Structural Studies of the Inner Membrane Ring of the Bacterial Type III Secretion System

Melanie A. McDowell

Department of Biochemistry

Sir William Dunn School of Pathology

Brasenose College

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity Term 2012

Structural Studies of the Inner Membrane Ring of the Bacterial Type III Secretion System

Abstract

Shigella flexneri attacks cells of the intestinal tract, causing over 1 million deaths annually from bacterial dysentery. A type III secretion system (T3SS) initiates the host-pathogen interaction and transports virulence factors directly into host cells via a needle complex (NC) comprising an extracellular needle and membrane-spanning basal body. Rings formed by the single-pass membrane proteins MxiG and MxiJ are arranged concentrically within the inner membrane ring (IMR) of the NC.

The N-terminal domain of MxiG (MxiG-N) is the predominant IMR cytoplasmic structure, however it was structurally and functionally uncharacterised. Determination of the solution structure of MxiG-N in this study revealed it to be a forkhead associated (FHA) domain, although subsequent analyses of conserved residues suggested it does not have the canonical role in cell-signalling via phospho-threonine recognition. Subsequent positioning of the structure in the electron microscopy (EM) density for the *S. flexneri* NC supported models with 24-fold symmetry in the IMR.

Both MxiG and MxiJ also have significant periplasmic domains, which were purified to homogeneity in this study, facilitating preliminary characterisation of their structures and intermolecular interactions. In addition, the entire IMR within the context of intact basal bodies was isolated and visualised *in vitro* by EM.

The essential function of MxiG-N could be to localise the putative cytoplasmic ring (C-ring) at the base of the T3SS. Although absolutely required for secretion, the C-ring component, Spa33, was structurally uncharacterised. The crystal structure of the C-terminal domain of Spa33 (Spa33-C) was determined in this study, showing an intertwined dimer that aligned with homologous structures and exhibited a novel interaction with the N-terminus of the ATPase regulator, MxiN. Subsequently, Spa33-C was identified as an alternative translation product of *spa33* that formed a 2:1 complex with Spa33 *in vitro*. This complex oligomerised further, demonstrating for the first time that Spa33 has the propensity to form the ordered, high molecular weight assemblies that would be required for C-ring formation in *S. flexneri*.

Acknowledgements

Firstly, I want to thank Susan Lea for being a fantastic role model and mentor during my time as her undergraduate and graduate student. I very much appreciate her unwavering support and belief in me throughout the highs and wobbles of my DPhil and her enthusiasm for my work has made these four years enjoyable and fulfilling - as a result, I am excited about a career in scientific research!

I would also like to thank all past and present members of the Lea Lab for providing a special working environment that is both productive and fun. Specifically, Steven Johnson has acted as a valuable sounding-board throughout my DPhil, providing key advice and insights for my project that the citations within this thesis can't do justice to. Patrizia Abrusci's lab experience has also been invaluable and was crucial for experimental troubleshooting on so many occasions. Pietro Roversi patiently helped me get to grips with X-ray crystallography, whilst Devon Sheppard and Janet Deane have provided great support in NMR spectroscopy. In addition, the Part II Projects from Harriet Allison and Miriam Friede provided a firm basis for my work on MxiJ and Spa33-C respectively.

Within the lab, I have also very much enjoyed and benefitted from the company of Phil Ward, Matthijs Jore, Joe Caesar, Matt Cottee, Paul Chappell, James Lillington and Kirstin Leath on a daily basis (along with the various Radio 2 presenters…). I want to say a special thank you to my Leigh girls, Sarah Rollauer and Rachel Everett: without your advice, support and friendship these past four years would have been much harder and a lot less fun!

I have also been fortunate to have had access to ample scientific advice and expertise from collaborators outside of the lab. In particular, I want to express my gratitude to Prof. Jim McDonnell for an excellent introduction to NMR spectroscopy and his co-supervision during the MxiG-N structure determination. Similarly, Dr Julien Marcoux provided native mass spectrometry data that has been invaluable for the characterisation of Spa33. In addition, Dr Ariel Blocker and Martin Cheung kindly assisted me with the MxiG-N *in vivo* assays in their

lab at the University of Bristol. I want to also thank the Wellcome Trust for generous financial support and guidance throughout my DPhil.

I would also like to thank my friends both within and outside of Oxford for their moral support during my DPhil. In particular, the past and present inhabitants of the 'Loveshack' (Caitlin, Sarah, Lydia, Jen, SJ, Eachan) have been so much more than just great housemates and I feel privileged to have shared my whole post-16 education with Ellen. Visiting the Marburg girls (Amanda, Jess, Lillian) and the Brasenose crowd (Matt, Karen, Jess, Sophie, Beth) has also provided valuable escapes from the bubble.

Finally, I want to thank my wonderful family for their limitless love, pride and support in all the years leading up to and during my DPhil, and for shaping the person that made it this far. Mum, Dad and Rose: no matter where the future takes me, you'll always be my home.

Publications resulting from this study

Structural and Functional Studies on the N-terminal Domain of the Shigella Type III Secretion Protein MxiG

McDowell, M. A., S. Johnson, J. E. Deane, M. Cheung, A. D. Roehrich, A. J. Blocker, J. M. McDonnell & S. M. Lea, (2011) Journal of Biological Chemistry **286:** 30606-30614.

The Salmonella Type III Secretion System Inner Rod Protein PrgJ is Partially Folded

Zhong, D., M. Lefebre, K. Kaur, McDowell, M. A., C. Gdowski, S. Jo, Y. Wang, S. H. Benedict, S. M. Lea, J. E. Galan & R. N. De Guzman, (2012) Journal of Biological Chemistry **287:** 25303-25311.

Abbreviations

CCP4	Collaborative Computational Project Number 4
CD	Circular Dichroism
CLEANEX-PM	Phase-Modulated CLEAN Chemical Exchange
CMC	Critical Micelle Concentration
CPMG	Carr-Purcell-Meiboom-Gill
CR	Congo Red
C-ring	Cytoplasmic ring
CSI	Chemical Shift Index
CV	Column Volume
DDM	n-dodecyl-β-D-maltopyranoside
DMA	Dimethyl Adipimidate
DNase	Deoxyribonuclease
DSP	Dithio-bis(Succinimidyl Propionate)
DTBP	Dithiobispropionimidate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGS	Ethylene Glycol bis(Succinimidylsuccinate)
EM	Electron Microscopy
EPEC	Enteropathogenic *Escherichia coli*
FFAS	Fold and Function Assignment Server
FHA	Forkhead Associated
GST	Glutathione S-Transferase
HSQC	Heteronuclear Single Quantum Coherance
IMR	Inner Membrane Ring
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani
LDAO	n-dodecyl-N,N-dimethylamine-N-oxide
LLG	Log-Likelihood Gain
MALS	Multiangle Light Scattering
MCS	Multiple Cloning Site
MR	Molecular Replacement
MxiG-C	C-terminal domain of MxiG
MxiG-N	N-terminal domain of MxiG
NC	Needle Complex

NCS	Non Crystallographic Symmetry
NMR	Nuclear Magentic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OD_{600}	Optical Density at 600 nm
OMR	Outer Membrane Ring
PA	Polyacrylamide
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
P-Thr	Phospho-Threonine
R1	Longitudinal Relaxation
R2	Transverse Relaxation
RBS	Ribosome Binding Site
RMSD	Root Mean Squared Deviation
S^2	Order Parameter
SDS	Sodium Dodecyl Sulphate
SEC	Size Exclusion Chromatography
Spa33-C	C-terminal domain of Spa33
SPI	*Salmonella* Pathogenicity Island
SPR	Surface Plasmon Resonance
SRF	Self-Rotation Function
SSM	Secondary Structure Matching
STPK	Serine/Threonine Protein Kinase
T_a	Annealing Temperature
τ_c	Correlation Time
TCSB	Tryptic Casein Soy Broth
t_e	Elongation Time
TEM	Transmission Electron Microscopy
ThermoFluor	Fluorescence-based thermal shift assay
TMAO	Trimethylamine-N-oxide
TOCSY	Total Correlation Spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
TROSY	Transverse Relaxation Optimised Spectroscopy
T3SS	Type III Secretion System

Table of Contents

1. Introduction 1

1.1 Bacterial Secretion Systems 1

1.2 The Translocation-Associated Type III Secretion System (T3SS) 5

Classification of Translocation-Associated T3SSs 8

Structural Approaches to Studying the Translocation-Associated T3SS 9

A Working Model for Hierarchical Assembly and Secretion by the T3SS 13

1.3 The *Shigella* T3SS 22

The Role of the *Shigella* T3SS in Pathogenesis 22

The Inner Membrane Ring of the *Shigella* T3SS 24

The C-ring of the *Shigella* T3SS 30

1.4 Thesis Outline 35

2. Structural characterisation of MxiG-N 36

2.1 Expression and Purification of MxiG-N 36

Expression of MxiG-N_{1-126} 36

Purification of MxiG-N_{1-126} 37

Limited Proteolysis of MxiG-N_{1-126} 41

Expression and Purification of $^{13}C/^{15}N$-MxiG-N_{6-112} 42

2.2 Assignment of MxiG-N_{6-112} 44

Backbone Assignment by the Triple Resonance Approach 46

Backbone Assignment by the Sequential Approach 50

Assignment of Side-Chain Protons 52

2.3 Secondary Structure Prediction for MxiG-N_{6-112} 53

Predictions from H^{α} Secondary Chemical Shift and $^{3}J_{HNH\alpha}$ Values 53

Measurement of Backbone Amide Proton Exchange 55

Summary of Secondary Structure Prediction for MxiG-N_{6-112} 57

2.4 Determination of Structural Restraints for MxiG-N_{6-112} 58

Identification of Secondary Structure via Medium and Long-Range NOEs 58

Direct Observation of Hydrogen Bonds 61

Identification of Long-Range NOEs in the $^{1}H,^{13}C$ NOESY-HSQC spectrum 63

Conversion of Experimental Parameters to Distance Restraints 64

Determination of Dihedral Angle Restraints for MxiG-N_{6-112} 65

2.5 Determination of the NMR Structure of MxiG-N_{6-112} 65

2.6 Backbone Dynamics of MxiG-N_{6-112} 71

2.7 Structural Analysis of MxiG-N_{6-112} 73
MxiG-N_{6-112} – A New Model System for NMR? 73
Comparison with the Crystal Structures of MxiG-N_{1-126} and YscD-N_{2-108} 74
Positioning of MxiG-N_{6-112} within the T3SS 76
2.8 Perspectives 79
3. Functional characterisation of MxiG-N 80
3.1 MxiG-N is an Essential Component of the T3SS 80
3.2 Does MxiG-N Function as a Canonical FHA Domain? 83
MxiG-N_{6-112} Has the Fold of a FHA Domain 83
MxiG-N_{6-112} Does Not Interact with P-Thr *in vitro* 86
Conserved P-Thr Binding Residues Are Not Required *in vivo* 87
MxiG-N Has a Non-Canonical Function for a FHA Domain 89
3.3 Investigation of Alternative Interaction Partners for MxiG-N 91
Can MxiG-N Pull Down *in vivo* Interaction Partners? 91
MxiG-N_{6-112} Does Not Interact With Phosphorylated Spa33 99
3.4 Does MxiG-N Structurally Stabilise the IMR? 104
MxiG-N_{6-112} Does Not Interact with $MxiJ_{232-241}$ 104
A 24 Subunit Ring Suggests Close Association of MxiG-N 106
MxiG-N is Monomeric in Solution 108
Low Affinity Interactions Could Still Be Important For Ring Assembly or Stabilisation 110
3.5 Perspectives 112
4. Purification and Characterisation of the MxiG and MxiJ Periplasmic Domains 113
4.1 Purification and Characterisation of MxiG-C 114
Purification of MxiG-$C_{142-371}$ from the Soluble Fraction 116
Purification of MxiG-$C_{142-371}$ from the Insoluble Fraction 124
Limited Proteolysis of MxiG-$C_{142-371}$ 129
Characterisation of MxiG-C 132
Crystallisation Trials With MxiG-$C_{161-362/4}$ 135
4.2 Purification and Characterisation of MxiJ Periplasmic Domains 139
4.3 Probing the Putative Interaction Between the Periplasmic Domains of MxiG and MxiJ 146
4.4 Perspectives 150
5. Purification and Characterisation of the T3SS Basal Body 151
5.1 Purification of the T3SS Basal Body 151
5.2 Optimisation of the Detergents used for T3SS Basal Body Purification 161
5.3 Purification of Complexes Lacking MxiG-N 165
5.4 Perspectives 169

6. Structural Characterisation of Spa33-C 170
6.1 Purification of $Spa33_{208-293}$ 171
6.2 Crystallisation Trials with $Spa33_{208-293}$ 174
6.3 Determination of the Crystal Structure of $Spa33_{208-293}$ 178
Data Collection 178
Data Processing 179
Characterisation of the Asymmetric Unit 183
Phasing by Molecular Replacement 185
Model Building and Structure Refinement 189
6.4 Structural Analysis of $Spa33_{208-293}$ 194
The Structure of $Spa33_{208-293}$ 194
Comparison with Homologous Structures 198
Improvement of the Crystallisation Efficiency 203
6.5 $Spa33_{220-293}$-His Interacts with the Extreme N-terminus of MxiN *in vitro* 205
6.6 Perspectives 211
7. Characterisation of the *Shigella* T3SS C-ring 212
7.1 $Spa33_{192-293}$ is an Alternative Translation Product of Spa33 214
Purification of Spa33 214
Identification of Contaminating Fragments 216
Optimisation of the Purification Protocol 220
7.2 Spa33-$(Spa33_{192-293})_2$ is the Building Block of the C-Ring 222
A His-Spa33-$(Spa33_{192-293})_2$ Complex Undergoes Further Oligomerisation 222
The Interactions and Alternative Conformations of Spa33-C 225
Oligomerisation of $(Spa33_{220-293})_2$ 233
7.3 Towards Structural Characterisation of the *Shigella* T3SS C-ring 239
Optimisation of His-Spa33-$(Spa33_{192-293})_2$ Complex Purification for Crystallisation Trials 239
Does Spa33-$(Spa33_{192-293})_2$ Form a C-ring *in vitro*? 243
The C-ring Within the Context of the *Shigella* T3SS 245
7.4 Future Perspectives 246
References 248
8. Appendix 248
8.1 Protein Sequences and Construct Directory 268
MxiG 268
MxiJ 269
Spa33 270

