.77 #QF 0.66
149 VAL H 149 VAL HB 3.97 #peak 839 #SUP 0.49 #QF 0.49
187 PHE H 188 THR H 4.85 #peak 840 #SUP 0.93 #QF 0.93
185 THR HB 187 PHE H 4.99 #peak 845 #SUP 0.44 #QF 0.44
49 THR HA 52 VAL H 3.82 #peak 846 #SUP 0.94 #QF 0.94
65 LYS HA 68 ALA H 3.85 #peak 848 #SUP 0.60 #QF 0.60
64 ILE HA 68 ALA H 4.55 #peak 849 #SUP 0.88 #QF 0.88
68 ALA H 70 ILE QD1 5.17 #peak 852 #SUP 0.83 #QF 0.83
54 VAL HB 55 ASN H 3.95 #peak 861 #SUP 0.92 #QF 0.92
165 ARG H 165 ARG HD2 5.16 #peak 862 #SUP 0.66 #QF 0.47
165 ARG H 165 ARG HD3 5.16 #peak 862 #SUP 0.66 #QF 0.47
163 THR QG2 165 ARG H 3.58 #peak 863 #SUP 0.41 #QF 0.41
61 ASP HA 63 GLN H 4.31 #peak 867 #SUP 0.81 #QF 0.81
125 ASP HB2 126 THR H 4.20 #peak 869 #SUP 0.82 #QF 0.82
67 GLU H 90 LEU QD1 3.94 #peak 881 #SUP 0.31 #QF 0.31
142 VAL H 143 LYS H 4.70 #peak 886 #SUP 0.89 #QF 0.89
46 ASP HA 48 GLY H 4.33 #peak 888 #SUP 0.64 #QF 0.64
48 GLY H 49 THR HB 5.12 #peak 890 #SUP 0.57 #QF 0.57
185 THR QG2 186 ALA H 4.36 #peak 918 #SUP 0.71 #QF 0.71
125 ASP H 127 TRP H 4.21 #peak 1047 #SUP 0.82 #QF 0.82
124 ASN H 127 TRP H 5.09 #peak 1048 #SUP 0.73 #QF 0.73
149 VAL H 149 VAL QG1 3.79 #peak 1058 #SUP 0.75 #QF 0.75
37 ASP H 38 PRO HD2 5.50 #peak 1078 #SUP 0.94 #QF 0.79
37 ASP H 38 PRO HD3 5.50 #peak 1078 #SUP 0.94 #QF 0.79
36 THR QG2 37 ASP H 4.27 #peak 1082 #SUP 0.87 #QF 0.44
80 LEU HB2 81 LEU H 4.19 #peak 1083 #SUP 0.43 #QF 0.43
57 ALA H 60 LYS HB3 5.25 #peak 1114 #SUP 0.78 #QF 0.78