MxiN 271
8.2 Standard Laboratory Methods 272
8.2.1 Transformation Protocol 272
8.2.2 SDS-PAGE 272
8.2.3 Site-Directed Mutagenesis 274
8.2.4 PCR For Insert Preparation 275
8.2.5 Agarose Gel Electrophoresis 276
8.2.6 ThermoFluor Assay 277
8.2.7 Western Blotting 278
8.3 NMR experiments 279
8.4 Additional Work 284
8.4.1 Cloning and Expression of the MxiD Periplasmic Domain 284
8.4.2 Characterisation of MxiIΔ5 285
8.4.3 SPR Analysis of the DNA Binding Properties of *N. meningitidis* ComP 287

Index of Tables and Figures

1. Introduction

Figure 1.1 *The diversity of bacterial secretion systems*. 2
Figure 1.2 *Comparison of translocation-associated and flagellar T3SSs*..... 6
Table 1.1 The components of translocation-associated and flagellar T3SSs..... 7
Table 1.2 The eight families of translocation-associated T3SSs……...8
Table 1.3 High resolution structures of isolated translocation-associated T3SS proteins and domains available before this study……10
Figure 1.3 *The three approaches to structural characterisation of the translocation-associated T3SS*. 11
Table 1.4 High resolution structures of sub-complexes of the translocation-associated T3SS....13
Figure 1.4 *The hierarchy of T3SS assembly and secretion*. 14
Figure 1.5 *The pathogenesis of Shigella*. 23
Table 1.5 The early effectors secreted by the *Shigella* T3SS……..24
Figure 1.6 *The topology of MxiG and MxiJ with respect to the inner membrane*..... 26
Figure 1.7 *The position of MxiG-N within the S. flexneri NC*.. 26
Figure 1.8 *The periplasmic region of the T3SS NC*..... 29
Figure 1.9 *EM characterisation of the C-rings from translocation-associated and flagellar T3SSs*..... 31
Table 1.6 The interaction network of Spa33 homologues……...32

2. Structural Characterisation of MxiG-N

Figure 2.1 *15% SDS-PA gel showing over-expression of MxiG-N_{1-126} in BL21 (DE3) pLysS*. ... 37
Figure 2.2 *Purification of MxiG-N_{1-126}*. 39
Figure 2.3 *Natural abundance $^{1}H,^{15}N$-HSQC spectrum of 3 mM MxiG-N_{1-126}*. 40
Figure 2.4 *15% SDS-PA gel showing limited proteolysis and purification of MxiG-N_{1-125}*. 41
Figure 2.5 *Tryptic digest sites enclosing MxiG-N_{6-112}*……..42
Table 2.1 Composition of minimal media……....43
Figure 2.6 *$^{1}H,^{15}N$-HSQC spectra for $^{13}C/^{15}N$-MxiG-N_{6-112}* 45
Figure 2.7 *Strips of the HN(CO)CA, CBCA(CO)NH, CBCANH and CC(CO)NH spectra corresponding to an isoleucine residue from the $^{1}H,^{15}N$-HSQC*. 46
Figure 2.8 *Strips from the CBCANH spectrum with sequence-specific assignment of residues I35 to L44 by the triple resonance approach*. 48
Figure 2.9 *Assignment of the $^{1}H,^{15}N$-HSQC spectrum of MxiG-N_{6-112}*. 49
Figure 2.10 *Strips from the $^{1}H,^{15}N$ NOESY-HSQC spectrum with sequence-specific assignment of residues I35 to L44 by the sequential approach*. 51
Figure 2.11 *Strips of the HNHA, HBHA(CO)NH, H(CCCO)NH and HCCH-TOCSY spectra corresponding to the side-chain of E30*. 52
Figure 2.12 *Secondary structure prediction for MxiG-N_{6-112} from H^{α} secondary chemical shift and $^{3}J_{HNH\alpha}$ values*..... 54
Figure 2.13 *Measurement of backbone amide proton exchange*..... 56
Figure 2.14 *Summary of secondary structure data for MxiG-N_{6-112}*..... 57

Figure 2.15 *Strips from the $^{1}H,^{15}N$ NOESY-HSQC spectrum identifying NOEs characteristic of secondary structure elements*. 59
Figure 2.16 *MxiG-N_{6-112} β-sheets identified from long-range NOEs and hydrogen bond couplings*. 60
Figure 2.17 *Selected region of the 2D long-range H(N)CO spectrum*. 62
Figure 2.18 *Strips from the $^{1}H,^{13}C$ NOESY-HSQC spectrum showing long-range NOEs between residues E97 and S101.* 63
Table 2.2 Models from iterative round of MxiG-N_{6-112} refinement……67
Figure 2.19 *The NMR structure of MxiG-N_{6-112}*..... 68
Table 2.3 Summary of restraints and structural statistics for MxiG-N_{6-112}……..69
Figure 2.20 *Solvent-exposed and slowly-exchanging amides mapped onto the MxiG-N_{6-112} structure*. 70
Figure 2.21 *The backbone dynamics of MxiG-N_{6-112}*. 72
Figure 2.22 *Comparison of the NMR structure of MxiG-N_{6-112} with high resolution structures of homologous T3SS proteins*. 75
Figure 2.23 *Positioning of MxiG-N_{6-112} in the C24 S. flexneri NC EM map*. 77
Figure 2.24 *Regions of the S. flexneri NC EM maps corresponding to portions of the MxiG-N ring*. 79

3. Functional Characterisation of MxiG-N

Table 3.1 Primers used to introduce the MxiGΔN mutation……...80
Table 3.2 *S. flexneri* strains used for overnight leakage and CR induction assays……..81
Figure 3.1 *Effect of MxiG-N knockout on T3SS assembly in S. flexneri*..... 82
Figure 3.2 *Comparison of the structure of MxiG-N_{6-112} with its closest structural homologue CT664 from C. trachomatis*..... 83
Figure 3.3 *Analysis of conserved FHA domain residues in MxiG-N_{6-112}*..... 85
Figure 3.4 *Titration of ^{15}N-MxiG-N_{6-112} with 50 mM P-Thr*. 86
Table 3.3 Primers used to create the R39A/S61A/S63A mutation……..87
Figure 3.5 *Effect of mutation of conserved FHA domain residues on T3SS assembly and function in S. flexneri*..... 88
Figure 3.6 *1% agarose gels of DNA fragments from pBAD-mxiGN_{2-125} and pGEX-mxiGN_{2-125} cloning*. 92
Figure 3.7 *15% SDS-PA gels showing overexpression of MxiG-N_{2-125} constructs in mxiG^{-}*. 95
Figure 3.8 *15% SDS-PA gels showing the pull-down assays with MxiG-N_{2-125}*. 96
Figure 3.9 *Comparison of SDS-PA gel profiles of purified MxiG-N extracted from S. flexneri cells that were subjected to in vivo crosslinking and from E. coli cells in the absence of reducing agent.* 100
Figure 3.10 *1D ^{1}H spectrum showing titration of ^{15}N-MxiG-N_{6-112} with P_5*. 101
Figure 3.11 *Titration of ^{15}N-MxiG-N_{6-112} with a saturated solution of P_5*. 102
Figure 3.12 *Titration of ^{15}N-MxiG-N_{6-112} with P_5 dissolved in DMSO*. 103
Figure 3.13 *Titration of ^{15}N-MxiG-N_{6-112} with 1 mM MxiJ$_{232-241}$*. 105
Figure 3.14 *The electrostatic surface of the NMR structure of MxiG-N_{6-112}*. 107
Figure 3.15 *MxiG-N_{1-126} dimer formed by crystal contacts*. 107
Figure 3.16 *MxiG-N_{6-112} tumbles as a monomer in solution*..... 108
Figure 3.17 *Analysis of the oligomeric state of MxiG-N_{1-126} by SEC and in-line multi-angle light scattering.* 109

4. Purification and Characterisation of the MxiG and MxiJ Periplasmic Domains

Figure 4.1 *Domain organisation of proteins contributing to the periplasmic region of the S. flexneri basal body.* 113
Figure 4.2 *Alignment of MxiG-C with PrgH$_{170-362}$.* 115
Figure 4.3 *15% SDS-PA gel showing overexpression of MxiG-C$_{142-371}$ in Origami 2 (DE3) cells.* 116
Figure 4.4 *Purification of MxiG-C$_{142-371}$ from Origami 2 (DE3) cells.* 118
Table 4.1 Primers used to amplify the *mxiGC$_{142-371}$* insert……....119
Figure 4.5 *1% agarose gel of mxiGC$_{142-371}$ insert amplified by touch-down PCR.* 119
Table 4.2 Primers used to create the C296S mutation in pET22b-*mxiGC$_{142-371}$*……...120
Table 4.3 *E. coli* cell-lines used for expression of MxiG-C$_{142-371}$(C296S)…….121
Figure 4.6 *13.5% SDS-PA reducing gel of periplasmic fractions from cells expressing pET22b-mxiGC$_{142-371}$(C296S)* 122
Figure 4.7 *Purification from pET22b-mxiGC$_{142-371}$(C296S).* 123
Table 4.4 Protocol for inclusion body preparation…….124
Figure 4.8 *Preparation of inclusion bodies from B834 (DE3) pLysS pET14b-mxiGC$_{142-371}$.* ... 125
Table 4.5 Refold buffers tested for MxiG-C$_{142-371}$…….125
Figure 4.9 *Refold trials for MxiG-C$_{142-371}$.* 126
Figure 4.10 *Optimisation of the SEC for MxiG-C$_{142-371}$ purification.* 128
Figure 4.11 *Limited proteolysis of MxiG-C$_{142-371}$.* 129
Figure 4.12 *Purification of MxiG-C$_{161-362/4}$.* 130
Figure 4.13 *CD spectroscopy of MxiG-C$_{142-371}$.* 133
Figure 4.14 *^{1}H,^{15}N-HSQC spectrum of 240 µM ^{15}N-MxiG-C$_{161-362/4}$.* 134
Figure 4.15 *Analysis of the oligomeric state of MxiG-C constructs by SEC and in-line multi-angle light scattering.* 134
Table 4.6 HT-96 broad screens from Molecular Dimensions used in crystallisation trials…...136
Table 4.7 Crystallisation trials with MxiG-C$_{161-362/4}$……..137
Figure 4.16 *Structural model for MxiJ$_{20-189}$.* 139
Figure 4.17 *The purification of ^{15}N-MxiJ$_{19-206}$.* 140
Figure 4.18 *^{1}H,^{15}N-HSQC spectrum of 150 µM ^{15}N-MxiJ$_{19-206}$.* 141
Figure 4.19 *Limited proteolysis of MxiJ$_{19-206}$.* 142
Figure 4.20 *Purification of MxiJ$_{19-100}$ and MxiJ$_{106-204}$.* 143
Figure 4.21 *^{1}H,^{15}N-HSQC spectrum of 200 µM ^{15}N-MxiJ$_{19-100}$.* 145
Figure 4.22 *NMR titrations with MxiJ$_{19-206}$ and MxiG-C$_{161-362/4}$.* 148

5. Purification and Characterisation of the T3SS Basal Body

Table 5.1 Growth medium used for *mxiG^{-}/mxiH^{-}* pBAD-*mxiG*…….153
Table 5.2 Protocol for the purification of basal bodies from *mxiG^{-}/mxiH^{-}* pBAD-*mxiG*……...154
Figure 5.1 *Purification of His-tagged basal bodies mxiG^{-}/mxiH^{-} pBAD-mxiG.* 156
Figure 5.2 *Elution profile for empty Triton X-100 micelles.* 157
Figure 5.3 *TEM images of the S. flexneri basal body.* 159
Figure 5.4 *Analysis of the mass of His-tagged basal bodies by SEC and in-line multi-angle light scattering.* 160
Figure 5.5 *Purification of His-tagged basal bodies with LDAO.* 162
Figure 5.6 *Purification of His-tagged basal bodies with DDM.* 163

Figure 5.7 *Purification of His-tagged basal bodies with Cymal-4*.....164
Figure 5.8 *Purification of complexes containing His-MxiGΔN*.....167

6. Structural Characterisation of Spa33-C

Figure 6.1 *Sequence alignments of Spa33-C with the C-termini of homologous T3SS proteins*170
Figure 6.2 *15% SDS-PA gel showing over-expression of Spa33$_{208-293}$-His in B834 (DE3) pLysS.* 171
Figure 6.2 *Purification of Spa33$_{208-293}$*.....173
Table 6.1 Optimisation Screen 1……175
Table 6.2 Optimisation Screen 2……175
Table 6.3 Crystallisation trials with Spa33$_{208-293}$……176
Figure 6.3 *Crystals of Spa33$_{208-293}$*.....177
Figure 6.4 *Diffraction patterns obtained for Spa33$_{208-293}$ crystals*.....178
Table 6.4 Data quality statistics used during processing……...180
Table 6.5 Statistics from processing of Spa33$_{208-293}$ datasets……....181
Figure 6.5 *Processing of Spa33$_{208-293}$ datasets.* 182
Table 6.6 Estimation of the number of copies of spa33$_{208-293}$ in the asymmetric unit……....183
Figure 6.6 *Spa33$_{208-293}$ self rotation function.*.....184
Figure 6.7 *Search models used during the molecular replacement of Spa33$_{208-293}$.* 187
Table 6.7 Molecular replacement solutions for Spa33$_{208-293}$……..187
Figure 6.8 *Initial electron density maps for Spa33$_{208-293}$.*.....188
Figure 6.9 *The use of difference electron density maps to modifiy the Spa33$_{208-293}$ model.* 190
Figure 6.10 *Structure refinement of Spa33$_{208-293}$.* 192
Table 6.8 Refinement statistics for the final Spa33$_{208-293}$ model……...193
Figure 6.11 *The final model and electron density difference map for Spa33$_{208-293}$.* 194
Figure 6.12 *The crystal structure of Spa33$_{208-293}$.*.....195
Figure 6.13 *Structural differences between the Spa33$_{208-293}$ monomers*.....196
Figure 6.14 *The arrangement of the Spa33$_{208-293}$ crystal lattice*.....197
Figure 6.15 *Structural overlays of the Spa33$_{208-293}$ dimer with homologous proteins*.....198
Figure 6.16 *Identification of functionally important residues within the Spa33$_{208-293}$ dimer.* ...200
Figure 6.17 *The electrostatic surfaces of Spa33$_{208-293}$ and homologues*.....202
Table 6.9 Primers used to create the pET28b-*spa33$_{220-293}$* construct……203
Figure 6.18 *Crystals of Spa33$_{220-293}$*.....204
Table 6.10 Comparison of unit cell dimensions of Spa33$_{208-293}$ and Spa33$_{220-203}$ crystals……..205
Figure 6.19 *Alignment of S. flexneri MxiN with S. typhimurium FliH*……...206
Figure 6.20 *^{1}H,^{15}N-HSQC spectra of ^{15}N-labelled Spa33-C constructs.* ……..208
Figure 6.21 *Titration of ^{15}N-Spa33$_{220-293}$-His with 1.5 mM MixN$_{1-18}$.*.....209

7. Characterisation of the *Shigella T3SS* C-ring

Figure 7.1 *Sequence alignments of Spa33 homologues*……213
Figure 7.2 *15% SDS-PA gel showing over-expression of His-Spa33 in B834 (DE3) pLysS.*.....214
Figure 7.3 *Co-purification of His-Spa33 with contaminants.* 215
Figure 7.4 *Identification of the contaminants from His-Spa33 purification.* 216
Figure 7.5 *Identification of Spa33-C.*.....217

Figure 7.6 *The alternative translation product of the spa33 gene*……218
Figure 7.7 *Optimised purification of the His-Spa33/Spa33$_{192-293}$ complex.* 221
Figure 7.8 *Analysis of the oligomeric state of the His-Spa33/Spa33$_{192-293}$ complex by SEC and in-line multi-angle light scattering.*..... 222
Figure 7.9 *Native Mass Spectrometry of the His-Spa33/Spa33$_{192-293}$ complex.*..... 223
Figure 7.10 *Dissociation of His-Spa33/Spa33$_{192-293}$ complexes during native mass spectrometry*. 226
Figure 7.11 *Models for monomeric Spa33-C.*..... 228
Figure 7.12 *Residues implicated in Spa33-(Spa33$_{192-293}$)$_2$ complex formation from FliM-FliN interaction analysis*. 231
Table 7.1 HADDOCK statistics from the docking of the Spa33-C monomer model and Spa33$_{208-293}$ dimer structure…….....232
Figure 7.13 *Docking of the Spa33-C monomer model and the Spa33$_{208-293}$ dimer structure.* ... 233
Figure 7.14 *Tetramer models for Spa33$_{208-293}$.* 234
Figure 7.15 *Analysis of the oligomeric state of Spa33$_{220-293}$ constructs by SEC and in-line multi-angle light scattering.* 235
Figure 7.16 *Native Mass Spectrometry of Spa33$_{220-293}$* 237
Table 7.2 Primers used to amplify *his-spa33* and *spa33$_{192-293}$*…….....240
Figure 7.17 *1% agarose gels of DNA fragments from pETDuet-his-spa33/spa33$_{192-293}$ cloning.* 241
Figure 7.18 *15% SDS-PA gel showing over-expression of His-Spa33 and Spa33$_{192-293}$ in B834 (DE3) pLysS.* 242
Figure 7.19 *Preliminary cross-linking of His-Spa33-(Spa33$_{192-293}$)$_2$ complexes.* 244
Figure 7.20 *A model for the Spa33-(Spa33$_{192-293}$)$_2$ complex in the context of the Shigella T3SS.* 247

8. Appendix

Table 8.1 Outline of MxiG constructs used in this study…… .268
Table 8.2 Outline of MxiJ constructs used in this study……269
Table 8.3 Outline of Spa33 constructs used in this study……..270
Table 8.4 Outline of MxiN constructs used in this study……..271
Table 8.5 The heat-shock protocol for transformation……..272
Table 8.6 Composition of solutions used in SDS-PAGE……..273
Table 8.7 PCR protocol used for site-directed mutagenesis……..274
Table 8.8 Standard PCR protocol for amplification of inserts…….....275
Table 8.9 Touch-down PCR protocol for amplification of inserts……276
Table 8.10 Calibration reactions used to optimise dye and protein dilutions for ThermoFluor277
Table 8.11 Treatment of western blotting membrane after protein transfer……..…….278
Table 8.12 Key parameters of NMR spectra collected in this study……..279
Table 8.13 Design of constructs for the expression of the MxiD periplasmic domain……..284
Figure 8.1 *MxiIΔ5 is partially unfolded.* 286
Table 8.14 Kinetic fits of SPR curves of ComP binding to DUS, SUD and SDU sequences...287
Figure 8.2 *SPR analysis of the DNA binding properties of N. meningitidis ComP.* 288

1. Introduction

1.1 Bacterial Secretion Systems

In order to survive, bacteria must be able to sense, adapt to and alter their biological environment, requiring the external action of proteins synthesised within the cell. Bacterial secretion systems are therefore required for the translocation of these effectors across the cell membranes and peptidoglycan layer, which act as physical and chemical barriers to the extracellular milieu, whilst maintaining cellular integrity. Despite the huge repertoire of secreted substrates, many of these specialised nanomachines can be classified into eight major families, designated type I-VIII secretion systems (**Figure 1.1**). Commonly, these secretion systems facilitate export of virulence factors to their site of action, making them a primary mediator of pathogenesis and therefore a key target for study to further understanding of human disease. (Kostakioti *et al.*, 2005, Waksman, 2012)

Whilst it is unclear how the type VII secretion system facilitates transport of substrate assemblies across the thick peptidoglycan cell wall of Gram positive bacteria (Das *et al.*, 2011), two broad strategies have evolved to allow protein transport across the inner and outer membrane of Gram negative bacteria. 'Two-step' systems, represented by the type II, V and VIII secretion systems (**Figure 1.1, asterisks**), require protein substrates to first be transported into the periplasm in an unfolded or folded state by the ubiquitous inner membrane Sec (Lycklama a Nijeholt & Driessen, 2012) or Tat (Fröbel *et al.*, 2012) systems respectively, before passage through a specialised outer membrane translocator. In contrast, 'one-step' systems, represented by the type I, III, IV and VI secretion systems, are more complex multi-protein structures that span both membranes and provide a continuous channel through which unfolded substrates can be exported without entering the periplasm. (Rêgo *et al.*, 2010)

The diversity of the secretion systems is partly a reflection on the huge repertoire of species- and environment-specific substrates. For example, the capacity of the type II secretion system to

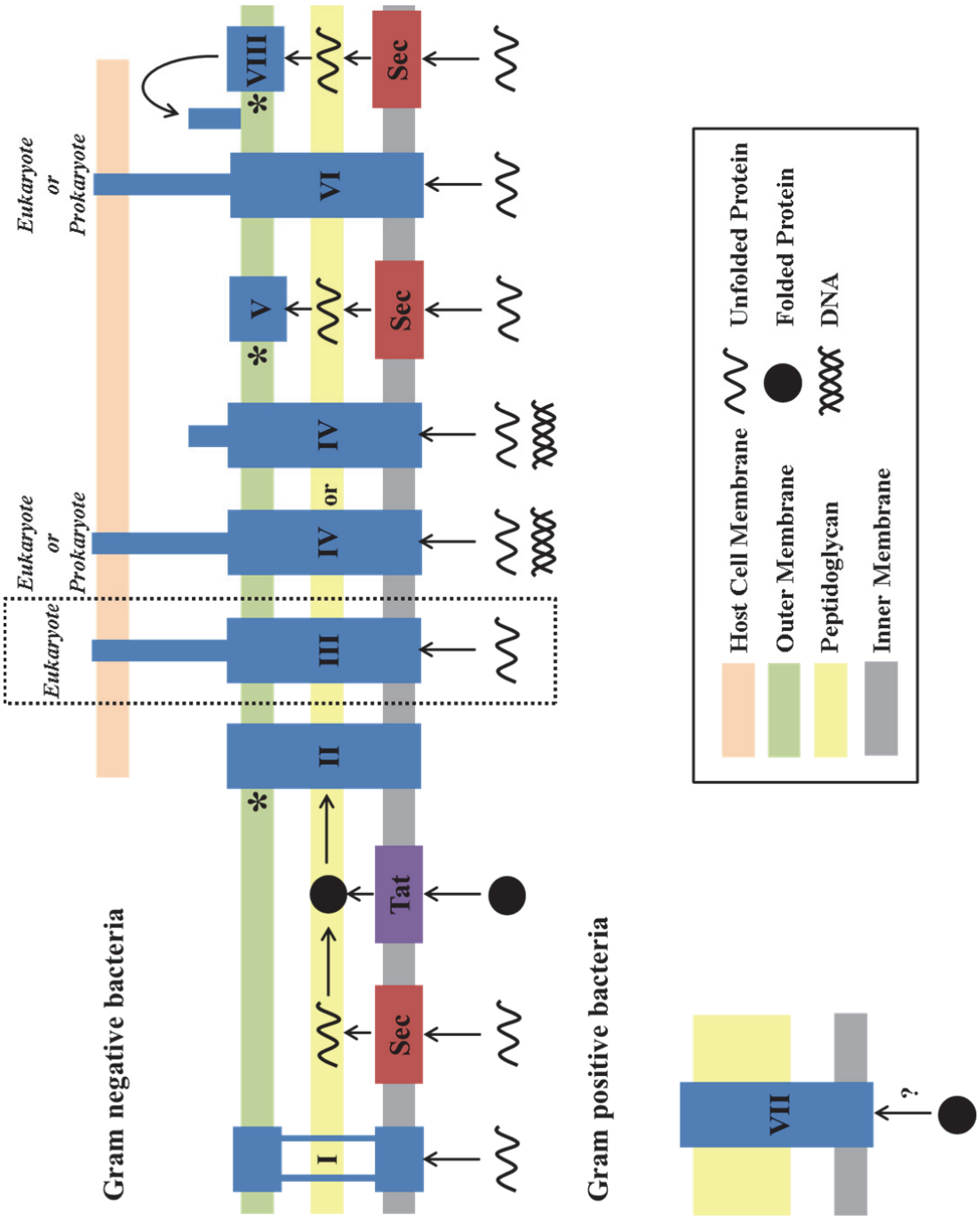


Figure 1.1 *The diversity of bacterial secretion systems*.

Type I-VIII secretion systems (blue) are found exclusively within the membrane of Gram negative or Gram positive bacteria as shown, with the exception of the type IV secretion system found in both types of bacteria. The nature of both the host cell for type III, IV and VI secretion systems and the transported substrate and is shown. Type II, V and VIII secretion systems are 'two-step' systems (asterisks), requiring initial transport to the periplasm by the inner membrane Sec (red) or Tat (purple) systems before transport to the extracellular milieu by the specialised outer membrane system. The type III secretion system (T3SS) is indicated by a dashed box.

transport fully-folded proteins across the outer membrane allows disulphide-bond formation in the periplasm, as was found to be essential for the proteolytic activity of elastase secreted by *Pseudomonas aeruginosa* (Braun *et al.*, 2001). In contrast, the same species uses a type VI secretion system for the one-step export of Tse1 and Tse3, as their peptidoglycan hydrolytic activity would be detrimental if allowed to enter the periplasm en route (Russell *et al.*, 2011).

In addition, the widely variable sites of action for extracellular proteins, as exemplified by virulence factors acting in the gut, have resulted in secretion systems that are adapted to deliver substrates to different locations. Whilst secretion of haemolysin A directly into the extracellular milieu by the uropathogenic *Escherichia coli* type I secretion system is sufficient to cause extra-intestinal infection (Delepelaire, 2004), type III, IV and VI secretion systems of other intestinal pathogens perforate eukaryotic host cell membranes and use a needle-like appendage to secrete virulence factors directly into their cytoplasm (Hodges & Hecht, 2012). In contrast, subunits secreted by the type VIII secretion system nucleate on the enterobacterial cell surface to form a filament or 'curli', which subsequently promotes cell adhesion and invasion (Barnhart & Chapman, 2006).

Despite the diversity of secreted substrates, there are several common requirements. Proteins for secretion must clearly be targeted to the correct system, requiring distinct sequence or structural motifs in the protein or mRNA. Whilst invariant N-terminal signal peptides are clearly defined for substrates of the Sec and Tat systems (Natale *et al.*, 2008), the secretion signal for systems such as the type III secretion system is less well-defined, although there is believed to be a characteristic amino acid composition bias at the N-termini of effector proteins (Arnold *et al.*, 2009). With the exception of the Tat, type II and potentially type VII secretion systems, most secretion systems transport unfolded proteins from the cytoplasm, requiring them to interact with chaperones to prevent premature folding or aggregation. Furthermore, these chaperones could also play a role in targeting, particularly to the type III and IV secretion systems. (Fattori *et al.*, 2011)

In addition, different classes of secretion system apparatus share similar structures and mechanisms. For example, type II, III and IV secretion systems appear to be at least partially energised by traffic-associated ATPases that form functional hexamers (Robien *et al.*, 2003, Claret *et al.*, 2003, Misic *et al.*, 2010). Similarly, type III, IV and VI secretion systems first need to mediate the assembly of either a flagellum-, pilus- or bacteriophage-like hollow needle respectively for transit of effector proteins. Therefore, stringent control of secretion hierarchy is required to separate needle subunit secretion from that of effectors, often in response to the extracellular environment. (Blocker *et al.*, 2008, Zechner *et al.*, 2012, Cascales & Cambillau, 2012). Similarities also exist between the core membrane-spanning components of the secretion systems. Notably, type II and type III secretion systems commonly use proteins from the secretin family to form the outer membrane portion of the translocator, displaying sequence and structural homology in the periplasmic region that is usually variable across other secretin family members (Korotkov *et al.*, 2009, Spreter *et al.*, 2009).

Clearly there are common themes between the different classes of bacterial secretion system, despite a huge diversity in function. Therefore, focused study of one system can be aided by established evidence from other bacterial secretion systems and in turn has the potential to uncover further general principles in their mechanism and structure. To this end, the type III secretion system (T3SS) was selected for further investigation in this study (**Figure 1.1, dashed box**).

1.2 The Translocation-Associated Type III Secretion System (T3SS)

The one-step T3SS of Gram negative bacteria represents a widely studied secretion system that can be sub-divided into two categories: the translocation-associated and flagellar T3SS. Translocation-associated T3SSs allow the injection of effector proteins directly into target eukaryotic cells to alter their function in favour of bacterial survival. These secretion systems mediate the assembly of an extracellular needle and subsequent insertion of a translocon into the host cell membrane, providing a conduit for effector secretion. In contrast, flagellar-T3SSs in motile bacteria mediate the sequential assembly of an extracellular hook and filament, which subsequently rotate to propel the cell through aqueous medium. (Büttner, 2012, Erhardt *et al.*, 2010) Therefore, the T3SS serves to produce two distinct appendages that perform very different functions for the bacterium. Indeed, some bacteria, as exemplified by the enteropathogen *Salmonella typhimurium*, use an arsenal of T3SSs to ensure their survival; whilst flagella allow the cell to move towards and adhere to gut epithelium, distinct translocation-associated T3SSs encoded within the *Salmonella* pathogenicity island (SPI) 1 and SPI-2 promote internalisation by the host cell and subsequent intracellular survival respectively (Ibarra & Steele-Mortimer, 2009).