20 ILE H 21 ALA QB 4.35 #peak 1118 #SUP 0.76 #QF 0.76
127 TRP HA 131 LYS H 4.32 #peak 1134 #SUP 0.66 #QF 0.66
131 LYS H 131 LYS HB2 4.15 #peak 1136 #SUP 0.91 #QF 0.91
130 THR QG2 131 LYS H 3.85 #peak 1137 #SUP 0.86 #QF 0.86
23 ALA H 24 VAL H 4.23 #peak 1211 #SUP 0.69 #QF 0.69
120 GLY H 122 ALA H 4.38 #peak 1255 #SUP 0.71 #QF 0.71
122 ALA H 124 ASN H 4.55 #peak 1258 #SUP 0.66 #QF 0.66
154 GLY H 155 GLU H 4.11 #peak 1262 #SUP 0.46 #QF 0.46
152 GLU H 155 GLU H 4.05 #peak 1263 #SUP 0.69 #QF 0.69
151 THR QG2 155 GLU H 3.72 #peak 1265 #SUP 0.63 #QF 0.63
149 VAL H 150 THR H 4.46 #peak 1269 #SUP 0.99 #QF 0.99
69 ARG H 70 ILE HG13 3.87 #peak 1273 #SUP 0.88 #QF 0.88
89 GLU H 92 ALA H 4.90 #peak 1276 #SUP 0.74 #QF 0.74
78 LYS H 78 LYS HD3 4.43 #peak 1313 #SUP 0.67 #QF 0.67
191 LYS HA 194 LEU H 3.79 #peak 1318 #SUP 0.27 #QF 0.27
49 THR HA 53 ARG H 3.99 #peak 1323 #SUP 0.83 #QF 0.83
77 GLY H 78 LYS H 3.92 #peak 1334 #SUP 0.84 #QF 0.84
75 TYR H 78 LYS H 4.40 #peak 1337 #SUP 0.41 #QF 0.41
121 GLU HA 125 ASP H 4.39 #peak 1355 #SUP 0.54 #QF 0.54
76 GLN HB2 78 LYS H 3.66 #peak 1363 #SUP 0.44 #QF 0.44
87 ASN HB2 89 GLU H 3.59 #peak 1394 #SUP 0.96 #QF 0.96
163 THR H 167 ALA H 4.51 #peak 1401 #SUP 0.69 #QF 0.69
165 ARG HA 167 ALA H 4.70 #peak 1407 #SUP 0.30 #QF 0.30
96 GLN HB2 97 ILE H 3.63 #peak 1409 #SUP 0.87 #QF 0.87
162 VAL QG2 167 ALA H 3.80 #peak 1411 #SUP 0.75 #QF 0.75
54 VAL H 55 ASN H 3.58 #peak 1417 #SUP 0.84 #QF 0.84
53 ARG H 54 VAL H 3.79 #peak 1418 #SUP 0.86 #QF 0.86
53 ARG HB2 54 VAL H 4.21 #peak 1427 #SUP 0.82 #QF 0.82
53 ARG HB3 54 VAL H 4.21 #peak 1428 #SUP 0.87 #QF 0.87
171 ALA H 174 ALA H 4.97 #peak 1435 #SUP 0.44 #QF 0.44
131 LYS HG2 132 VAL H 3.90 #peak 1445 #SUP 0.98 #QF 0.82
172 ASP H 173 ILE H 3.67 #peak 1454 #SUP 0.48 #QF 0.48
90 LEU H 93 ARG H 4.82 #peak 1462 #SUP 0.41 #QF 0.41
159 MET H 159 MET HG2 5.50 #peak 1473 #SUP 0.81 #QF 0.61
159 MET H 159 MET HG3 5.50 #peak 1473 #SUP 0.81 #QF 0.61
30 VAL H 31 GLY H 4.02 #peak 1511 #SUP 0.26 #QF 0.26
159 MET HA 187 PHE H 4.53 #peak 1520 #SUP 0.82 #QF 0.82
51 GLU H 52 VAL H 3.57 #peak 1524 #SUP 0.92 #QF 0.92
186 ALA H 187 PHE H 4.10 #peak 1534 #SUP 0.61 #QF 0.61
37 ASP HA 39 ARG H 4.42 #peak 1538 #SUP 0.55 #QF 0.55
66 LYS HA 68 ALA H 4.80 #peak 1539 #SUP 0.62 #QF 0.62
39 ARG H 39 ARG HD2 5.50 #peak 1540 #SUP 0.51 #QF 0.34
39 ARG H 39 ARG HD3 5.50 #peak 1540 #SUP 0.51 #QF 0.34
167 ALA H 170 ALA H 5.03 #peak 1546 #SUP 0.97 #QF 0.95
170 ALA H 173 ILE H 5.50 #peak 1546 #SUP 0.97 #QF 0.95
55 ASN H 98 ALA QB 4.65 #peak 1552 #SUP 0.59 #QF 0.59
65 LYS H 68 ALA H 4.49 #peak 1558 #SUP 0.74 #QF 0.74
169 ALA QB 172 ASP H 4.80 #peak 1596 #SUP 0.62 #QF 0.62
96 GLN H 97 ILE H 4.09 #peak 1597 #SUP 0.48 #QF 0.48
94 ALA H 96 GLN H 4.53 #peak 1598 #SUP 0.51 #QF 0.51
96 GLN H 98 ALA QB 4.77 #peak 1611 #SUP 0.61 #QF 0.61
171 ALA H 172 ASP H 3.73 #peak 1612 #SUP 0.87 #QF 0.87
63 GLN H 66 LYS H 4.74 #peak 1634 #SUP 0.87 #QF 0.87
49 THR HA 51 GLU H 4.97 #peak 1637 #SUP 0.80 #QF 0.80
63 GLN H 65 LYS HG3 4.77 #peak 1657 #SUP 0.42 #QF 0.42
130 THR H 131 LYS H 3.58 #peak 1664 #SUP 0.86 #QF 0.86
127 TRP H 130 THR H 4.96 #peak 1665 #SUP 0.75 #QF 0.75
126 THR QG2 130 THR H 4.26 #peak 1667 #SUP 0.33 #QF 0.33
95 LYS H 111 ILE QD1 5.50 #peak 1668 #SUP 0.86 #QF 0.69
65 LYS H 67 GLU H 4.00 #peak 1704 #SUP 0.31 #QF 0.31
126 THR H 151 THR QG2 5.02 #peak 1737 #SUP 0.95 #QF 0.95
126 THR H 128 ILE QD1 5.04 #peak 1738 #SUP 0.88 #QF 0.88
162 VAL H 162 VAL QG2 3.98 #peak 1740 #SUP 1.00
125 ASP HB3 126 THR H 4.20 #peak 1749 #SUP 0.86 #QF 0.86
137 LEU H 139 SER H 4.46 #peak 1771 #SUP 0.94 #QF 0.94
32 THR H 32 THR HB 3.84 #peak 1776 #SUP 0.82 #QF 0.82
32 THR H 32 THR QG2 4.04 #peak 1786 #SUP 0.82 #QF 0.82
145 SER H 146 ASN H 3.76 #peak 1788 #SUP 0.46 #QF 0.46
52 VAL HA 56 SER H 4.08 #peak 1802 #SUP 0.85 #QF 0.85
66 LYS HG2 67 GLU H 4.06 #peak 1880 #SUP 0.57 #QF 0.57
72 VAL H 82 VAL H 5.01 #peak 1954 #SUP 0.56 #QF 0.56

```
 82 VAL  H    110 GLU  H     4.68       #peak  1963 #SUP  0.60  #QF  0.60
 27 THR  H     28 ALA  H     5.11       #peak  1996 #SUP  0.84  #QF  0.77
134 SER  H    136 LEU  H     4.16       #peak  1997 #SUP  0.51  #QF  0.51
130 THR  HA   134 SER  H     4.17       #peak  2008 #SUP  0.89  #QF  0.38
 36 THR  H     37 ASP  HA    5.03       #peak  2040 #SUP  0.95  #QF  0.95
 89 GLU  H     91 SER  H     4.40       #peak  2078 #SUP  0.56  #QF  0.56
 90 LEU  HG    91 SER  H     4.01       #peak  2087 #SUP  0.62  #QF  0.62
 90 LEU  HB3   91 SER  H     3.88       #peak  2089 #SUP  0.65  #QF  0.65
 91 SER  H     94 ALA  H     5.36       #peak  2098 #SUP  1.00  #QF  0.98
 87 ASN  H     87 ASN  HB2   4.01       #peak  2116 #SUP  0.97  #QF  0.97
 87 ASN  H     87 ASN  HB3   3.94       #peak  2117 #SUP  0.84  #QF  0.84
 99 MET  H     99 MET  HG3   4.09       #peak  2200 #SUP  0.91  #QF  0.91
162 VAL  HB   163 THR  H     3.97       #peak  2288 #SUP  0.52  #QF  0.52
162 VAL  QG1  163 THR  H     3.77       #peak  2297 #SUP  0.92  #QF  0.92
162 VAL  QG2  163 THR  H     4.20       #peak  2298 #SUP  0.80  #QF  0.80
163 THR  H    190 ILE  QG2   3.97       #peak  2299 #SUP  0.96  #QF  0.96
148 LYS  H    149 VAL  H     4.65       #peak  2324 #SUP  0.75  #QF  0.75
 55 ASN  H     56 SER  HB2   5.36       #peak  2334 #SUP  0.85  #QF  0.85
 56 SER  H     58 LEU  H     4.32       #peak  2335 #SUP  0.87  #QF  0.87
 58 LEU  H     60 LYS  H     4.31       #peak  2341 #SUP  0.95  #QF  0.95
 61 ASP  HB3   63 GLN  H     4.40       #peak  2346 #SUP  0.81  #QF  0.81
 62 GLU  H     65 LYS  H     5.28       #peak  2351 #SUP  0.73  #QF  0.59
130 THR  H    133 ARG  H     5.50       #peak  2351 #SUP  0.73  #QF  0.59
150 THR  H    157 PHE  H     4.33       #peak  2352 #SUP  0.82  #QF  0.82
 69 ARG  H     84 GLN  H     4.36       #peak  2353 #SUP  0.58  #QF  0.58
100 GLY  H    104 ALA  QB    4.71       #peak  2377 #SUP  0.92  #QF  0.92
126 THR  H    127 TRP  HB3   4.66       #peak  2386 #SUP  0.71  #QF  0.71
133 ARG  H    134 SER  HB2   5.50       #peak  2403 #SUP  0.82  #QF  0.40
133 ARG  H    134 SER  HB3   5.50       #peak  2403 #SUP  0.82  #QF  0.40
137 LEU  H    138 THR  QG2   4.32       #peak  2406 #SUP  0.96  #QF  0.96
136 LEU  H    138 THR  H     4.28       #peak  2407 #SUP  0.84  #QF  0.84
164 GLU  H    190 ILE  QG2   4.98       #peak  2433 #SUP  0.94  #QF  0.94
176 ARG  HB2  177 VAL  H     4.43       #peak  2442 #SUP  0.92  #QF  0.92
162 VAL  QG2  188 THR  H     4.82       #peak  2449 #SUP  0.90  #QF  0.90
 78 LYS  HA   106 GLU  H     4.57       #peak  2629 #SUP  0.98  #QF  0.53
163 THR  HA   190 ILE  H     4.40       #peak  2650 #SUP  0.88  #QF  0.88
 20 ILE  H     20 ILE  QG2   4.04       #peak  2681 #SUP  0.92  #QF  0.92
 30 VAL  QG1   31 GLY  H     4.87       #peak  2700 #SUP  0.98  #QF  0.94
 30 VAL  QG2   31 GLY  H     4.87       #peak  2700 #SUP  0.98  #QF  0.94
 46 ASP  H     49 THR  QG2   3.57       #peak  2746 #SUP  0.64  #QF  0.64
 47 ASP  H     49 THR  H     4.81       #peak  2761 #SUP  0.70  #QF  0.70
 54 VAL  H     55 ASN  HB2   5.20       #peak  2786 #SUP  0.79  #QF  0.79
 54 VAL  H     98 ALA  QB    4.65       #peak  2789 #SUP  0.66  #QF  0.66
 53 ARG  HG3   54 VAL  H     4.97       #peak  2790 #SUP  0.56  #QF  0.56
 55 ASN  H     56 SER  HA    5.50       #peak  2793 #SUP  0.81  #QF  0.81
 57 ALA  H     59 SER  H     4.51       #peak  2799 #SUP  0.99  #QF  0.99
 57 ALA  H     60 LYS  H     4.93       #peak  2801 #SUP  0.99  #QF  0.99
 54 VAL  HA    57 ALA  H     3.92       #peak  2804 #SUP  0.72  #QF  0.72
 57 ALA  H     60 LYS  HB2   5.25       #peak  2806 #SUP  0.72  #QF  0.72
 57 ALA  QB    61 ASP  H     4.55       #peak  2813 #SUP  0.71  #QF  0.71
 59 SER  H     61 ASP  H     4.65       #peak  2815 #SUP  0.90  #QF  0.90
 62 GLU  HA    65 LYS  H     3.54       #peak  2837 #SUP  0.62  #QF  0.62
 62 GLU  HA    66 LYS  H     4.26       #peak  2844 #SUP  0.69  #QF  0.69
 64 ILE  H     66 LYS  H     4.12       #peak  2847 #SUP  0.50  #QF  0.50
 65 LYS  HE2   66 LYS  H     5.35       #peak  2848 #SUP  0.81  #QF  0.36
 65 LYS  HE3   66 LYS  H     5.35       #peak  2848 #SUP  0.81  #QF  0.36
 68 ALA  H     86 PRO  HD2   5.50       #peak  2852 #SUP  0.92  #QF  0.80
 68 ALA  H     86 PRO  HD3   5.50       #peak  2852 #SUP  0.92  #QF  0.80
 73 THR  QG2   74 ALA  H     3.80       #peak  2875 #SUP  0.76  #QF  0.76
 75 TYR  H     79 VAL  HA    4.42       #peak  2884 #SUP  0.94  #QF  0.94
 76 GLN  H     76 GLN  HG2   4.69       #peak  2894 #SUP  0.63  #QF  0.63
 76 GLN  H     76 GLN  HG3   4.69       #peak  2895 #SUP  0.95  #QF  0.86
 76 GLN  HB2   77 GLY  H     4.22       #peak  2898 #SUP  0.84  #QF  0.84
 71 ASN  H     82 VAL  H     4.39       #peak  2912 #SUP  0.72  #QF  0.72
 82 VAL  H     82 VAL  QG1   4.02       #peak  2923 #SUP  0.80  #QF  0.80
 84 GLN  H     85 SER  H     4.06       #peak  2933 #SUP  0.73  #QF  0.73
 69 ARG  H     85 SER  H     4.65       #peak  2939 #SUP  0.81  #QF  0.81
 87 ASN  H     90 LEU  HB2   4.30       #peak  2958 #SUP  0.49  #QF  0.49
 88 ALA  H     89 GLU  H     3.87       #peak  2961 #SUP  0.81  #QF  0.81
 87 ASN  HB2   88 ALA  H     4.20       #peak  2967 #SUP  0.80  #QF  0.80
 87 ASN  HB3   88 ALA  H     3.75       #peak  2968 #SUP  0.77  #QF  0.77
```

88 ALA H 89 GLU HG2 4.83 #peak 2969 #SUP 0.50 #QF 0.50
87 ASN H 89 GLU H 5.38 #peak 2971 #SUP 1.00
89 GLU HG2 90 LEU H 4.09 #peak 2983 #SUP 0.61 #QF 0.38
89 GLU HG3 90 LEU H 4.22 #peak 2983 #SUP 0.61 #QF 0.38
90 LEU HB2 91 SER H 3.64 #peak 3000 #SUP 0.49 #QF 0.49
90 LEU QD1 91 SER H 4.44 #peak 3001 #SUP 0.66 #QF 0.66
90 LEU HB2 92 ALA H 4.97 #peak 3010 #SUP 0.77 #QF 0.77
92 ALA H 94 ALA QB 4.72 #peak 3012 #SUP 0.72 #QF 0.72
92 ALA H 111 ILE QD1 4.28 #peak 3014 #SUP 0.68 #QF 0.68
93 ARG H 93 ARG HD2 4.99 #peak 3022 #SUP 0.97 #QF 0.97
89 GLU HB2 93 ARG H 4.94 #peak 3024 #SUP 0.76 #QF 0.76
95 LYS H 95 LYS HB3 4.14 #peak 3031 #SUP 0.78 #QF 0.78
96 GLN H 98 ALA H 4.44 #peak 3047 #SUP 0.89 #QF 0.89
95 LYS HA 98 ALA H 4.43 #peak 3049 #SUP 0.78 #QF 0.78
99 MET H 100 GLY H 3.79 #peak 3054 #SUP 0.83 #QF 0.83
97 ILE H 100 GLY H 4.98 #peak 3058 #SUP 0.71 #QF 0.71
98 ALA QB 100 GLY H 4.46 #peak 3064 #SUP 0.92 #QF 0.92
104 ALA H 105 ASN H 4.58 #peak 3078 #SUP 0.86 #QF 0.86
101 VAL HA 104 ALA H 4.32 #peak 3079 #SUP 0.46 #QF 0.46
102 ASP HB3 104 ALA H 4.09 #peak 3080 #SUP 0.80 #QF 0.80
105 ASN H 106 GLU H 3.79 #peak 3086 #SUP 0.64 #QF 0.64
113 GLN HA 115 GLN H 5.10 #peak 3103 #SUP 0.84 #QF 0.84
122 ALA QB 126 THR H 5.34 #peak 3108 #SUP 0.56 #QF 0.56
117 ILE H 118 GLY H 4.33 #peak 3113 #SUP 0.91 #QF 0.91
118 GLY H 121 GLU HG2 4.67 #peak 3115 #SUP 0.80 #QF 0.80
118 GLY H 121 GLU HG3 4.67 #peak 3116 #SUP 0.91 #QF 0.91
118 GLY H 122 ALA QB 5.09 #peak 3120 #SUP 0.72 #QF 0.72
58 LEU H 59 SER H 3.73 #peak 3122 #SUP 0.90 #QF 0.90
120 GLY H 122 ALA QB 4.92 #peak 3131 #SUP 0.76 #QF 0.76
120 GLY H 123 SER H 4.86 #peak 3142 #SUP 0.89 #QF 0.89
70 ILE HG13 85 SER H 4.98 #peak 3146 #SUP 0.48 #QF 0.48
127 TRP H 128 ILE HB 5.20 #peak 3156 #SUP 0.95 #QF 0.95
133 ARG H 133 ARG HB3 3.68 #peak 3181 #SUP 0.97 #QF 0.97
133 ARG H 133 ARG HB2 3.68 #peak 3182 #SUP 0.65 #QF 0.65
132 VAL QG2 133 ARG H 3.68 #peak 3185 #SUP 0.97 #QF 0.97
133 ARG H 149 VAL QG2 4.48 #peak 3186 #SUP 0.66 #QF 0.66
135 GLN HG3 136 LEU H 4.71 #peak 3199 #SUP 0.94 #QF 0.94
136 LEU H 138 THR QG2 4.57 #peak 3200 #SUP 0.87 #QF 0.87
140 ASP H 141 LEU QD1 4.48 #peak 3220 #SUP 0.72 #QF 0.72
140 ASP HA 142 VAL H 5.16 #peak 3226 #SUP 0.71 #QF 0.71
142 VAL H 142 VAL HB 4.06 #peak 3228 #SUP 0.86 #QF 0.86
142 VAL HB 143 LYS H 4.45 #peak 3235 #SUP 0.85 #QF 0.85
39 ARG HB3 40 SER H 4.49 #peak 3257 #SUP 0.88 #QF 0.88
146 ASN H 146 ASN HB3 4.07 #peak 3261 #SUP 0.93 #QF 0.93
146 ASN H 146 ASN HB2 4.01 #peak 3262 #SUP 0.95 #QF 0.95
147 VAL H 148 LYS H 4.50 #peak 3281 #SUP 0.71 #QF 0.71
148 LYS H 159 MET H 3.84 #peak 3282 #SUP 0.94 #QF 0.94
129 THR QG2 150 THR H 4.90 #peak 3303 #SUP 0.97 #QF 0.97
153 ASN HA 155 GLU H 4.73 #peak 3311 #SUP 0.36 #QF 0.36
124 ASN H 128 ILE QD1 4.89 #peak 3320 #SUP 0.93 #QF 0.93
150 THR QG2 158 LEU H 5.25 #peak 3331 #SUP 0.67 #QF 0.67
161 LEU H 162 VAL H 4.65 #peak 3351 #SUP 0.56 #QF 0.56
162 VAL H 167 ALA QB 4.65 #peak 3368 #SUP 0.84 #QF 0.78
162 VAL H 188 THR QG2 4.97 #peak 3368 #SUP 0.84 #QF 0.78
163 THR HB 165 ARG H 4.73 #peak 3387 #SUP 0.80 #QF 0.80
164 GLU HA 167 ALA H 3.80 #peak 3397 #SUP 0.64 #QF 0.64
162 VAL HB 167 ALA H 4.07 #peak 3400 #SUP 0.71 #QF 0.71
167 ALA H 170 ALA QB 4.60 #peak 3404 #SUP 0.88 #QF 0.88
166 GLU HA 169 ALA H 3.90 #peak 3414 #SUP 0.80 #QF 0.80
169 ALA H 172 ASP H 4.92 #peak 3418 #SUP 0.87 #QF 0.87
168 LYS HA 170 ALA H 4.75 #peak 3420 #SUP 0.89 #QF 0.89
166 GLU HA 170 ALA H 4.06 #peak 3421 #SUP 0.93 #QF 0.93
170 ALA H 173 ILE HG12 5.11 #peak 3424 #SUP 0.49 #QF 0.49
170 ALA H 172 ASP H 4.38 #peak 3428 #SUP 0.94 #QF 0.94
169 ALA HA 172 ASP H 3.91 #peak 3443 #SUP 0.80 #QF 0.80
172 ASP H 172 ASP HB2 3.56 #peak 3445 #SUP 0.71 #QF 0.71
173 ILE HG12 174 ALA H 4.43 #peak 3462 #SUP 0.46 #QF 0.46
23 ALA QB 24 VAL H 3.77 #peak 3508 #SUP 0.52 #QF 0.52
186 ALA QB 187 PHE H 3.90 #peak 3510 #SUP 0.55 #QF 0.55
161 LEU H 188 THR H 5.15 #peak 3512 #SUP 0.47 #QF 0.47
183 VAL H 183 VAL HB 3.83 #peak 3518 #SUP 0.30 #QF 0.30