Despite the clear difference between the function of their extracellular appendages, translocation-associated and flagellar T3SSs have a similar overall architecture (**Figure 1.2**), with many components being conserved between the systems in either sequence or function (**Table 1.1**). This is in agreement with phylogenetic studies that suggest the two types of T3SS share a common ancestor (Gophna *et al.*, 2003) and the finding that the translocation-associated T3SS of enteropathogenic *E. coli* (EPEC) can secrete the flagellar T3SS substrate FliC to stimulate a pro-inflammatory response in the gut (Badea *et al.*, 2009). Unsurprisingly, the key players involved in the secretion process are the most conserved components, including the five integral membrane proteins of the export apparatus and the cytoplasmic ATPase. In contrast, the most striking differences can be attributed to the specific functional requirements for the fully

assembled structure, namely the contiguous translocon inserted into the host cell membrane by the translocation-associated T3SS (**Figure 1.2, green**) and the membrane-embedded stator that drives flagellum rotation using energy from proton influx (**Figure 1.2, red**)(Yonekura *et al.*, 2011). Intermediate conservation is shown by the dual-membrane spanning basal bodies, which essentially act as scaffolds for the export apparatus and extracellular appendages; within both systems they comprise an inner and outer membrane ring (IMR and OMR respectively) and an internal rod structure, but differ in complexity and composition likely due to the additional requirement for the flagellar basal body to rotate *in situ*. Similarly, the cytoplasmic-ring (C-ring) has an additional role in the flagellar T3SS as the switch complex, acting in concert with FliG to generate torque and control direction switching during flagellum rotation (Stock *et al.*, 2012). A knowledge of these similarities and differences allowed the study of translocation-associated T3SSs presented in this thesis to be informed by and related to the flagellar T3SS, providing insights into both nanomachines.

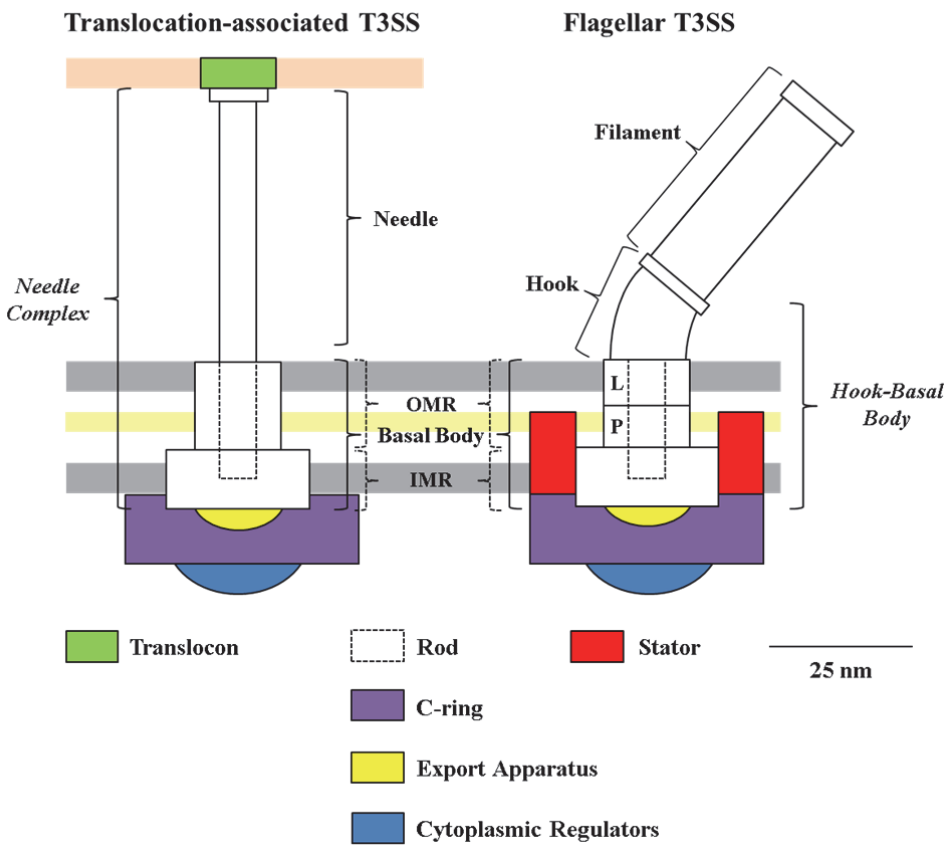


Figure 1.2 *Comparison of translocation-associated and flagellar T3SSs*. Protein components of regions labeled in bold are shown in **Table 1.1**.

Table 1.1 The components of translocation-associated and flagellar T3SSs.

Translocation-associated T3SS				Flagellar T3SS	Function
Shigella spp.	*Salmonella* SPI-1	*Yersinia* spp.	EPEC		
					Basal Body
MxiG	PrgH	YscD	EscD	---	Inner membrane ring
MxiJ	PrgK	YscJ	EscJ	FliF	Inner membrane ring*
MxiD	InvG	YscC	EscC	---	Outer membrane ring (Secretin)
MxiM	InvH	YscW	?	---	Pilotin
---	---	---	---	FlgH	Outer membrane ring (L ring)
---	---	---	---	FlgI	Outer membrane ring (P ring)
MxiI	PrgJ	YscI	EscI	---	Inner rod
---	---	---	---	FlgB, FlgC, FlgF	Proximal rod
---	---	---	---	FlgG	Distal rod
---	---	---	---	FlgJ	Rod cap
					Needle
MxiH	PrgI	YscF	EscF		Needle protein
IpaD	SipD	LcrV	---		Needle tip
					Hook
				FlgE	Hook protein
				FlgD	Hook cap
				FlgL, FlgK	Hook-filament junction
					Filament
			EspA	FliC	Filament protein
				FliD	Filament cap
					Translocon
IpaB, IpaC	SipB, SipC	YopB, YopD	EspD, EspB		Pore-forming proteins inserted into host cell membrane
Spa33	SpaO	YscQ	EscQ	FliM, FliN (FliG)	**C-ring*** (& switch complex)
MxiA	InvA	YscV	EscV	FlhA	
Spa40	SpaS	YscU	EscU	FlhB	
Spa9	SpaQ	YscS	EscS	FliQ	**Export Apparatus***
Spa24	SpaP	YscR	EscR	FliP	
Spa29	SpaR	YscT	EscT	FliR	
				FliO	
					Cytoplasmic Regulators
Spa47	InvC	YscN	EscN	FliI	ATPase*
MxiN	OrgA	YscL	EscL	FliH	Regulator of needle assembly*
MxiK	OrgB	YscK	EscK	FliG?	Regulator of needle assembly
Spa32	InvJ	YscP	Orf16	FliK	Needle length regulator
Spa13	InvI	YscO	EscA	(FliJ)	Chaperone (ATPase γ stalk)*
MxiC	InvE	YopN/ TyeA	SepL/ SepD	---	Regulation of effector secretion
				---	
				MotA/MotB	**Stator**

Components are classified according to their location in the translocation and/or flagellar T3SS (**Figure 1.2**). Orthologues are on the same line. Protein families relevant to this study are highlighted in yellow. *Proteins which display sequence homology between translocation-associated and flagellar T3SS.

Classification of Translocation-Associated T3SSs

Translocation-associated T3SSs are mostly distributed in animal-pathogenic, plant-pathogenic and symbiotic bacteria, allowing classification into eight major families (**Table 1.2**). Interestingly, there is evidence to suggest that *Chlamydia* spp. have the most ancient T3SS and could be evolutionary progenitors of the system (Kim, 2001). Due to their widespread contribution to human disease, the T3SSs of intestinal pathogens are the most widely studied (**Table 1.2, blue**)(Reis & Horn, 2010). Therefore, our understanding of the animal-pathogenic T3SS as outlined in this section is largely derived from four model organisms: *Shigella* spp. and related *S. typhimurium* SPI-1, *Yersinia* spp. and EPEC. Apart from the EspA helical filament that sits atop the needle in the EPEC T3SS (Sekiya *et al.*, 2001), the components of these archetypal T3SSs are largely conserved (**Table 1.1**), allowing species-specific studies to contribute to the general principles involved in T3SS function and assembly.

Table 1.2 The eight families of translocation-associated T3SSs.

Family	Model Organism	Role of Bacteria-Host Interaction
Inv-Mxi-Spa	*Shigella* spp. (*S. flexneri*) *S. typhimurium* SPI-1	Animal pathogen (intestinal)
Ysc	*Yersinia* spp. (*Y. pestis, Y. pseudotuberculosis*)	Animal pathogen (intestinal)
Ssa-Esc	EPEC (*E. coli* T3SS 1 (ETT1)) *S. typhimurium* SPI-2	Animal pathogen (intestinal)
Eiv-Epa	*E. coli* T3SS 2 (ETT2)	Animal pathogen
Hrp-1	*Pseudomonas syringae* *Erwinia* spp.	Plant pathogen
Hrp-2	*Xanthomonas* spp. *R. solanacearum*	Plant pathogen
Chlamydiales	*Chlamydia* spp.	Animal pathogen (target uncharacterised)
Rhizobiales	*Rhizobium* spp.	Symbiotic (target uncharacterised)

Structural Approaches to Studying the Translocation-Associated T3SS

In order to fully understand mechanism and assembly, the translocation-associated T3SS must be characterised at the structural level. Furthermore, given the widespread involvement of the system in pathogenesis, a knowledge of its structure is required to inform design of antivirulence therapies (Cegelski *et al.*, 2008). Clearly, the ultimate goal would be the determination of a high resolution structure of the entire secretion system from one of the model animal pathogens. However, as the translocation-associated T3SS is a large macromolecular complex embedded in both the inner and outer bacterial membrane (**Figure 1.2**), it provides a challenging target for structural biology. Indeed, the intact needle complex (NC) of the translocation-associated T3SS, comprising the extracellular needle and membrane-spanning basal body, has only been characterised at low resolution by electron microscopy (EM). To date, the best reconstruction of the NC has been obtained for the *S. typhimurium* SPI-1 T3SS by cryo-EM at 10 Å resolution with C3 symmetry applied, providing a detailed representation of overall morphology (**Figure 1.3A**)(Schraidt & Marlovits, 2011). However, the export apparatus, C-ring and cytoplasmic regulators are largely lost during the stringent purification required to obtain a homogeneous sample of the complex for structural studies, resulting in visualisation of only partial sub-complexes within *S. typhimurium in vitro* complexes (Wagner *et al.*, 2010) and of amorphous assemblies within native membranes (Blocker *et al.*, 1999). Therefore, the interactions within and between certain sub-complexes of the translocation-associated T3SS could be too weak for their *in vitro* isolation and characterisation.

Therefore, a common approach to T3SS structural characterisation has been the determination of high resolution structures of isolated proteins and their composite domains, using both X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (**Figure 1.3B**). Whilst such structures are available for many components of the system, the regulators of needle assembly and length, inner rod and much of the export apparatus remain uncharacterised (**Table 1.3**). In addition, despite the translocation-associated T3SS being a membrane-embedded complex, structures are only available for soluble domains of integral membrane proteins

(**Table 1.3, asterisks**), requiring further evidence to understand their structure within the context of the intact protein in its native environment. Therefore, there is much scope for further characterisation of the translocation-associated T3SS by this approach.

Table 1.3 High resolution structures of isolated translocation-associated T3SS proteins and domains available before this study

Function	Structure
Basal Body	
Inner membrane ring (MxiG)*	*S. typhimurium* $PrgH_{170-362}$ (Spreter *et al.*, 2009) (periplasmic domain only)
Inner membrane ring (MxiJ)*	EPEC $EscJ_{21-190}$ (Yip *et al.*, 2005), $EscJ_{20-190}$ (Crepin *et al.*, 2005) (periplasmic domain only)
Outer membrane ring (MxiD)*	EPEC $EscC_{21-174}$ (Spreter *et al.*, 2009) (periplasmic domain only)
Pilotin (MxiM)	*S. flexneri* $MxiM_{28-142}$ (Lario *et al.*, 2005, Okon *et al.*, 2008)
Inner Rod	-----
Needle	
Needle protein (MxiH)	*S. flexneri* MxiH (Deane *et al.*, 2006); *S. typhimurium* PrgI (Poyraz *et al.*, 2010); *Burkholderia pseudomallei* BsaL (Zhang *et al.*, 2006)
Needle tip (IpaD)	*S. flexneri* IpaD & *B. pseudomallei* BipD (Johnson *et al.*, 2007); *Y. pestis* LcrV (Derewenda *et al.*, 2004); *S. typhimurium* SipD (Chatterjee *et al.*, 2011)
C-ring (Spa33)	*P. syringae* $HrcQ_B$-C (Fadouloglou *et al.*, 2004)
Export Apparatus	
(MxiA)*	*S. typhimurium* $InvA_{357-685}$ (Worrall *et al.*, 2010); $MxiA_{318-686}$ (Abrusci *et al.*, In Press)
(Spa40)*	*S. flexneri* $Spa40_{237-338}$ (Deane *et al.*, 2008a); *Y. pestis* $YscU_{211-354}$ (Lountos *et al.*, 2009); *S. typhimurium* $SpaS_{211-356}$ & EPEC $EscU_{215-345}$ (Zarivach *et al.*, 2008)
(Spa9)	-----
(Spa24)	-----
(Spa29)	-----
Cytoplasmic Regulators	
ATPase (Spa47)	EPEC EscN (Zarivach *et al.*, 2007)
Regulators of needle assembly (MxiN & MxiK)	-----
Needle length regulator (Spa32)	-----
Regulation of effector secretion (MxiC)	*S. flexneri* MxiC (Deane *et al.*, 2008b); *Y. pestis* YopN/TyeA (Schubot *et al.*, 2005)

Components are classified according to the fuctional categories assigned in **Figure 1.2** and **Table 1.1**. The homologue from *Shigella* spp. is shown in brackets after each function. Existing structures of relevance to this study are highlighted in yellow.
*Integral membrane proteins for which structures of soluble domains only are available.

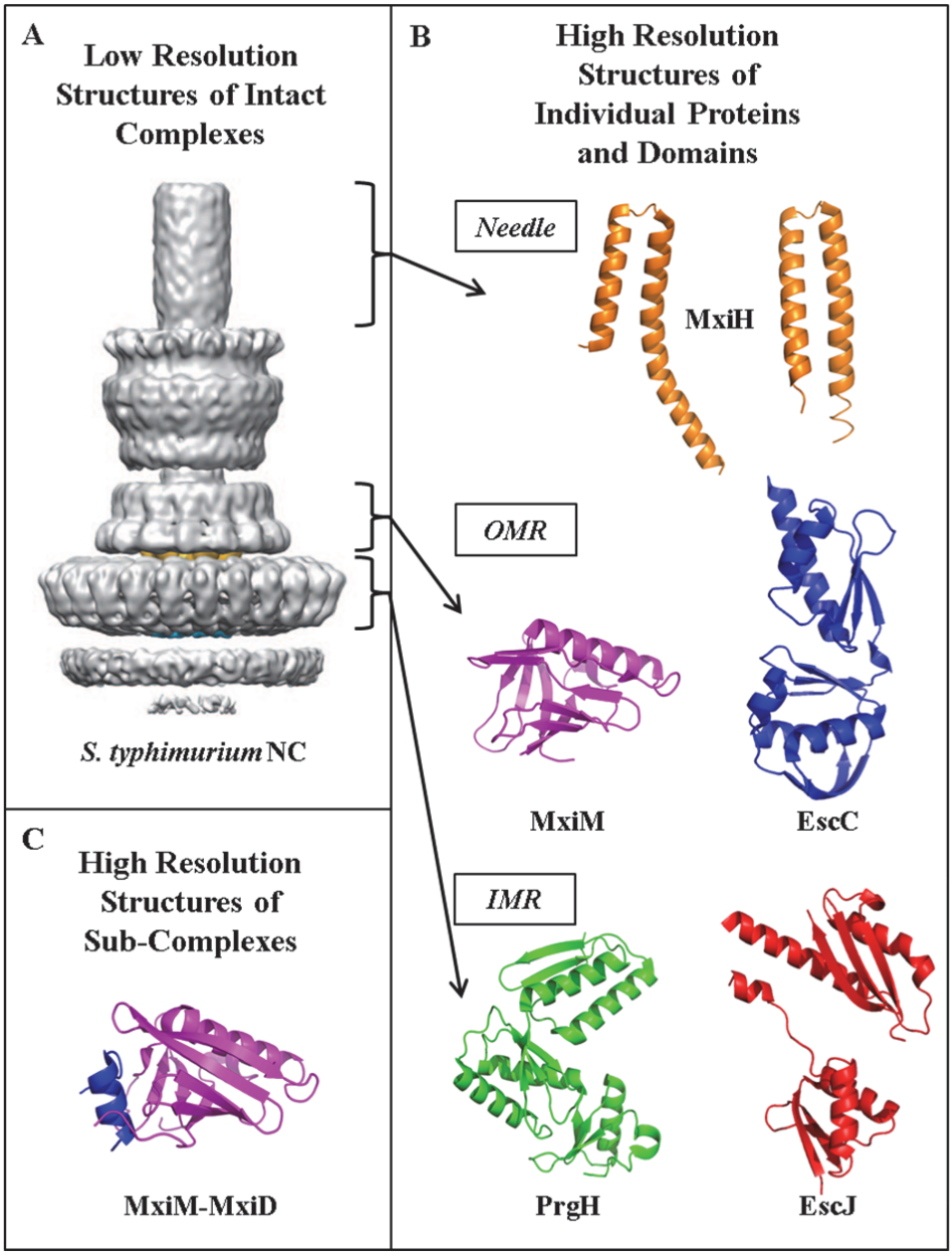


Figure 1.3 *The three approaches to structural characterisation of the translocation-associated T3SS.*

A Cryo-EM reconstruction of the *S. typhimurium* NC at 10 Å resolution with C3 symmetry applied (EMDB 1875) taken from Schraidt and Marlovits, 2011.

B Crystal structures of *S. flexneri* MxiH (orange; pdb id 2ca5; left conformation = residues 20-81, right conformation = residues 15-75)(Deane *et al.*, 2006), *S. flexneri* MxiM residues 28-142 (magenta; pdb id 1y9l)(Lario *et al.*, 2005), EPEC EscC residues 21-174 (blue; pdb id 3gr5)(Spreter *et al.*, 2009), *S. typhimurium* PrgH residues 170-362 (green; pdb id 3gr0)(Spreter *et al.*, 2009), EPEC EscJ residues 21-190 (red; pdb id 1yj7)(Yip *et al.*, 2005). Structures are shown relative to their positions within the EM reconstruction shown in **A**.

C The solution structure of *S. flexneri* MxiM residues 28-142 (magenta) complexed with MxiD residues 553-570 (blue; pdb id 2jw1)(Okon *et al.*, 2008).

As the translocation-associated T3SS is a large protein complex, these high resolution structures of individual proteins and their domains must be placed within the context of the entire secretion system in order to fully understand their biological relevance. In the case of the components of the NC, this has been achieved by docking their high resolution structures (**Figure 1.3B**) into the lower resolution envelope of the intact complex (**Figure 1.3A**)(Hodgkinson *et al.*, 2009, Spreter *et al.*, 2009, Schraidt & Marlovits, 2011). However, such studies rely on matching defined surface features between EM reconstructions and high resolution structures and assume the isolated component has the same structure within the intact complex. Indeed, docking of components into EM density can be ambiguous, requiring additional biochemical data to validate their orientations within the assembly. For example, docking of the MxiH crystal structure (**Figure 1.3, orange**) into the 16 Å EM reconstruction of the *S. flexneri* needle (Cordes *et al.*, 2003) suggested only one of the observed conformations was compatible with the observed density and was arranged with the N-terminus pointing into the centre of the hollow needle (Deane *et al.*, 2006). However, solid state NMR derived restraints between PrgI monomers within recombinant *S. typhimurium* needle fragments suggested an opposing model with the N-terminus pointing towards the exterior of the needle (Loquet *et al.*, 2012). In addition, this docking approach cannot be used to elucidate interactions within the C-ring, export apparatus and cytoplasmic regulators, as EM data are not available for their complexes.

Alternatively, high resolution structure determination of sub-complexes of the translocation-associated T3SS would enable characterisation of their interaction interfaces at the atomic level. However, this approach has previously been used with only limited success (**Table 1.4**), with the NMR structure of *S. flexneri* MxiM complexed with a small fragment of MxiD (**Figure 1.3C**)(Okon *et al.*, 2008) and the crystal structure of self-associated $EscJ_{21-190}$ (Yip *et al.*, 2005) representing the only high resolution structures of sub-complexes from the hetero-oligomeric basal body. Indeed, characterisation of putative interaction interfaces within the T3SS has been largely reliant on biochemical approaches, such as cross-linking and site-directed mutagenesis,

leaving a niche for techniques that would define the molecular nature of the interaction such as chemical-shift mapping from NMR titrations.

Table 1.4 High resolution structures of sub-complexes of the translocation-associated T3SS

Function	Structure
Basal Body	
Inner membrane ring (MxiG)	-----
Inner membrane ring (MxiJ)	Superhelical arrangement of EPEC $EscJ_{21-190}$ (Yip *et al.*, 2005)
Outer membrane ring (MxiD)	-----
Pilotin (MxiM)	*S. flexneri* $MxiM_{28-142}$ and $MxiD_{553-570}$ (Okon *et al.*, 2008)
Inner Rod	
Needle	
Needle protein (MxiH) / Needle tip (IpaD)	*S. typhimurium* PrgI-SipD (Lunelli *et al.*, 2011)
C-ring (Spa33)	*P. syringae* $HrcQ_B$-C tetramer (Fadouloglou *et al.*, 2004)
Export Apparatus	-----
Cytoplasmic Regulators	-----

Components are classified according to the categories assigned in **Figure 1.2** and **Table 1.1**. The homologue from *Shigella* spp. is shown in brackets after each function. Exisiting structures of particular interest for this thesis are highlighted in yellow.

In summary, this multi-pronged approach to structural characterisation of the translocation-associated T3SS (**Figure 1.3**) has already been used to some success, shaping our current model of secretion as described next. However, there is clearly still much to be learnt structurally about the system, requiring that these themes be taken forward for studies such as those carried out in this thesis.

A Working Model for Hierarchical Assembly and Secretion by the T3SS

As the translocation-associated T3SS is responsible for both the assembly of its extracellular appendages and the subsequent injection of effector proteins into host cells, secretion by the system must occur with a defined hierarchy to prevent premature release of substrates into the extracellular milieu. A secretion-competent complex is initially formed from the membrane-spanning basal body, the cytoplasmic components and the export apparatus, which then successively exports substrates required to form the inner rod, needle, needle tip and translocon

and finally effector proteins (**Figure 1.4**). Regulatory proteins therefore have to mediate both the targeting of the required substrates for secretion and the substrate-specificity switches that are required to progress between each mode of secretion. (Deane *et al.*, 2010) A model for this assembly and regulation of the T3SS based on current evidence is outlined in this section.

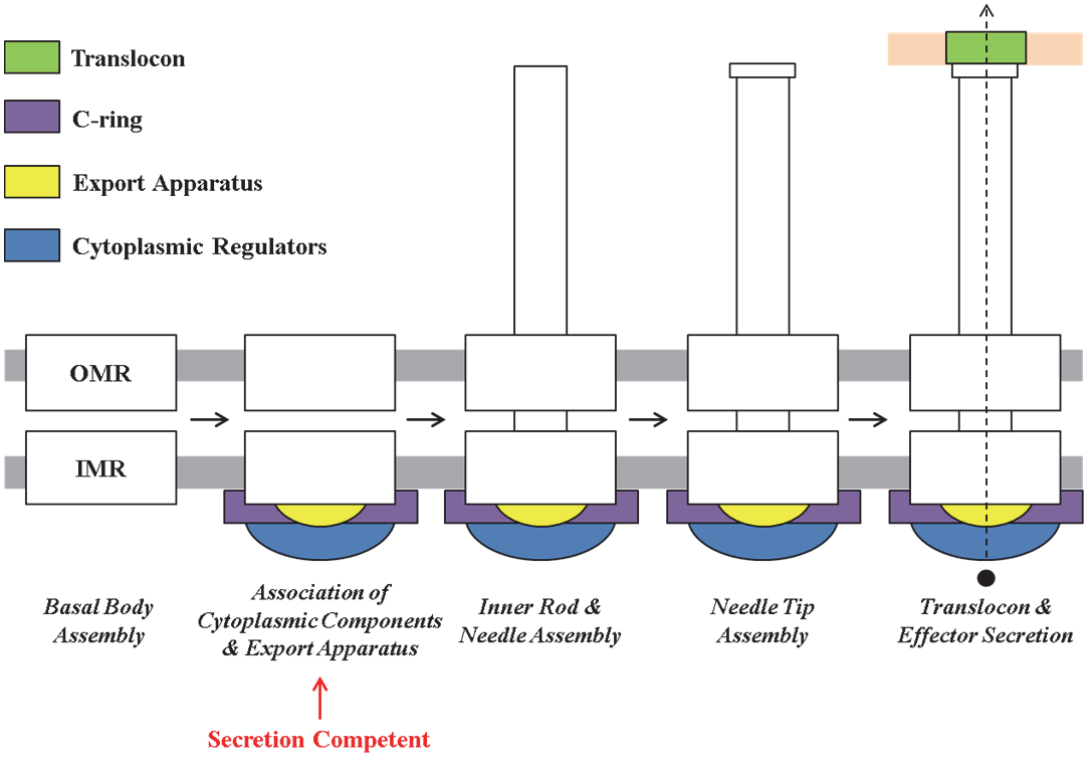


Figure 1.4 *The hierarchy of T3SS assembly and secretion.*

Assembly of the Basal Body

The overall morphology of the basal body of the translocation-associated T3SS shows distinct IMR and OMR rings, found by immunoblotting and N-terminal sequencing to comprise MxiG, MxiJ and MxiD in *S. flexneri* (Tamano *et al.*, 2000, Kubori *et al.*, 1998). Subsequently, an EM reconstruction of NCs purified from a *invG*⁻ *S. typhimurium* strain showed no density for the OMR, indicating that MxiG and MxiJ homologues together form the IMR and MxiD homologues form the OMR (Schraidt *et al.*, 2010). These proteins each exhibit a targeting

sequence for the Sec pathway, implicating this as the route for inner membrane transport (Sukhan *et al.*, 2001). Indeed, MxiD and MxiJ each have a stretch of 22 and 17 hydrophobic amino acids at their N-termini respectively (Allaoui *et al.*, 1992, Allaoui *et al.*, 1993) that are absent in intact NCs (Tamano *et al.*, 2000). In contrast, the N-terminus of MxiG is not cleaved (Tamano *et al.*, 2000), suggesting the single transmembrane region acts as the signal for inner membrane insertion (Allaoui *et al.*, 1995).

The C-terminus of MxiD shows homology to the outer membrane-spanning region found in the secretin family of proteins (Allaoui *et al.*, 1993), indicating an analogous OMR to that found within the type II secretion system (Nouwen *et al.*, 1999). As is typical for secretins, MxiD that has been transported into the periplasm requires the lipidated pilot protein MxiM for stability and subsequent association with the outer membrane (Schuch & Maurelli, 2001, Diepold *et al.*, 2010). High resolution structures of MxiM show a hydrophobic cavity that is predicted to bind both its own N-terminal lipid moiety and a C-terminal α-helix of MxiD in a mutually exclusive manner, with displacement of the lipid moiety by MxiD allowing its subsequent membrane insertion to facilitate localisation (**Figure 1.3**)(Lario *et al.*, 2005, Okon *et al.*, 2008). Indeed, MxiM was detected in purified NCs by immunoblotting (Zenk *et al.*, 2007) and putative density for the 15 kDa protein has been observed adjacent to the OMR in EM reconstructions of the *S. flexneri* NC (Sani *et al.*, 2007).

When *S. typhimurium* PrgH and PrgK were recombinantly overexpressed in *E. coli*, ring structures that were virtually identical in size and shape to the IMR could be purified and visualised by EM (Kimbrough & Miller, 2000). Similarly, ring structures reminiscent of the OMR could be purified from a *prgK⁻ prgH⁻* mutant. It was also found that a lower proportion of intracellular InvG was present in high molecular weight complexes in the mutant, indicating PrgK and PrgH enhance ring assembly. (Sukhan *et al.*, 2001) Together these results could suggest a model for basal body assembly whereby the IMR and OMR form independently then self-associate to form a more stable complex.

However, this model is contradicted by results from *Yersinia enterocolitica* showing that N-terminally-tagged YscD cannot localise to the membrane or copurify YscJ in the absence of YscC (Diepold *et al.*, 2010). This implicates an alternative stepwise model for basal body assembly, whereby initial formation of the OMR is required to nucleate the assembly of MxiG and MxiJ homologues within the IMR. Therefore, the mechanism of basal body assembly is currently unclear and could be found to vary between species.