191 LYS HA 193 GLY H 3.87 #peak 3535 #SUP 0.56 #QF 0.56
156 VAL H 183 VAL QG2 5.38 #peak 3617 #SUP 0.74 #QF 0.74
81 LEU QD1 110 GLU H 5.50 #peak 3673 #SUP 0.84 #QF 0.57
81 LEU QD2 110 GLU H 5.50 #peak 3673 #SUP 0.84 #QF 0.57
87 ASN H 90 LEU H 4.62 #peak 3841 #SUP 0.90 #QF 0.90
168 LYS HA 171 ALA H 4.28 #peak 3865 #SUP 0.32 #QF 0.32
128 ILE QG2 131 LYS H 4.73 #peak 4019 #SUP 0.75 #QF 0.75
33 LYS H 34 ALA H 4.81 #peak 4028 #SUP 0.78 #QF 0.44
20 ILE HG12 21 ALA H 4.93 #peak 4038 #SUP 0.88 #QF 0.88
51 GLU H 51 GLU HB3 3.90 #peak 4059 #SUP 0.96 #QF 0.96
51 GLU H 74 ALA QB 3.87 #peak 4060 #SUP 0.94 #QF 0.94
52 VAL H 55 ASN H 4.88 #peak 4061 #SUP 0.87 #QF 0.87
51 GLU HB2 52 VAL H 3.72 #peak 4065 #SUP 0.94 #QF 0.94
52 VAL HB 53 ARG H 3.79 #peak 4068 #SUP 1.00
61 ASP HB2 64 ILE H 3.80 #peak 4079 #SUP 0.72 #QF 0.72
66 LYS H 68 ALA H 4.00 #peak 4081 #SUP 0.74 #QF 0.74
69 ARG H 70 ILE QD1 3.63 #peak 4083 #SUP 0.76 #QF 0.76
72 VAL QG2 73 THR H 4.16 #peak 4086 #SUP 0.94 #QF 0.94
93 ARG H 94 ALA HA 5.50 #peak 4120 #SUP 0.78 #QF 0.78
94 ALA H 97 ILE H 4.98 #peak 4124 #SUP 0.94 #QF 0.94
95 LYS HB3 96 GLN H 4.54 #peak 4129 #SUP 0.75 #QF 0.75
95 LYS HB2 96 GLN H 4.54 #peak 4130 #SUP 0.96 #QF 0.92
93 ARG HA 97 ILE H 4.30 #peak 4131 #SUP 0.35 #QF 0.35
110 GLU H 111 ILE QD1 4.39 #peak 4151 #SUP 0.95 #QF 0.95
114 GLY H 115 GLN H 3.77 #peak 4171 #SUP 0.72 #QF 0.72
130 THR H 132 VAL H 4.85 #peak 4189 #SUP 0.91 #QF 0.61
132 VAL H 134 SER H 5.13 #peak 4189 #SUP 0.91 #QF 0.61
129 THR QG2 133 ARG H 4.43 #peak 4191 #SUP 0.91 #QF 0.91
139 SER HA 141 LEU H 4.59 #peak 4197 #SUP 0.82 #QF 0.82
170 ALA HA 173 ILE H 4.33 #peak 4232 #SUP 0.61 #QF 0.61
189 PHE H 189 PHE HD2 3.60 #peak 3 #SUP 0.69 #QF 0.69
189 PHE HD1 190 ILE H 4.13 #peak 4 #SUP 0.71 #QF 0.71
164 GLU HA 189 PHE HD1 3.51 #peak 6 #SUP 0.87 #QF 0.87
127 TRP HD1 128 ILE HA 4.72 #peak 18 #SUP 0.95 #QF 0.95
127 TRP H 127 TRP HD1 4.64 #peak 22 #SUP 0.56 #QF 0.56
127 TRP HD1 128 ILE H 4.65 #peak 23 #SUP 0.81 #QF 0.81
75 TYR HD2 78 LYS HD2 3.47 #peak 34 #SUP 0.89 #QF 0.89
75 TYR HD1 188 THR QG2 4.60 #peak 36 #SUP 0.87 #QF 0.87
75 TYR HD1 161 LEU HG 4.63 #peak 37 #SUP 0.84 #QF 0.84
75 TYR H 75 TYR HD2 4.54 #peak 39 #SUP 0.47 #QF 0.47
75 TYR HE1 161 LEU H 3.98 #peak 49 #SUP 0.90 #QF 0.90
75 TYR HE1 162 VAL H 5.50 #peak 50 #SUP 0.98 #QF 0.98
75 TYR HE2 78 LYS HD2 3.63 #peak 52 #SUP 0.95 #QF 0.95
75 TYR HE1 188 THR QG2 4.28 #peak 54 #SUP 0.81 #QF 0.81
127 TRP HH2 131 LYS HA 5.02 #peak 76 #SUP 0.70 #QF 0.70
75 TYR HB2 75 TYR HE1 5.26 #peak 97 #SUP 0.99 #QF 0.96
75 TYR HB2 75 TYR HD1 3.65 #peak 105 #SUP 0.98 #QF 0.96
78 LYS HD2 108 TYR HD1 5.06 #peak 107 #SUP 0.97 #QF 0.69
80 LEU HB2 108 TYR HD1 4.89 #peak 109 #SUP 0.58 #QF 0.58
188 THR HA 189 PHE HD1 5.50 #peak 130 #SUP 0.65 #QF 0.21
188 THR HA 189 PHE HD2 5.50 #peak 130 #SUP 0.65 #QF 0.21
167 ALA H 189 PHE HD1 4.75 #peak 133 #SUP 0.91 #QF 0.91
168 LYS H 189 PHE HD1 5.24 #peak 135 #SUP 0.81 #QF 0.81
75 TYR HE1 188 THR H 5.01 #peak 152 #SUP 0.85 #QF 0.85
75 TYR HE1 160 GLY H 5.25 #peak 153 #SUP 0.93 #QF 0.82
75 TYR HE2 160 GLY H 5.50 #peak 153 #SUP 0.93 #QF 0.82
20 ILE HA 20 ILE QG1 3.22 #peak 1420
20 ILE QG2 20 ILE QG1 3.03 #peak 210
24 VAL H 24 VAL QQG 4.20 #peak 227
25 VAL QQG 26 GLY QA 3.44 #peak 247
29 ALA HA 30 VAL QQG 3.37 #peak 10
30 VAL QQG 31 GLY H 4.08 #peak 2700
33 LYS H 33 LYS QE 5.34 #peak 4467
33 LYS QB 33 LYS QE 3.88 #peak 3596
33 LYS QB 34 ALA H 4.04 #peak 834
36 THR H 37 ASP QB 4.83 #peak 2712
36 THR HA 37 ASP QB 4.88 #peak 232
37 ASP H 38 PRO QD 4.76 #peak 1078
37 ASP QB 38 PRO HA 4.76 #peak 1777
37 ASP QB 38 PRO HB2 5.34 #peak 3276
37 ASP QB 38 PRO QG 3.98 #peak 1249

37 ASP QB 38 PRO QD 3.15 #peak 1255
37 ASP QB 39 ARG H 3.97 #peak 1592
37 ASP QB 40 SER H 4.39 #peak 3255
39 ARG HA 39 ARG QD 3.27 #peak 2688
39 ARG QB 39 ARG QD 3.33 #peak 2853
39 ARG QB 40 SER H 3.68 #peak 3257
40 SER H 41 VAL QQG 4.51 #peak 877
40 SER QB 41 VAL QQG 3.96 #peak 1766
41 VAL QQG 44 GLN QB 3.55 #peak 1486
41 VAL QQG 44 GLN QG 4.17 #peak 1281
45 VAL H 45 VAL QQG 4.35 #peak 4880
45 VAL HB 50 LEU QQD 3.71 #peak 1290
45 VAL QQG 46 ASP QB 5.07 #peak 4388
45 VAL QQG 47 ASP H 4.49 #peak 1274
45 VAL QQG 103 GLY HA2 3.87 #peak 9
45 VAL QQG 103 GLY HA3 3.38 #peak 2030
46 ASP H 50 LEU QQD 4.35 #peak 837
46 ASP QB 48 GLY H 3.84 #peak 892
46 ASP QB 49 THR H 3.78 #peak 2769
46 ASP QB 49 THR HB 3.40 #peak 1312
46 ASP QB 49 THR QG2 3.66 #peak 262
48 GLY HA2 51 GLU QB 3.96 #peak 3804
49 THR HA 53 ARG QD 3.76 #peak 4426
49 THR HB 50 LEU QQD 4.66 #peak 2208
49 THR QG2 53 ARG QD 4.13 #peak 256
50 LEU H 50 LEU QB 3.54 #peak 1159
50 LEU HA 50 LEU QQD 3.34 #peak 3777
50 LEU QB 74 ALA QB 3.11 #peak 1704
50 LEU QB 104 ALA QB 4.97 #peak 380
50 LEU QQD 51 GLU H 4.44 #peak 2777
50 LEU QQD 102 ASP HB2 4.73 #peak 2024
50 LEU QQD 102 ASP HB3 4.49 #peak 2025
50 LEU QQD 103 GLY H 4.84 #peak 396
50 LEU QQD 103 GLY HA2 4.78 #peak 2027
50 LEU QQD 103 GLY HA3 3.89 #peak 2031
50 LEU QQD 104 ALA HA 3.48 #peak 2034
51 GLU H 51 GLU QB 3.36 #peak 4059
51 GLU H 51 GLU QG 3.78 #peak 119
51 GLU H 54 VAL QQG 4.06 #peak 2776
51 GLU HA 51 GLU QG 3.24 #peak 2735
51 GLU HA 54 VAL QQG 3.15 #peak 1755
51 GLU HA 72 VAL QQG 4.63 #peak 1334
51 GLU QB 52 VAL H 3.26 #peak 125
51 GLU QB 52 VAL QG2 3.59 #peak 3287
51 GLU QB 54 VAL QQG 4.46 #peak 1157
51 GLU QB 72 VAL QQG 4.83 #peak 1157
51 GLU QB 74 ALA QB 4.16 #peak 1702
51 GLU QG 54 VAL QQG 4.41 #peak 1344
51 GLU QG 72 VAL QQG 3.67 #peak 4417
51 GLU QG 73 THR HA 3.62 #peak 1669
51 GLU QG 73 THR QG2 5.13 #peak 1684
51 GLU QG 74 ALA H 4.35 #peak 260
51 GLU QG 74 ALA QB 3.13 #peak 1703
52 VAL H 54 VAL QQG 4.05 #peak 1532
53 ARG H 54 VAL QQG 3.99 #peak 1331
53 ARG HA 53 ARG QD 3.54 #peak 3761
53 ARG QB 53 ARG QD 2.88 #peak 2603
53 ARG QB 54 VAL H 3.47 #peak 1427
53 ARG QB 54 VAL QQG 4.33 #peak 1120
53 ARG QB 55 ASN H 5.30 #peak 4073
53 ARG QD 102 ASP HB2 4.58 #peak 372
53 ARG QD 102 ASP HB3 5.13 #peak 371
54 VAL H 54 VAL QQG 3.12 #peak 4217
54 VAL HA 54 VAL QQG 3.13 #peak 3815
54 VAL HA 72 VAL QQG 4.28 #peak 3826
54 VAL HB 81 LEU QQD 4.20 #peak 1991
54 VAL QQG 55 ASN H 4.12 #peak 427
54 VAL QQG 55 ASN HB2 3.83 #peak 1367
54 VAL QQG 55 ASN HB3 4.53 #peak 1370
54 VAL QQG 56 SER H 4.34 #peak 149
54 VAL QQG 72 VAL HB 4.47 #peak 1655

54 VAL QQG 81 LEU QQD 3.28 #peak 4294
54 VAL QQG 98 ALA H 4.56 #peak 3050
55 ASN H 56 SER QB 4.64 #peak 138
55 ASN H 58 LEU QB 5.18 #peak 4074
55 ASN HA 58 LEU QB 3.72 #peak 3875
55 ASN HA 58 LEU QQD 4.01 #peak 1408
55 ASN HA 72 VAL QQG 3.36 #peak 3252
55 ASN HB2 72 VAL QQG 4.53 #peak 1368
55 ASN HB3 56 SER QB 4.39 #peak 1356
55 ASN HB3 58 LEU QB 5.26 #peak 3288
55 ASN HB3 72 VAL QQG 4.99 #peak 1370
56 SER QB 57 ALA QB 4.03 #peak 4038
57 ALA H 60 LYS QB 4.46 #peak 1114
57 ALA H 60 LYS QD 5.19 #peak 2472
57 ALA HA 60 LYS QB 3.53 #peak 1385
57 ALA HA 60 LYS QD 3.22 #peak 1386
57 ALA HA 60 LYS QE 4.60 #peak 1383
57 ALA QB 60 LYS QB 3.58 #peak 1436
57 ALA QB 60 LYS QD 3.67 #peak 1922
58 LEU H 58 LEU QQD 4.19 #peak 2418
58 LEU H 72 VAL QQG 4.89 #peak 319
58 LEU QB 60 LYS H 5.01 #peak 178
58 LEU QB 72 VAL QQG 3.92 #peak 1397
58 LEU QQD 59 SER H 4.18 #peak 1872
58 LEU QQD 60 LYS H 5.01 #peak 866
60 LYS H 60 LYS QD 4.77 #peak 1445
60 LYS HA 60 LYS QE 4.35 #peak 2746
60 LYS QB 60 LYS QE 3.49 #peak 1574
62 GLU HA 65 LYS QE 3.42 #peak 3557
63 GLN HA 66 LYS QB 4.02 #peak 1515
63 GLN HG2 64 ILE QG1 3.90 #peak 3880
63 GLN HG3 64 ILE QG1 4.02 #peak 1477
64 ILE H 64 ILE QG1 3.64 #peak 2832
64 ILE HA 64 ILE QG1 3.40 #peak 3521
64 ILE QG2 64 ILE QG1 3.19 #peak 3435
64 ILE QD1 93 ARG QB 4.99 #peak 4943
64 ILE QD1 93 ARG QD 3.19 #peak 1949
65 LYS QB 65 LYS QE 3.97 #peak 3198
65 LYS QE 66 LYS H 4.64 #peak 2848
65 LYS QE 66 LYS QB 4.56 #peak 2772
66 LYS QB 66 LYS HD2 3.56 #peak 1527
66 LYS QB 66 LYS HD3 3.41 #peak 1527
66 LYS QB 66 LYS HE3 4.89 #peak 2773
68 ALA H 86 PRO QD 4.75 #peak 2852
68 ALA HA 69 ARG QB 4.88 #peak 1561
68 ALA HA 86 PRO QD 3.89 #peak 1556
68 ALA QB 86 PRO QD 3.26 #peak 302
69 ARG H 69 ARG QB 3.39 #peak 248
70 ILE QG2 71 ASN QB 4.42 #peak 1613
70 ILE HG13 91 SER QB 4.35 #peak 4038
71 ASN HA 72 VAL QQG 4.01 #peak 3665
71 ASN QB 72 VAL H 4.26 #peak 786
71 ASN QB 82 VAL H 4.35 #peak 4479
71 ASN QB 82 VAL QG1 5.25 #peak 2199
72 VAL H 81 LEU QQD 4.72 #peak 789
72 VAL HA 72 VAL QQG 2.88 #peak 2803
72 VAL HA 81 LEU QB 4.11 #peak 1654
72 VAL HA 81 LEU QQD 4.58 #peak 637
72 VAL HB 81 LEU QQD 4.26 #peak 3902
72 VAL QQG 73 THR H 3.26 #peak 1752
72 VAL QQG 74 ALA H 4.61 #peak 2876
72 VAL QQG 81 LEU QB 3.99 #peak 4158
73 THR H 80 LEU QQD 4.40 #peak 4529
73 THR H 81 LEU QQD 3.95 #peak 794
73 THR HA 80 LEU QQD 5.39 #peak 4531
73 THR HB 80 LEU QQD 3.38 #peak 1681
73 THR HB 81 LEU QQD 4.76 #peak 3376
73 THR QG2 80 LEU QQD 3.56 #peak 330
74 ALA H 79 VAL QQG 5.28 #peak 2876
74 ALA HA 79 VAL QQG 3.91 #peak 1693
75 TYR H 79 VAL QQG 4.38 #peak 2889

75 TYR HA 76 GLN QG 3.85 #peak 576
75 TYR HB2 76 GLN QG 4.17 #peak 1729
75 TYR HB3 76 GLN QG 4.64 #peak 1730
75 TYR HD1 76 GLN QG 3.54 #peak 35
75 TYR HE1 76 GLN QG 3.51 #peak 1735
75 TYR HE1 161 LEU QB 3.62 #peak 53
75 TYR HE1 161 LEU QQD 3.74 #peak 2417
75 TYR HE2 78 LYS QE 4.18 #peak 5073
75 TYR HD2 78 LYS QE 4.84 #peak 5075
76 GLN H 76 GLN QG 3.87 #peak 2895
76 GLN QG 77 GLY H 5.12 #peak 274
76 GLN QG 78 LYS H 5.34 #peak 1738
77 GLY QA 105 ASN QB 4.26 #peak 83
78 LYS H 78 LYS QE 5.04 #peak 4059
78 LYS HA 78 LYS QE 3.97 #peak 3228
78 LYS HA 106 GLU QB 4.29 #peak 3834
78 LYS HA 106 GLU QG 4.34 #peak 1038
78 LYS QE 79 VAL H 4.03 #peak 3290
78 LYS QE 79 VAL HA 4.17 #peak 344
78 LYS QE 106 GLU QB 4.21 #peak 4649
78 LYS QE 106 GLU QG 4.76 #peak 3851
78 LYS QE 108 TYR HD1 3.58 #peak 144
79 VAL H 79 VAL QQG 3.42 #peak 1753
79 VAL HA 79 VAL QQG 3.14 #peak 3202
79 VAL QQG 80 LEU H 3.60 #peak 4433
79 VAL QQG 99 MET QG 3.35 #peak 1756
80 LEU H 80 LEU QQD 4.20 #peak 1187
80 LEU HA 80 LEU QQD 4.05 #peak 1189
80 LEU QQD 81 LEU H 4.64 #peak 4528
80 LEU QQD 82 VAL QG1 3.22 #peak 1193
80 LEU QQD 108 TYR HD1 5.44 #peak 3351
81 LEU H 81 LEU QQD 3.63 #peak 664
81 LEU HA 81 LEU QQD 3.62 #peak 3425
81 LEU QB 82 VAL H 4.29 #peak 3955
81 LEU QB 109 ASN HA 4.65 #peak 5081
81 LEU QB 111 ILE QD1 4.55 #peak 877
81 LEU QQD 82 VAL H 4.10 #peak 665
81 LEU QQD 82 VAL QG1 5.17 #peak 3375
81 LEU QQD 95 LYS H 4.17 #peak 3034
81 LEU QQD 95 LYS HA 3.61 #peak 3302
81 LEU QQD 95 LYS QB 3.89 #peak 631
81 LEU QQD 98 ALA H 5.24 #peak 3345
81 LEU QQD 98 ALA QB 3.13 #peak 633
81 LEU QQD 99 MET H 4.54 #peak 3344
81 LEU QQD 109 ASN HA 4.98 #peak 637
81 LEU QQD 109 ASN QB 4.25 #peak 4353
85 SER H 91 SER QB 4.45 #peak 1896
86 PRO QD 87 ASN H 3.49 #peak 296
86 PRO QD 90 LEU QD1 4.10 #peak 3917
89 GLU HG3 93 ARG QD 4.26 #peak 583
90 LEU QD2 93 ARG QD 3.89 #peak 1869
91 SER QB 92 ALA H 3.79 #peak 3008
91 SER QB 93 ARG H 5.34 #peak 4594
91 SER QB 95 LYS H 5.34 #peak 2017
91 SER QB 111 ILE H 4.83 #peak 2016
91 SER QB 111 ILE HB 4.07 #peak 4661
91 SER QB 111 ILE QD1 2.86 #peak 2318
93 ARG H 93 ARG QG 5.25 #peak 1170
93 ARG HA 93 ARG QG 3.31 #peak 4147
93 ARG HA 93 ARG QD 4.81 #peak 4145
93 ARG QB 96 GLN HG3 4.21 #peak 4617
93 ARG QG 94 ALA H 4.35 #peak 1171
93 ARG QG 94 ALA HA 3.79 #peak 1913
93 ARG QG 95 LYS H 5.16 #peak 3032
93 ARG QD 94 ALA H 4.83 #peak 1943
93 ARG QD 94 ALA HA 5.10 #peak 1940
93 ARG QD 94 ALA QB 4.57 #peak 1948
95 LYS H 95 LYS QB 3.55 #peak 343
95 LYS H 95 LYS QD 4.50 #peak 4174
95 LYS HA 95 LYS QD 4.27 #peak 3531
95 LYS QB 95 LYS QD 2.91 #peak 4606