Association of Cytoplasmic Components and Export Apparatus

The Sec pathway is also likely to mediate insertion of the five conserved integral membrane proteins of the export apparatus, all of which are essential for secretion by both flagellar and translocation-associated T3SSs. Whilst complexes isolated from the *invG* ⁻ strain of *S. typhimurium* have an assembled needle (Schraidt *et al.*, 2010), PrgI was not detected in complexes purified from a *prgK*⁻ *prgH*⁻ mutant (Sukhan *et al.*, 2001), suggesting the IMR of the basal body could be acting as a scaffold for association of the export apparatus. Indeed, Spa40, Spa24 and Spa29 homologues have been identified in NC preparations, with the cup structure within the basal body corresponding to putative density for Spa24 and Spa29 (Wagner *et al.*, 2010, Zenk *et al.*, 2007). This was supported by further evidence showing that YscV moved laterally in the membrane in the absence of the basal body and interacted either directly or indirectly via other export apparatus components with YscJ (Diepold *et al.*, 2011). In addition, a reduced level of basal body assembly was observed in the absence of the export apparatus, indicating either incorporation of the export apparatus into the basal body base or assembly of the IMR around the export apparatus increases the stability of the NC (Wagner *et al.*, 2010). Indeed, oligomers of FlhA were found to promote oligomerisation of the IMR component FliF in the flagellar T3SS (Li & Sourjik, 2011). Furthermore, the large cytoplasmic C-terminal domain of *S. flexneri* MxiA has recently been shown to form a homo-nonameric ring with residues essential for secretion lining the central pore, suggesting MxiA homologues could act as an export channel for the T3SS (Abrusci *et al.*, In Press).

The assembly of the basal body also appears to be a prerequisite for localisation of the C-ring to the base of the T3SS, as membrane localisation of fluorescently-tagged YscQ was abolished in the absence of any basal body components (Diepold *et al.*, 2010). Indeed, pull-down assays have indicated an interaction between MxiG and Spa33, suggesting C-ring localisation in *S. flexneri* could be mediated by the IMR (Morita-Ishihara *et al.*, 2006). Furthermore, basal body but not export apparatus formation was shown to be required for association of the YscN ATPase with the membrane. As C-ring and ATPase localisation were also dependent on each other and the regulators of needle assembly, YscK and YscL, these four proteins could be expected to form a complex at the base of the NC. (Diepold *et al.*, 2010) Indeed, several studies on the *S. flexneri* system have reported the isolation of soluble complexes of Spa33, MxiN, Spa47 and MxiK (Jouihri *et al.*, 2003, Morita-Ishihara *et al.*, 2006, Johnson & Blocker, 2008). In addition, knockout mutations of each of these proteins resulted in a failure to form the T3SS needle (Tamano *et al.*, 2000, Morita-Ishihara *et al.*, 2006, Jouihri *et al.*, 2003), indicating that this structural unit could be mediating secretion by the translocation-associated T3SS.

Within the flagellar T3SS, FliJ shows structural similarity to the γ-subunit of the F_1-ATPase, promoting hexamerisation of the FliI ATPase *in vitro* (Ibuki *et al.*, 2011). However, although FliJ homologues are present within the translocation-associated T3SS, they are relatively understudied, with *S. typhimurium* InvI and *Y. enterocolitica* YscO being assigned a role in substrate-targeting by chaperone binding (Evans & Hughes, 2009). In addition, it was recently found that YscO did not promote membrane localisation of fluorescently-tagged YscN (Diepold *et al.*, 2012), suggesting these proteins may not have an analogous role to their counterparts in the flagellar T3SS.

Inner Rod and Needle Assembly

The first substrates to be secreted by the secretion-competent T3SS are the subunits of the inner rod and needle, which share a conserved C-terminal domain (Blocker *et al.*, 2001). Crystal and solution structures of isolated needle subunits from a variety of T3SSs (**Table 1.3**; **Figure 1.3B**)

have been used to model hollow helical assemblies with a ~25 Å lumen (Deane *et al.*, 2006, Loquet *et al.*, 2012, Fujii *et al.*, 2012, Galkin *et al.*, 2010), providing a molecular representation of the translocation-associated T3SS needle. In contrast, although EM reconstructions of the *S. typhimurium* NC show the inner rod to be a major feature within the basal body (Marlovits *et al.*, 2004, Marlovits *et al.*, 2006), the structure of the assembly remains largely uncharacterised and the isolated PrgJ subunit was found to be predominantly unfolded in solution (Zhong *et al.*, 2012). However, EPEC EscI was found to oligomerise and associate with the OMR secretin, indicating it can participate in interactions that would presumably be required for inner rod formation (Sal‐Man *et al.*, 2012).

The length of the translocation-associated T3SS needle is precisely controlled in a species-dependent manner that is thought to correlate to the size of appendages on the bacterial and host cell surface (Mota *et al.*, 2005). Therefore, secretion of needle subunits must be controlled to allow measurement of the needle and cessation of assembly once the defined length is reached. Knockout mutations of *S. flexneri* Spa32 homologues led to the formation of abnormally long needles, implicating this protein as the regulator of needle length (Tamano *et al.*, 2002, Journet *et al.*, 2003). In *Yersinia* spp., insertions and deletions within YscP showed a linear correlation between its residue number and the mean needle length (Journet *et al.*, 2003). This suggested extended YscP could function as a molecular ruler, either by being periodically secreted or by permanently occupying the needle channel throughout assembly. However, the finding that equivalent mutations of Spa33 did not alter needle length suggested that such a model may not apply in *Shigella* spp.. Indeed, overexpression of PrgJ and PrgI resulted in shorter and longer needles respectively, suggesting *S. typhimurium* could control needle length through the relative stoichiometry of inner rod and needle subunits, with completion of the rod signalling completion of the needle (Marlovits *et al.*, 2006). In this alternative model, the molecular ruler protein would exert its control on needle length by regulating the secretion of the inner rod subunits, as suggested by elevated levels of YscI in the absence of YscP (Wood *et al.*, 2008).

The highly conserved cytoplasmic domain of Spa40 homologues, one of the components of the export apparatus, is structurally characterised for all four model animal-pathogenic T3SSs (**Table 1.3**), revealing autocleavage of a NPTH surface loop upon protein folding. As non-cleavable mutants were unable to secrete tip proteins and translocators in *S. typhimurium*, *Y. enterocolitica* and EPEC, Spa40 homologues could have a role in the substrate specificity switch between needle subunit and needle tip protein secretion (Zarivach *et al.*, 2008, Sorg *et al.*, 2007). However, the needle length regulator is also required for this switch in secreted proteins (Magdalena *et al.*, 2002) and directly interacts with Spa40 via a functionally essential region at its C-terminus (Botteaux *et al.*, 2008), suggesting the two proteins likely work in concert to initiate the secretion of needle tip proteins.

Needle Tip Assembly

The needle tip has been visualised by EM in both *Y. enterocolitica* and *S. flexneri* and was recognised by antibodies against LcrV and IpaD respectively, suggesting they are the main constituents of the complex (Mueller *et al.*, 2005, Espina *et al.*, 2006, Epler *et al.*, 2012). Several high resolution structures are available for needle tip proteins (**Table 1.3**), allowing pentamers of LcrV and IpaD to be modelled into their respective electron density from the EM maps (Broz *et al.*, 2007, Epler *et al.*, 2012). Furthermore, the N-terminal domain of the tip protein, which was proposed to function as a self-chaperone from the structures of isolated IpaD and BipD (Johnson *et al.*, 2007), was found to be displaced by an interaction with the needle subunit in the structure of a PrgI-SipD fusion (Lunelli *et al.*, 2011), demonstrating a potential mechanism by which a premature interaction between the needle and tip subunits in the cytosol is prevented. However, this does not appear to be a universal mechanism, as tip proteins in *Y. pestis* and *P. aeruginosa* have been found to interact with a cytoplasmic chaperone (Matson & Nilles, 2001, Lee *et al.*, 2010b).

As the needle tip is present in the absence of effector secretion (Mueller *et al.*, 2005), the complex could potentially play important roles in preventing premature translocon and effector

secretion, sensing host cell contact and insertion of the translocon. Indeed, deletion of *ipaD* in *S. flexneri* led to constitutive effector secretion (Parsot *et al.*, 1995) and ring models of the needle protein homopentamer have shown both closed and open conformations (Mueller *et al.*, 2005, Johnson *et al.*, 2007, Lunelli *et al.*, 2011), suggesting the needle tip plugs the needle. In addition, construction of the IpaD homopentamer model for the *S. flexneri* needle tip required substantial rearrangement of the crystal structure (Epler *et al.*, 2012), suggesting an earlier model for the needle tip with four copies of IpaD and one copy of the translocon component IpaB could be more plausible (Johnson *et al.*, 2007). This model would also be consistent with biochemical evidence showing the localisation of IpaB at the needle tip (Veenendaal *et al.*, 2007) and the requirement of IpaD for efficient translocon insertion (Picking *et al.*, 2005). In addition, the observation that bile salts induce conformational changes in IpaD (Dickenson *et al.*, 2010) could implicate a further role for the needle tip in host cell sensing and triggering of the next substrate specificity switch.

Translocon and Effector Secretion

S. flexneri IpaB and IpaC are both required for secretion of effector proteins into host cells and formed a ~25 Å pore within the membranes of red blood cells (Blocker *et al.*, 1999), indicating they likely form the translocon within the host cell membrane. Indeed, the N-terminal coiled-coil domains of IpaB and SipB were found to be structurally similar to pore-forming proteins from other Gram negative bacteria (Barta *et al.*, 2012). Interestingly, the association of IpaC with the proposed IpaB/IpaD tip complex suggested formation of a ‘pre-pore’ complex that could be inserted into the membrane upon host cell contact (Veenendaal *et al.*, 2007). However, YopB and YopC were not found to be associated with the *Yersinia* tip complex (Mueller *et al.*, 2005), indicating the mechanism of translocon insertion could be species-specific. This would correlate with the diverse environmental triggers for translocon secretion, which have been suggested to be bile salts for *S. flexneri* (Olive *et al.*, 2007) and serum albumin for *Y. enterocolitica* (Lee *et al.*, 2001b).

Although the mechanisms involved in host-cell sensing and translocon insertion are poorly characterised, a change in conformation or chemical environment as a result of translocon completion presumably triggers the substrate-specificity switch required to initiate effector secretion. Whilst effector secretion could simply be initiated by opening of an extracellular 'plug', an alternative cytosolic gating mechanism would require signals from the tip of the needle to be communicated to the base of the NC, either by an unidentified sensor polypeptide inside the needle or by conformational changes within the needle itself (Blocker *et al.*, 2008). Indeed, secretion of effector proteins was selectively enhanced in the absence of *S. flexneri* MxiC, which is itself secreted upon host-cell contact, suggesting MxiC homologues could act to block effector secretion until the translocon is inserted (Botteaux *et al.*, 2009). Furthermore, mutations in the needle subunit that selectively abolished effector protein and MxiC secretion were bypassed by introducing a *mxiC*$^{-}$ mutation, suggesting that the signal for MxiC release and effector secretion could be transmitted by the needle (Martinez-Argudo & Blocker, 2010). Interestingly, a YopN-TyeA complex in *Y. pestis* performs an equivalent role and is structurally similar to MxiC, but shows no sequence conservation on the surface, suggesting their interaction partners are species-specific (Deane *et al.*, 2008b, Schubot *et al.*, 2005).

Taken together, the evidence suggests that T3SS assembly and secretion is not only hierarchical, but also highly cooperative, reliant on the formation of structural assemblies and multiple signalling events for progression to the next stage. It is also clear that many aspects of structure and mechanism are far from understood and show species-specific variation. Therefore, a few key research questions were chosen for study in the *Shigella* T3SS, as outlined in the next section.

1.3 The *Shigella* T3SS

The Role of the *Shigella* T3SS in Pathogenesis

Shigella spp. cause over 1 million deaths annually from bacterial dysentery or Shigellosis, with over 99% cases occurring in developing countries and 60% of these being caused by *S. flexneri* (Kotloff *et al.*, 1999). Given the poor availability of antibiotics in affected countries and the high prevalence of resistant strains (Sack *et al.*, 1997), the pathogen is extensively studied with a view to informing vaccine development (Levine *et al.*, 2007). Within *S. flexneri* most of the genes required for pathogenicity are present on a 214 kb virulence plasmid, with the translocation-associated T3SS encoded by the 30 kb *ipa/mxi-spa* locus being absolutely required for virulence (Buchrieser *et al.*, 2000).

Shigella spp. cause disease by invasion of a variety of cells lining the colon and rectum (**Figure 1.5**). The primary site of action is the M cells of follicle-associated epithelium, where internalisation of the bacteria is followed by vacuole escape and translocation to the lymphoid follicle. Despite the high level of immune cells in this region, *Shigella* spp. are able to survive by triggering macrophage apoptosis. Although the subsequent release of interleukins mounts an immune response against the infection, the induced rupture of the basolateral membrane encourages further epithelial invasion. In addition, *Shigella* spp. are able to proliferate in the cytosol and induce actin polymerisation at one pole of the cell via the surface-exposed outer membrane protein IcsA (Suzuki & Sasakawa, 2001), allowing the bacteria to spread into adjacent epithelial cells and evade the immune system. (Sansonetti, 2001, Sansonetti, 2004).

Several key stages in the *Shigella* pathogenicity cycle are a direct result of the secretion of 'early effectors' (**Table 1.5**) by its T3SS, namely host cell invasion, vacuole escape and macrophage apoptosis (**Figure 1.5, asterisks**). Prior to formation of a translocon in the host cell membrane, IpaB and IpaC bound to the needle tip (Veenendaal *et al.*, 2007) could contribute to initial cell adhesion by binding of surface receptors. Together, secreted IpaB, IpaC, IpaA, IpgD

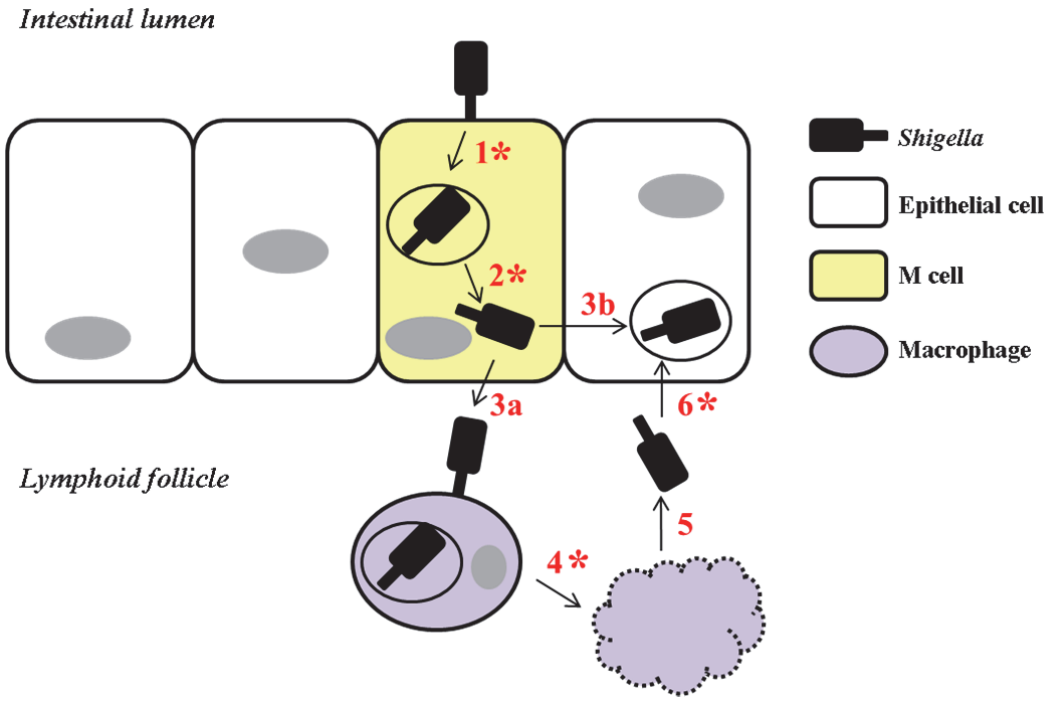


Figure 1.5 *The pathogenesis of Shigella*.