95 LYS QB 95 LYS QE 4.01 #peak 4607
95 LYS QB 96 GLN H 3.81 #peak 4129
95 LYS QB 109 ASN QB 3.72 #peak 3641
95 LYS QD 96 GLN H 4.02 #peak 1620
95 LYS QD 109 ASN HA 5.34 #peak 4175
95 LYS QE 96 GLN HA 4.34 #peak 4183
98 ALA QB 99 MET QB 4.20 #peak 1976
98 ALA QB 99 MET QG 3.82 #peak 75
99 MET H 99 MET QB 3.50 #peak 368
99 MET H 99 MET QG 3.58 #peak 77
99 MET QB 104 ALA QB 3.77 #peak 379
99 MET QG 104 ALA HA 5.34 #peak 1985
99 MET QG 104 ALA QB 3.98 #peak 1981
101 VAL QQG 102 ASP HB2 4.18 #peak 2023
101 VAL QQG 102 ASP HB3 4.94 #peak 79
101 VAL QQG 103 GLY H 3.97 #peak 395
104 ALA HA 105 ASN QB 5.34 #peak 3942
105 ASN QB 106 GLU H 3.65 #peak 408
105 ASN QB 106 GLU QB 4.21 #peak 2041
105 ASN QB 106 GLU QG 3.56 #peak 81
106 GLU H 106 GLU QB 3.56 #peak 854
106 GLU QB 108 TYR HD1 5.17 #peak 107
106 GLU QB 108 TYR HE1 3.53 #peak 68
106 GLU QG 108 TYR HE1 4.13 #peak 1041
113 GLN HA 114 GLY QA 4.26 #peak 4674
113 GLN QB 114 GLY QA 5.18 #peak 2067
113 GLN QG 114 GLY H 4.84 #peak 2061
113 GLN QG 114 GLY QA 3.40 #peak 90
113 GLN QG 115 GLN H 3.80 #peak 3105
114 GLY QA 115 GLN QB 5.18 #peak 2067
114 GLY QA 115 GLN HG2 4.66 #peak 2097
117 ILE HB 121 GLU QG 4.88 #peak 2101
117 ILE QG2 121 GLU QB 4.20 #peak 3790
117 ILE QG2 121 GLU QG 3.21 #peak 2108
117 ILE QG2 153 ASN QB 4.37 #peak 2105
117 ILE QD1 121 GLU QB 3.84 #peak 4724
117 ILE QD1 121 GLU QG 3.47 #peak 595
117 ILE QD1 153 ASN QB 4.86 #peak 1011
118 GLY H 121 GLU QG 3.91 #peak 3116
118 GLY QA 119 LEU HG 4.39 #peak 2121
118 GLY QA 119 LEU QQD 4.16 #peak 5264
118 GLY QA 120 GLY H 4.75 #peak 3126
118 GLY QA 121 GLU QB 5.05 #peak 4706
118 GLY QA 121 GLU QG 3.87 #peak 4727
118 GLY QA 122 ALA H 4.18 #peak 463
118 GLY QA 122 ALA QB 4.25 #peak 2134
119 LEU QQD 120 GLY H 4.35 #peak 3132
120 GLY QA 123 SER H 4.19 #peak 3145
121 GLU H 121 GLU QG 3.72 #peak 459
121 GLU HA 124 ASN QB 3.69 #peak 2150
121 GLU QB 122 ALA QB 5.34 #peak 4721
121 GLU QB 153 ASN HA 3.72 #peak 488
121 GLU QG 122 ALA H 4.02 #peak 1260
121 GLU QG 122 ALA QB 3.26 #peak 2847
121 GLU QG 123 SER H 4.67 #peak 1905
121 GLU QG 153 ASN HA 4.47 #peak 489
122 ALA HA 125 ASP QB 3.42 #peak 3399
122 ALA QB 125 ASP QB 4.04 #peak 2131
123 SER H 124 ASN QB 4.96 #peak 883
123 SER H 125 ASP QB 4.76 #peak 882
123 SER QB 124 ASN HA 4.22 #peak 4736
123 SER QB 124 ASN QB 4.96 #peak 4436
124 ASN QB 125 ASP H 3.63 #peak 481
124 ASN QB 125 ASP HA 5.00 #peak 3318
124 ASN QB 127 TRP H 5.34 #peak 2389
124 ASN QB 127 TRP HB2 5.34 #peak 3317
124 ASN QB 128 ILE H 4.88 #peak 3548
124 ASN QB 128 ILE QD1 3.43 #peak 4006
125 ASP H 125 ASP QB 3.54 #peak 482
125 ASP QB 126 THR H 3.42 #peak 869
125 ASP QB 127 TRP H 5.34 #peak 2389

125 ASP QB 128 ILE QD1 4.60 #peak 594
125 ASP QB 151 THR HB 3.69 #peak 1034
125 ASP QB 151 THR QG2 3.57 #peak 1302
127 TRP HD1 131 LYS QE 4.62 #peak 952
127 TRP HZ2 131 LYS QE 3.90 #peak 951
127 TRP HH2 131 LYS QE 4.13 #peak 953
128 ILE H 131 LYS QE 4.68 #peak 954
128 ILE HA 131 LYS QB 3.93 #peak 2162
128 ILE HA 131 LYS QE 3.47 #peak 18
128 ILE QG2 131 LYS QE 4.42 #peak 4480
129 THR HA 149 VAL QQG 3.45 #peak 743
129 THR HA 156 VAL QQG 4.24 #peak 3949
129 THR QG2 133 ARG QD 3.08 #peak 712
130 THR H 149 VAL QQG 4.81 #peak 3167
130 THR HA 133 ARG QB 3.45 #peak 271
130 THR HA 133 ARG QD 3.19 #peak 5106
131 LYS H 131 LYS QB 3.49 #peak 1136
131 LYS H 131 LYS QE 4.27 #peak 3493
131 LYS QB 131 LYS HD3 3.13 #peak 3984
131 LYS QB 131 LYS QE 3.28 #peak 3274
131 LYS QB 132 VAL HA 5.34 #peak 3750
131 LYS QB 132 VAL QG1 4.16 #peak 1109
131 LYS QB 135 GLN HG3 5.34 #peak 126
131 LYS HG2 131 LYS QE 3.18 #peak 21
131 LYS QE 132 VAL QG1 4.95 #peak 950
132 VAL HA 135 GLN QB 3.84 #peak 3749
133 ARG H 133 ARG QB 3.21 #peak 3181
133 ARG H 149 VAL QQG 3.83 #peak 534
133 ARG HA 137 LEU QQD 3.82 #peak 2263
133 ARG QB 133 ARG QD 3.28 #peak 5105
133 ARG QB 134 SER H 3.51 #peak 2014
133 ARG QD 137 LEU QQD 4.99 #peak 2259
134 SER H 137 LEU QQD 4.66 #peak 2019
135 GLN QB 138 THR H 5.34 #peak 3211
136 LEU H 137 LEU QQD 4.96 #peak 3201
137 LEU H 137 LEU QB 3.43 #peak 555
137 LEU H 142 VAL QQG 4.55 #peak 558
137 LEU HA 142 VAL QQG 3.58 #peak 478
137 LEU QB 138 THR H 3.53 #peak 571
137 LEU QB 138 THR QG2 3.89 #peak 2252
139 SER QB 140 ASP H 4.14 #peak 588
139 SER QB 141 LEU H 4.34 #peak 3222
139 SER QB 142 VAL H 4.33 #peak 2413
139 SER QB 142 VAL HB 3.47 #peak 4838
139 SER QB 142 VAL QQG 2.88 #peak 4817
140 ASP QB 141 LEU H 3.83 #peak 596
140 ASP QB 141 LEU QD1 4.12 #peak 2291
140 ASP QB 141 LEU QD2 4.79 #peak 3363
141 LEU H 165 ARG QB 5.34 #peak 2224
141 LEU QB 141 LEU QD2 3.09 #peak 132
141 LEU QB 142 VAL H 3.73 #peak 602
141 LEU QB 142 VAL HA 4.12 #peak 2304
141 LEU QD1 165 ARG QB 3.14 #peak 2292
141 LEU QD1 165 ARG QD 3.32 #peak 2290
141 LEU QD2 165 ARG QD 3.63 #peak 3364
142 VAL H 165 ARG QB 5.34 #peak 2223
142 VAL HA 166 GLU QB 4.10 #peak 2303
142 VAL HA 166 GLU QG 3.35 #peak 4212
142 VAL QQG 146 ASN HB3 5.38 #peak 2987
142 VAL QQG 162 VAL QG1 4.11 #peak 484
142 VAL QQG 162 VAL QG2 5.24 #peak 4845
142 VAL QQG 166 GLU QB 4.12 #peak 4942
142 VAL QQG 166 GLU QG 3.36 #peak 473
143 LYS H 166 GLU QG 4.27 #peak 3236
146 ASN HA 147 VAL QQG 4.29 #peak 455
146 ASN HB2 147 VAL QQG 4.33 #peak 4857
146 ASN HB3 147 VAL QQG 4.21 #peak 986
147 VAL HA 148 LYS QB 5.16 #peak 460
147 VAL HA 159 MET QB 5.34 #peak 2100
147 VAL QQG 148 LYS H 3.26 #peak 471
147 VAL QQG 148 LYS QB 4.31 #peak 3652