Key steps involved in the pathogenesis of *Shigella* in the intestine are shown, with asterisks denoting those steps triggered by secretion of effectors by the T3SS:
1 Internalisation by M cells
2 Vacuole escape and cytoplasmic proliferation
3a Traversal of M cells and exposure to macrophages in the lymphoid follicle
3b Actin polymerisation and invasion of neighbouring cells
4 Macrophage apoptosis
5 Release of *Shigella* and IL-1β/IL-18 mediated inflammation
6 Basolateral invasion of ruptured epithelial cell

and VirA then mediate invasion of M cells and, later in the process, epithelial cells by triggering the actin rearrangements that ultimately lead to formation of a macropinocytic pocket. IpaB and IpaC are also together responsible for the membrane disruption that leads to vacuole escape, whilst the N-terminus of IpaB triggers the apoptotic signalling pathway in macrophages (Guichon *et al.*, 2001). (Reis & Horn, 2010) In addition to these key players in cellular disruption, the repertoire of *Shigella* effectors also includes the OspB-G and IpaH protein families that modulate host cell-signalling pathways and transcription, particularly to down-regulate the inflammatory response against the bacteria. For example, IpaH and OspG appear to

target the inflammatory signalling molecule NF-κB for degradation (Rohde *et al.*, 2007, Kim *et al.*, 2005).

Table 1.5 The early effectors secreted by the *Shigella* T3SS

Protein	General Function	Binding Partners
IpaB	Translocon pore Cell adhesion *Invasion:* Actin polymerisation (**1,6**) *Disruption of vacuole membrane* (**2**) *Macrophage apoptosis* (**4**)	IpaC CD44, fibronectin receptor Activated CD44 binds ezrin Caspase-1
IpaC	Translocon pore Cell adhesion *Invasion:* Actin polymerisation (**1,6**) *Disruption of vacuole membrane* (**2**)	IpaB Fibronectin receptor Cdc42
VirA	*Invasion:* Microtuble destabilisation resulting in RhoGTPase activation (**1,6**)	a/b-tubulin
IpaA	*Invasion:* F actin depolymerisation (**1,6**)	Vinculin
IpgD	*Invasion:* Phosphatase – activates PI-3 kinase/Akt pathway to promote cell survival (**1,6**)	$PI(4,5)P_2 \rightarrow PI(5)P$

Numbers shown in red correspond to stages in *Shigella* pathogenesis shown in **Figure 1.5**.

Therefore, the T3SS is the prime weapon used by *Shigella* spp. to cause Shigellosis and requires extensive structural and functional characterisation.

The Inner Membrane Ring of the *Shigella* T3SS

As mentioned previously, the IMR of the *S. flexneri* T3SS comprises two proteins: MxiG and MxiJ. EM reconstructions of NCs purified from *S. flexneri* and related *S. typhimurium* show the presence of two concentric rings in the IMR (Marlovits *et al.*, 2004, Hodgkinson *et al.*, 2009, Sani *et al.*, 2007, Schraidt & Marlovits, 2011), with the greater surface accessibility of PrgH lysine residues to biotinylation and acetylation suggesting MxiG homologues form an outer ring that encompasses an inner ring formed by MxiJ homologues (Yip *et al.*, 2005, Schraidt *et al.*,

2010). The stoichiometry of the IMR has been a controversial subject in the field, largely due to the low resolution of EM images and the bias from the rotational symmetry that must be imposed to obtain 3D reconstructions. Indeed, the most recent EM reconstructions of the *S. flexneri* NC did not allow distinction between 24-fold or pseudo 24-fold symmetry for the IMR (Hodgkinson *et al.*, 2009). Initial EM reconstructions of the *S. typhimurium* NC concluded a variable 19-22-fold symmetry for the IMR (Marlovits *et al.*, 2004), whilst the highest resolution 10 Å reconstruction with C3 symmetrisation suggested 24-fold symmetry (**Figure 1.3A**)(Schraidt & Marlovits, 2011). Indeed, a soluble portion of EscJ from EPEC crystallised as a superhelix with 24-fold symmetry, providing further evidence for 24-member rings in the IMR (Yip *et al.*, 2005). However, not all crystal contacts were maintained in the construction of a flat 24-fold ring model, which subsequently could not be reconciled with the IMR density in the *S. flexneri* EM reconstruction (Hodgkinson *et al.*, 2009), indicating further evidence of *S. flexneri* IMR stoichiometry is required.

MxiG is a 42 kDa protein with a 15 residue transmembrane region separating two soluble domains (Allaoui *et al.*, 1995). Ni-NTA nanogold labelling of the *S. typhimurium* homologue and subsequent EM visualisation of NCs indicated the 14.3 kDa N-terminal domain of MxiG (MxiG-N) is localised in the cytosol, whilst the 26.9 kDa C-terminal domain of MxiG (MxiG-C) is in the periplasm (**Figure 1.6**)(Schraidt *et al.*, 2010). MxiJ also has a single transmembrane region located towards the C-terminus of the protein (Allaoui *et al.*, 1992). Labelling of the PrgK C-terminus showed localisation in the cytoplasm, indicating the 22 kDa soluble domain of MxiJ resides in the periplasm (Schraidt *et al.*, 2010). In addition, MxiJ possesses a conserved lipobox sequence (LIGCEQR) within its N-terminal signal peptide, promoting lipidation and cleavage by signal peptidase II at the cysteine residue after transfer into the periplasm (Paetzel *et al.*, 2002). Although the ability to N-terminally sequence MxiJ isolated from NCs suggested acylation had proceeded no further than attachment of a single diacylglyceride (Tamano *et al.*, 2000, Blocker *et al.*, 2001), the requirement of the N-terminus for correct localisation of YscJ suggests MxiJ homologues are additionally anchored to the inner membrane via this N-terminal

lipid moiety (Silva-Herzog *et al.*, 2008)(**Figure 1.6**). As high resolution structures were not available for any portion of MxiG or MxiJ, they were attractive targets for further characterisation.

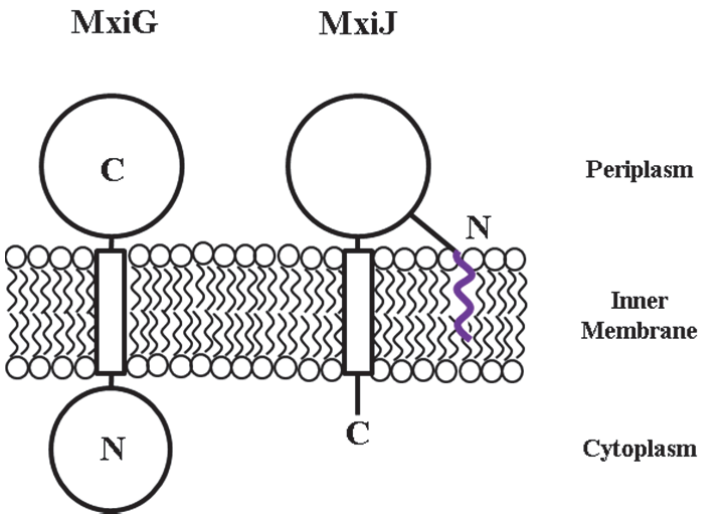


Figure 1.6 *The topology of MxiG and MxiJ with respect to the inner membrane*. Soluble domains are shown by circles, whilst transmembrane regions are shown by rectangles. The lipid moiety of MxiJ is shown in purple.

The Cytoplasmic Region of the IMR

Given that the cytoplasmic region of MxiJ is only ten residues long, MxiG-N is the predominant basal body structure within the cytoplasm and likely accounts for the density of the 'leg domains' visualised at the base of the *S. flexneri* NC (**Figure 1.7**)(Hodgkinson *et al.*, 2009).

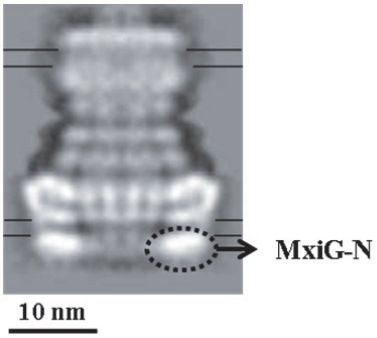


Figure 1.7 *The position of MxiG-N within the S. flexneri NC*. Reprojection from the C24 EM reconstruction taken from Hodgkinson et al., 2009. The lines show the hypothetical positions of the inner and outer membranes.

Within EM reconstructions of both *S. typhimurium* and *S. flexneri* NCs equivalent regions of density have shown positional variation (Marlovits *et al.*, 2004, Hodgkinson *et al.*, 2009, Schraidt & Marlovits, 2011) and Ni-NTA nanogold attached to the PrgH N-terminus showed a diffuse localisation (Schraidt *et al.*, 2010), implying conformational flexibility of MxiG-N. Specifically this density was observed to undergo a clamping movement

when EM images of *Salmonella* basal bodies and NCs were compared (Marlovits *et al.*, 2004). Therefore, MxiG-N speculatively may contribute to substrate-specificity switching by the apparatus and could have a role in communicating regulatory signals from the cytoplasm to the rest of the NC or vice versa.

Alternatively, MxiG-N could function to structurally stabilise the IMR through self-association. However, it was unclear from current EM reconstructions whether MxiG-N domains were in close enough proximity to oligomerise, as some indicated a continuous ring of density at the base of the NC (Marlovits *et al.*, 2004, Hodgkinson *et al.*, 2009) and others indicated a discontinuous ring (Sani *et al.*, 2007).

In *Y. enterocolitica* YscD was required for co-purification of YscJ with the OMR ring component (Diepold *et al.*, 2010), suggesting the MxiJ ring is associated with basal bodies via an interaction with the MxiG ring. Indeed, the formation of IMR-like structures by recombinantly overexpressed PrgH and PrgK and cross-linking of these proteins within purified NCs suggested the *S. typhimurium* MxiG and MxiJ homologues were closely associated (Schraidt *et al.*, 2010, Sanowar *et al.*, 2010, Kimbrough & Miller, 2000). Interestingly, the observation of cross-links between the cytoplasmic regions of PrgH and PrgK (Schraidt *et al.*, 2010) could implicate MxiG-N in IMR assembly and stabilisation via an interaction with the extreme C-terminus of MxiJ.

As discussed previously, the IMR could also act as a scaffold for the association of the remaining components of the secretion-competent T3SS. In particular, MxiG-N could interact with cytoplasmic components to mediate their assembly, with the identification of Spa33, MxiN and Spa47 as potential binding partners from pull-down assays in *S. flexneri* and *Chlamydophila pneumoniae* (Morita-Ishihara *et al.*, 2006, Johnson *et al.*, 2008, Stone *et al.*, 2008), although these putative interactions have subsequently remained relatively unexplored. In particular, a function for MxiG-N in localising the putative C-ring to the base of the *S. flexneri* NC via an

interaction with Spa33 would be plausible, given the FliG-mediated interaction between the IMR and C-ring found within the flagellar T3SS (Paul *et al.*, 2011, Levenson *et al.*, 2012).

Although the MxiG-N homologue in *Yersinia pestis* was shown to be essential for protein secretion by the T3SS (Ross & Plano, 2011), potential roles for MxiG-N in cell-signalling, IMR stabilisation and T3SS assembly could only be postulated. As the MxiG-N sequence gave no clue about the domain function, structural homology had to be relied upon to gain an insight. There were no published high-resolution structures available for any domain homologous to MxiG-N (**Table 1.3**), identifying it as an important structural target.

The Periplasmic Domains of the IMR

The periplasmic region of the *S. flexneri* basal body comprises significant soluble domains from both MxiG and MxiJ in the IMR (**Figure 1.6**) and the N-terminus of MxiD in the OMR (Allaoui *et al.*, 1993). Comparison of EM reconstructions obtained from wild-type and *invG* ¯ *S. typhimurium* strains clearly demonstrated the relative positions of the IMR and OMR density within the NC, showing that the InvG secretin reaches deep into the periplasm to contact the IMR (**Figure 1.8A**)(Schraidt *et al.*, 2010). Indeed, an interaction between YscC and the extreme C-terminus of YscD was observed in *Y. pestis* (Ross & Plano, 2011), whilst loss of the four C-terminal residues of PrgH destabilised the interaction between the IMR and OMR (Schraidt *et al.*, 2010). This suggests a direct interaction between MxiG and MxiD that could facilitate the OMR-mediated assembly of the IMR observed in *Y. enterocolitica* (Diepold *et al.*, 2010).

High resolution structures are available for homologues of the MxiG, MxiJ and MxiD periplasmic domains from *S. typhimurium* PrgH (Spreter *et al.*, 2009), EPEC EscJ (Crepin *et al.*, 2005, Yip *et al.*, 2005) and EscC (Spreter *et al.*, 2009) respectively (**Figure 1.3B**). Each periplasmic domain has a modular architecture and comprises 2-3 small α/β sub-domains that exhibit striking similarity, leading to the proposal of a wedge-shaped 'ring-building' motif with two α-helices packed against a β-sheet (Spreter *et al.*, 2009). Indeed, crystal contacts within the $EscJ_{21-190}$ superhelical assembly were mediated by this structurally conserved motif (Yip *et al.*,

2005). Furthermore, homology models based on the structures of these periplasmic domains could be precisely docked into the periplasmic region of the 10 Å EM reconstruction of the *S. typhimurium* NC, providing a precise indication of the relative positions of each domain within the NC that was corroborated by knowledge of surface-accessible regions of PrgH and a destabilising mutation at a putative interaction interface within the PrgH ring (Schraidt & Marlovits, 2011)(**Figure 1.8B**).