147 VAL QQG 159 MET H 3.56 #peak 470
147 VAL QQG 160 GLY HA3 3.84 #peak 465
147 VAL QQG 161 LEU H 4.45 #peak 3361
147 VAL QQG 162 VAL QG1 3.90 #peak 2997
147 VAL QQG 162 VAL QG2 3.63 #peak 2337
147 VAL QQG 187 PHE QB 4.39 #peak 433
148 LYS H 148 LYS QB 3.68 #peak 819
148 LYS HA 148 LYS QG 3.71 #peak 4222
148 LYS HA 148 LYS QD 3.27 #peak 4223
148 LYS HA 148 LYS QE 3.98 #peak 4017
148 LYS HA 149 VAL QQG 3.72 #peak 4876
148 LYS HA 159 MET QB 5.34 #peak 2100
148 LYS QB 148 LYS QD 3.15 #peak 3649
148 LYS QB 148 LYS QE 3.42 #peak 1063
148 LYS QB 149 VAL H 4.22 #peak 626
148 LYS QB 159 MET H 4.51 #peak 1489
148 LYS QB 159 MET QB 4.02 #peak 3650
148 LYS QB 159 MET QG 4.76 #peak 3651
148 LYS QG 149 VAL H 4.78 #peak 340
148 LYS QD 149 VAL H 4.27 #peak 1115
148 LYS QE 159 MET H 5.34 #peak 939
148 LYS QE 159 MET QB 4.56 #peak 2394
149 VAL QQG 150 THR H 3.61 #peak 738
149 VAL QQG 150 THR HA 3.41 #peak 2338
149 VAL QQG 151 THR H 4.41 #peak 2327
149 VAL QQG 157 PHE H 5.44 #peak 741
150 THR H 156 VAL QQG 3.72 #peak 3401
150 THR HA 156 VAL QQG 4.19 #peak 4878
150 THR QG2 156 VAL QQG 4.04 #peak 2366
151 THR H 156 VAL QQG 3.94 #peak 636
151 THR HA 156 VAL QQG 3.91 #peak 4298
152 GLU H 156 VAL QQG 5.41 #peak 4219
153 ASN QB 155 GLU H 4.35 #peak 1242
154 GLY QA 181 LYS QG 3.72 #peak 4281
154 GLY QA 181 LYS QE 4.30 #peak 4889
156 VAL H 156 VAL QQG 3.41 #peak 486
156 VAL H 183 VAL QQG 4.51 #peak 3617
156 VAL HA 156 VAL QQG 3.05 #peak 2137
156 VAL QQG 157 PHE H 3.32 #peak 314
159 MET H 159 MET QG 4.79 #peak 1473
159 MET H 187 PHE QB 4.79 #peak 1528
159 MET HA 159 MET QG 3.50 #peak 5145
159 MET HA 187 PHE QB 4.96 #peak 747
159 MET QB 186 ALA QB 4.45 #peak 4905
159 MET QG 186 ALA QB 3.63 #peak 522
160 GLY HA2 161 LEU QB 5.34 #peak 4919
161 LEU H 161 LEU QB 3.38 #peak 656
161 LEU H 161 LEU QQD 4.44 #peak 3359
161 LEU HA 190 ILE QG1 4.38 #peak 2535
161 LEU QB 161 LEU QQD 2.69 #peak 3018
161 LEU QB 190 ILE HA 4.94 #peak 2407
161 LEU QB 190 ILE HB 4.99 #peak 2530
161 LEU QB 190 ILE QG1 3.63 #peak 5022
161 LEU QQD 162 VAL H 3.54 #peak 3366
161 LEU QQD 162 VAL HA 3.92 #peak 1409
161 LEU QQD 162 VAL QG1 4.60 #peak 1414
161 LEU QQD 162 VAL QG2 4.17 #peak 1415
161 LEU QQD 188 THR H 3.71 #peak 741
161 LEU QQD 188 THR HA 4.50 #peak 1408
161 LEU QQD 189 PHE H 3.86 #peak 1189
161 LEU QQD 189 PHE HA 3.98 #peak 2419
161 LEU QQD 189 PHE HD1 5.01 #peak 2416
161 LEU QQD 190 ILE HA 3.35 #peak 151
162 VAL H 190 ILE QG1 4.00 #peak 3369
162 VAL HA 166 GLU QB 4.46 #peak 4262
162 VAL HA 190 ILE QG1 4.04 #peak 189
162 VAL QG1 166 GLU QB 3.42 #peak 725
162 VAL QG2 166 GLU QB 4.12 #peak 1124
162 VAL QG2 187 PHE QB 3.32 #peak 822
163 THR H 166 GLU QB 3.81 #peak 2293
163 THR HA 164 GLU QG 4.75 #peak 2432

163 THR HA 190 ILE QG1 5.34 #peak 1167
163 THR HA 193 GLY QA 5.34 #peak 2547
163 THR HB 165 ARG QB 5.34 #peak 2222
163 THR QG2 166 GLU QG 4.13 #peak 5196
164 GLU H 164 GLU QG 3.53 #peak 665
164 GLU HA 164 GLU QG 3.71 #peak 3045
164 GLU QB 167 ALA H 5.34 #peak 1481
164 GLU QG 165 ARG H 4.01 #peak 1618
164 GLU QG 167 ALA H 5.34 #peak 2435
164 GLU QG 193 GLY QA 3.76 #peak 5202
165 ARG H 165 ARG QD 4.37 #peak 862
165 ARG HA 165 ARG QG 3.37 #peak 5149
165 ARG QB 167 ALA H 5.04 #peak 1789
166 GLU H 166 GLU QG 4.57 #peak 1360
166 GLU QB 167 ALA H 3.47 #peak 3402
166 GLU QB 167 ALA QB 4.63 #peak 5211
166 GLU QB 169 ALA QB 5.34 #peak 4265
166 GLU QG 167 ALA H 4.57 #peak 4956
167 ALA QB 187 PHE QB 4.07 #peak 2463
168 LYS H 168 LYS QG 3.61 #peak 682
168 LYS HA 168 LYS QG 3.50 #peak 5192
168 LYS QG 169 ALA H 3.72 #peak 1397
173 ILE HA 176 ARG QB 4.37 #peak 5228
174 ALA H 177 VAL QQG 4.41 #peak 2002
174 ALA H 183 VAL QQG 5.44 #peak 5006
175 SER H 183 VAL QQG 5.18 #peak 5005
176 ARG H 176 ARG QG 4.55 #peak 2606
176 ARG HA 176 ARG QG 3.51 #peak 178
176 ARG QB 176 ARG HD2 3.63 #peak 3711
176 ARG QB 176 ARG HD3 3.01 #peak 3702
176 ARG HD3 177 VAL QQG 4.57 #peak 3705
177 VAL H 177 VAL QQG 3.11 #peak 2000
177 VAL QQG 178 SER H 4.24 #peak 3246
181 LYS HA 181 LYS QG 3.50 #peak 4275
181 LYS HA 181 LYS QE 4.74 #peak 3240
181 LYS QB 181 LYS QE 4.43 #peak 4116
181 LYS QG 182 ARG H 5.10 #peak 3494
183 VAL QQG 184 THR H 3.49 #peak 512
183 VAL QQG 185 THR HA 4.54 #peak 3082
183 VAL QQG 185 THR HB 4.53 #peak 1025
187 PHE QB 188 THR H 4.13 #peak 740
189 PHE H 189 PHE QB 3.25 #peak 745
189 PHE HA 190 ILE QG1 4.90 #peak 2536
189 PHE QB 190 ILE H 3.62 #peak 755
190 ILE H 190 ILE QG1 4.17 #peak 1159
191 LYS QB 193 GLY H 4.52 #peak 3537
194 LEU H 194 LEU QQD 4.25 #peak 2301
194 LEU H 195 GLU QB 4.42 #peak 1304
194 LEU HA 194 LEU QQD 3.57 #peak 3143
194 LEU HA 195 GLU QB 4.34 #peak 2560
194 LEU HA 195 GLU QG 4.53 #peak 2569
194 LEU QB 195 GLU HA 4.99 #peak 5030
195 GLU HA 196 HIS QB 4.88 #peak 1433
195 GLU QG 196 HIS HA 4.24 #peak 2570
195 GLU QG 196 HIS QB 4.70 #peak 3896
196 HIS QB 197 HIS H 4.39 #peak 1141

APPENDIX IV
List of residues that gave a ^{15}N HSQC resonance signal after exchange in D_2O

I20
A28
V30
K33
T36
D37
R39
V41
Q44
V45
D47
T49
E51
V52
R53
V54
N55
S56
A57
L58
S59
D61
Q63
I64
K65
A68
R69
N71
V72
T73
A74
Y75
K78
V79
L80
L81
V82
G83
Q84
S85
E89
A92
R93
A94
K95
Q96

I97
A98
M99
G100
V101
A104
E106
Y108
E110
I111
R112
Q113
L119
A122
N124
D125
I128
T129
T130
K131
V132
R133
S134
Q135
L136
L137
L141
S144
N146
K148
V149
T150
G154
E155
V156
F157
L158
M159
T163
R165
E166
A167
K168
A169
A170
A171
D172
I173
A174
S175
R176
V177

V180
T184
F189
K191
L194
E195
H196
H197

Appendix V List of Hydrogen bonds used to determine the structure of YraP

```
!1
assign ( residue   51 and name HN  ) ( residue   47  and name O )  1.80  0.00
0.50
assign ( residue   51 and name N  )  ( residue   47  and name O )  2.80  0.00
0.50

!3
assign ( residue   53 and name HN  ) ( residue   49  and name O )  1.80  0.00
0.50
assign ( residue   53 and name N  )  ( residue   49  and name O )  2.80  0.00
0.50

!5
assign ( residue   55 and name HN  ) ( residue   51  and name O )  1.80  0.00
0.50
assign ( residue   55 and name N  )  ( residue   51  and name O )  2.80  0.00
0.50

!8
assign ( residue   57 and name HN  ) ( residue   53  and name O )  1.80  0.00
0.50
assign ( residue   57 and name N  )  ( residue   53  and name O )  2.80  0.00
0.50

!2
assign ( residue   65 and name HN  ) ( residue   61  and name O )  1.80  0.00
0.50
assign ( residue   65 and name N  )  ( residue   61  and name O )  2.80  0.00
0.50

!10
assign ( residue   68 and name HN  ) ( residue   64  and name O )  1.80  0.00
0.50
assign ( residue   68 and name N  )  ( residue   64  and name O )  2.80  0.00
0.50

!11
assign ( residue   84 and name HN  ) ( residue   69  and name O )  1.80  0.00
0.50
assign ( residue   84 and name N  )  ( residue   69  and name O )  2.80  0.00
0.50

!12
assign ( residue   82 and name HN  ) ( residue   71  and name O )  1.80  0.00
0.50
assign ( residue   82 and name N  )  ( residue   71  and name O )  2.80  0.00
0.50

!13
assign ( residue   80 and name HN  ) ( residue   73  and name O )  1.80  0.00
0.50
assign ( residue   80 and name N  )  ( residue   73  and name O )  2.80  0.00
0.50

!14
assign ( residue   78 and name HN  ) ( residue   75  and name O )  1.80  0.00
0.50
assign ( residue   78 and name N  )  ( residue   75  and name O )  2.80  0.00
0.50

!15
assign ( residue   75 and name HN  ) ( residue   78  and name O )  1.80  0.00
0.50
```

```
assign ( residue   75 and name N  )  ( residue   78  and name O )  2.80  0.00
0.50

!90
assign ( residue   73 and name HN  ) ( residue   80  and name O )  1.80  0.00
0.50
assign ( residue   73 and name N  )  ( residue   80  and name O )  2.80  0.00
0.50

!16
assign ( residue   71 and name HN  ) ( residue   82  and name O )  1.80  0.00
0.50
assign ( residue   71 and name N  )  ( residue   82  and name O )  2.80  0.00
0.50

!18
assign ( residue   93 and name HN  ) ( residue   89  and name O )  1.80  0.00
0.50
assign ( residue   93 and name N  )  ( residue   89  and name O )  2.80  0.00
0.50

!46
assign ( residue   96 and name HN  ) ( residue   92  and name O )  1.80  0.00
0.50
assign ( residue   96 and name N  )  ( residue   92  and name O )  2.80  0.00
0.50

!20
assign ( residue   97 and name HN  ) ( residue   93  and name O )  1.80  0.00
0.50
assign ( residue   97 and name N  )  ( residue   93  and name O )  2.80  0.00
0.50

!99
assign ( residue   98 and name HN  ) ( residue   94  and name O )  1.80  0.00
0.50
assign ( residue   98 and name N  )  ( residue   94  and name O )  2.80  0.00
0.50

!21
assign ( residue  100 and name HN  ) ( residue   96  and name O )  1.80  0.00
0.50
assign ( residue  100 and name N  )  ( residue   96  and name O )  2.80  0.00
0.50

!22
assign ( residue  128 and name HN  ) ( residue  124  and name O )  1.80  0.00
0.50
assign ( residue  128 and name N  )  ( residue  124  and name O )  2.80  0.00
0.50

!23
assign ( residue  129 and name HN  ) ( residue  125  and name O )  1.80  0.00
0.50
assign ( residue  129 and name N  )  ( residue  125  and name O )  2.80  0.00
0.50

!25
assign ( residue  132 and name HN  ) ( residue  128  and name O )  1.80  0.00
0.50
assign ( residue  132 and name N  )  ( residue  128  and name O )  2.80  0.00
0.50

!26
assign ( residue  133 and name HN  ) ( residue  129  and name O )  1.80  0.00
0.50
assign ( residue  133 and name N  )  ( residue  129  and name O )  2.80  0.00
0.50
```

```
!27
assign ( residue  134 and name HN  ) ( residue  130  and name O )  1.80  0.00
0.50
assign ( residue  134 and name N  )  ( residue  130  and name O )  2.80  0.00
0.50

!30
assign ( residue  137 and name HN  ) ( residue  133  and name O )  1.80  0.00
0.50
assign ( residue  137 and name N  )  ( residue  133  and name O )  2.80  0.00
0.50

!31
assign ( residue  159 and name HN  ) ( residue  148  and name O )  1.80  0.00
0.50
assign ( residue  159 and name N  )  ( residue  148  and name O )  2.80  0.00
0.50

!32
assign ( residue  157 and name HN  ) ( residue  150  and name O )  1.80  0.00
0.50
assign ( residue  157 and name N  )  ( residue  150  and name O )  2.80  0.00
0.50

!33
assign ( residue  184 and name HN  ) ( residue  156  and name O )  1.80  0.00
0.50
assign ( residue  184 and name N  )  ( residue  156  and name O )  2.80  0.00
0.50

!34
assign ( residue  150 and name HN  ) ( residue  157  and name O )  1.80  0.00
0.50
assign ( residue  150 and name N  )  ( residue  157  and name O )  2.80  0.00
0.50

!35
assign ( residue  148 and name HN  ) ( residue  159  and name O )  1.80  0.00
0.50
assign ( residue  148 and name N  )  ( residue  159  and name O )  2.80  0.00
0.50

!36
assign ( residue  158 and name HN  ) ( residue  184  and name O )  1.80  0.00
0.50
assign ( residue  158 and name N  )  ( residue  184  and name O )  2.80  0.00
0.50

!37
assign ( residue  169 and name HN  ) ( residue  165  and name O )  1.80  0.00
0.50
assign ( residue  169 and name N  )  ( residue  165  and name O )  2.80  0.00
0.50

!38
assign ( residue  170 and name HN  ) ( residue  166  and name O )  1.80  0.00
0.50
assign ( residue  170 and name N  )  ( residue  166  and name O )  2.80  0.00
0.50

!39
assign ( residue  172 and name HN  ) ( residue  168  and name O )  1.80  0.00
0.50
assign ( residue  172 and name N  )  ( residue  168  and name O )  2.80  0.00
0.50

!40
```

```
assign ( residue  173 and name HN  ) ( residue  169  and name O )  1.80  0.00
0.50
assign ( residue  173 and name N  )  ( residue  169  and name O )  2.80  0.00
0.50

!41
assign ( residue  110 and name HN  ) ( residue  81  and name O )  1.80  0.00
0.50
assign ( residue  110 and name N  )  ( residue  81  and name O )  2.80  0.00
0.50

!42
assign ( residue  108 and name HN  ) ( residue  79  and name O )  1.80  0.00
0.50
assign ( residue  108 and name N  )  ( residue  79  and name O )  2.80  0.00
0.50

!43
assign ( residue  112 and name HN  ) ( residue  83  and name O )  1.80  0.00
0.50
assign ( residue  112 and name N  )  ( residue  83  and name O )  2.80  0.00
0.50

!44
assign ( residue  184 and name HN  ) ( residue  156  and name O )  1.80  0.00
0.50
assign ( residue  184 and name N  )  ( residue  156  and name O )  2.80  0.00
0.50

!45
assign ( residue  182 and name HN  ) ( residue  154  and name O )  1.80  0.00
0.50
assign ( residue  182 and name N  )  ( residue  154  and name O )  2.80  0.00
0.50

!46
assign ( residue  186 and name HN  ) ( residue  158  and name O )  1.80  0.00
0.50
assign ( residue  186 and name N  )  ( residue  158  and name O )  2.80  0.00
0.50
```