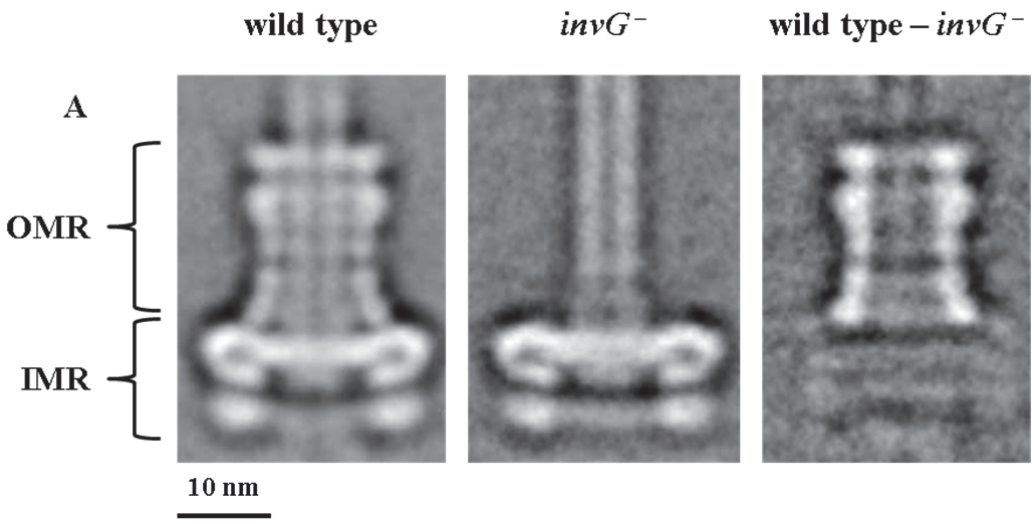


Figure 1.8 *The periplasmic region of the T3SS NC.*

A EM class averages of NCs purified from wild-type and Δ*invG S. typhimurium* and their density difference taken from Schraidt et al., 2010. The implicated positions of the IMR and OMR are shown.

B Docking of homology models of PrgK (yellow) and InvG (orange) and the crystal structure of PrgH residues 170-362 (grey)(Spreter *et al.*, 2009) into the 10 Å C3 EM reconstruction of the *S. typhimurium* NC taken from Schraidt & Marlovits, 2011. PrgK and InvG homology models are based on EPEC EscJ residues 21-190 (Yip *et al.*, 2005) and EscC residues 21-174 (Spreter *et al.*, 2009) respectively.

However, this docking only provides us with a model for association of PrgH and PrgK periplasmic domains in the IMR and requires supporting experimental evidence to confirm the molecular nature of the interaction interfaces. Although, additional characterisation of this PrgH-PrgK interface has been obtained via cross-linking, all the PrgK lysine residues implicated

were found within the structurally uncharacterised C-terminus of the domain (Sanowar *et al.*, 2010). Therefore, there is a need for high resolution structural information on their interaction interface.

In addition, as each periplasmic domain only has one representative high resolution structure, it is difficult to take species-specific variation in NC arrangement or experimental artefacts in the determined structures into account during docking into EM reconstructions. Indeed, docking of the crystal structure of the EscJ periplasmic domain (Yip *et al.*, 2005) into the *S. flexneri* EM reconstruction resulted in three diverse positions of the domain (Hodgkinson *et al.*, 2009), indicating that the currently available NC EM reconstruction and representative structures are insufficient to accurately predict the location of MxiJ and MxiG periplasmic domains within the IMR. Therefore, high resolution structural information on MxiJ and MxiG periplasmic domains and their sub-complexes would be optimally required to predict the molecular nature of their interaction within the *S. flexneri* IMR.

The C-ring of the *Shigella* T3SS

The putative C-ring of translocation-associated T3SSs has never been visualised in the highest resolution EM reconstructions, due to detachment from the IMR during NC purification. However, early visualisations of the *S. flexneri* NC within inner membrane ghosts produced from osmotically-shocked cells revealed a prominent cytoplasmic bulb with a ~45 nm diameter, suggesting the T3SS has a significant cytoplasmic structure at its base (**Figure 1.9A**)(Blocker *et al.*, 1999). Indeed, subsequent *in situ* immuno-gold labelling of the predicted C-ring component Spa33 demonstrated that it was localised at the base of the T3SS (Morita-Ishihara *et al.*, 2006). Furthermore, subcellular fractionation and blue native polyacrylamide gel electrophoresis (PAGE) analysis of *S. typhimurium* SpaO showed the protein forms large heterogeneous molecular weight complexes (Lara-Tejero *et al.*, 2011). Therefore, the localisation and oligomerisation of Spa33 homologues together indicate the probable formation of a C-ring structure *in vivo*.

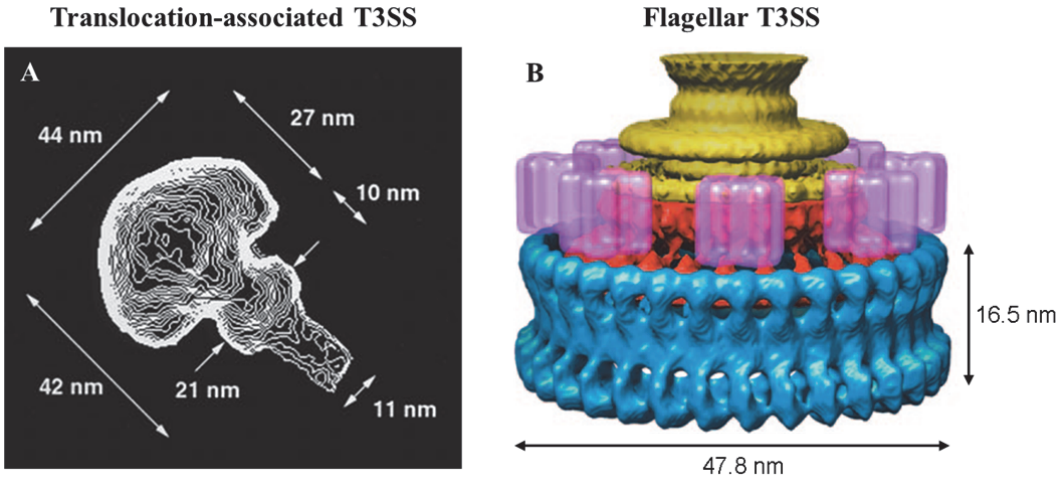


Figure 1.9 *EM characterisation of the C-rings from translocation-associated and flagellar T3SSs*.

A Average EM image of the *S. flexneri* T3SS at 28 Å resolution obtained from the complete membrane ghosts of osmotically-shocked and negatively stained bacteria showing the prominent cytoplasmic bulb. Taken from Blocker et al., 1999.

B Cryo-EM reconstruction of the flagellar rotor from a clockwise-locked mutant of *S. typhimurium* showing the 34-fold symmetry averaged C-ring (blue), the 25-fold symmetry averaged M-ring comprising FliG (red) and FliF (yellow) and a model for the MotA/MotB stator position (pink). Taken from Thomas et al., 2006.

The absence of a needle in complexes isolated from a *spa33*$^{-}$ strain suggested the C-ring was essential for secretion (Morita-Ishihara *et al.*, 2006), with subsequent interaction studies suggesting Spa33 is involved in a complicated interaction network at the base of the T3SS (**Table 1.6**). In addition to a potential localising interaction with the IMR mentioned previously, Spa33 homologues from most species have been consistently found to interact with homologues of the Spa47 ATPase and its regulators, MxiN and MxiK, suggesting a central role in mediating formation of the cytoplasmic assembly at the base of the T3SS. In addition, several studies have indicated a greater promiscuity in the interactions of Spa33 that are not observed in every species and with all binding assays. In particular, a couple of recent analyses in *S. typhimurium* and *Chlamydia trachomatis* have suggested an interaction between Spa33 homologues and the chaperones of translocon and effector proteins, leading to the suggestion that the C-ring of translocation-associated T3SSs may act as a 'sorting platform' in establishing the correct secretion hierarchy (Spaeth *et al.*, 2009, Lara-Tejero *et al.*, 2011). However, these experiments

Table 1.6 The interaction network of Spa33 homologues

Species & Spa33 homologue	Interaction Partners	Method	Reference
S. flexneri Spa33	**IMR:** MxiG, MxiJ **Cytoplasmic regulators:** MxiN, MxiK, Spa47, Spa32 **Effectors**: VirA, IcsB, IpaC, IpgB1	Pull down assays, co-purification, yeast two-hybrid	(Morita-Ishihara *et al.*, 2006, Johnson & Blocker, 2008, Jouihri *et al.*, 2003)
S. typhimurium SpaO	**Cytoplasmic regulators:** OrgA, OrgB, InvC **Chaperones:** SicA, InvE, SicP	Blue native PAGE of whole cell lysate, no effector binding in absence of its chaperone(!)	(Lara-Tejero *et al.*, 2011)
Y. pestis YscQ	**Cytoplasmic regulators:** YscL, YscK	Yeast two-hybrid	(Jackson & Plano, 2000)
C. pneumoniae CdsQ	**IMR:** CdsD **Cytoplasmic regulators:** CdsL, CdsN	Pull-down assays	(Johnson *et al.*, 2008, Stone *et al.*, 2008)
C. trachomatis	**Cytoplasmic regulators:** CdsL **Export Apparatus:** CdsS, CdsT **Chaperones:** Ct260 **Other:** Ct677 (ribosome recycling factor), Ct824 (protease), Ct560, Ct567, Ct085, Ct676	Yeast two-hybrid	(Spaeth *et al.*, 2009)
EPEC EscQ	**Cytoplasmic regulators:** EscL, EscN	Co-purification	(Biemans-Oldehinkel *et al.*, 2011)
Flagellar FliM & FliN	**Cytoplasmic regulators:** FliH, FliG **Chemotaxis pathway:** CheY-P	Co-crystallisation, NMR titration, pull-down assays, co-purification,	(Brown *et al.*, 2007, Paul *et al.*, 2011, Paul *et al.*, 2006, McMurry *et al.*, 2006, Minamino *et al.*, 2009, González-Pedrajo *et al.*, 2006, Sarkar *et al.*, 2010a, Dyer *et al.*, 2009, Lee *et al.*, 2001a)

Homologues of *S. flexneri* MxiN, MxiK and Spa47 are shown in red, blue and green respectively.

are far from conclusive, with non-specific binding, weaknesses in methodology and subjective interpretation potentially biasing the results, indicating further work is required to ascertain the function of Spa33 in the *S. flexneri* T3SS.

Some insight into the structure and function of the *S. flexneri* C-ring could be gained from the flagellar C-ring, which is comparatively better characterised. Indeed, the C-ring remains associated with hook-basal bodies during their purification from *S. typhimurium* (Francis *et al.*, 1994), enabling cryo-EM to be used to provide a model for the intact structure (**Figure 1.9B, blue**)(Thomas *et al.*, 2006). This flagellar C-ring model corroborates the approximate diameter for the *S. flexneri* C-ring suggested by negative stain EM (**Figure 1.9A**) and reveals ~34 fold symmetry. In addition, the flagellar C-ring is required for assembly of the hook and flagellum (Yamaguchi *et al.*, 1986, Vogler *et al.*, 1991) and has been shown to interact with the MxiN homologue FliH (Paul *et al.*, 2006, Minamino *et al.*, 2009), suggesting a similar role to the *S. flexneri* C-ring in the T3SS. Interestingly, the interaction between the flagellar C-ring and ATPase is mediated by FliH (McMurry *et al.*, 2006), suggesting observed interactions between Spa33 and Spa47 homologues could be indirect (**Table 1.6**).

However, there are significant structural and functional differences between the flagellar and *S. flexneri* C-ring, which suggests they may not be directly comparable. Whilst Spa33 was presumed to be the sole C-ring component in *S. flexneri*, the flagellar C-ring comprises FliM and FliN proteins. Although both sequence homologues of Spa33, only FliM is an equivalent size, with FliN encompassing the C-terminal third of the protein sequence (Hueck, 1998). FliM and FliN have been found to form a FliM-$(FliN)_4$ complex in solution (Brown *et al.*, 2005), with the current $(FliN)_4$-FliM-$(FliN)_4$ model based on cross-linking and mutagenesis data suggesting the flagellar C-ring forms through each FliM molecule bridging two FliN tetramers (Sarkar *et al.*, 2010b). Therefore, it was unclear how a single protein in the *S. flexneri* T3SS could structurally and functionally perform the same role as two proteins in the flagellar T3SS. Although high resolution structures of portions of *Thermotoga maritima* FliN (Brown *et al.*, 2005) and homologous *P. syringae* $HrcQ_B$ (Fadouloglou *et al.*, 2004) show that the domains

form highly intertwined dimers, it was unknown whether the homologous C-terminal domain of Spa33 (Spa33-C) would adopt the same conformation in the context of a C-ring with a single protein component. Together with FliG, the flagellar C-ring also forms the switch complex responsible for torque generation and direction switching in response to CheY-P binding during flagellum rotation (Stock *et al.*, 2012), a function presumably not required of the *S. flexneri* C-ring.

Although an intact C-ring can be isolated and studied *in vitro* within the hook-basal body of the flagellar T3SS, recent fluorescence recovery after photobleaching (FRAP) studies suggest that tagged FliM and FliN both turnover on timescales of tens of minutes *in vivo* (Fukuoka *et al.*, 2010, Li & Sourjik, 2011). Therefore, the *S. flexneri* C-ring could also be a dynamic structure to the extent that it precludes *in vitro* isolation and characterisation. However, within the *E. coli* C-ring, a discrete population of ~20 FliM molecules was shown to undergo turnover on a timescale of ~40 s in the presence of CheY-P, suggesting that this dynamic population could be involved in motor switching and therefore may not exist in the putative *S. flexneri* C-ring (Delalez *et al.*, 2010).

Therefore, many aspects of the structure and function of the *S. flexneri* C-ring cannot be assumed or predicted from knowledge of the flagellar C-ring. Indeed, equivalent characterisation of the two systems would be required to determine the extent to which their different functions have resulted in a diversification of C-ring structure. As there were no high resolution structures for Spa33 homologues from any of the model animal-pathogenic T3SSs, structural characterisation of Spa33 and its intermolecular interactions had the potential to provide much insight into C-ring assembly and structure.

1.4 Thesis Outline

The cytoplasmic and periplasmic regions of the *S. flexneri* IMR and its associated C-ring were chosen as targets for further structural characterisation using the three main approaches outlined in **Figure 1.3**. This study could be divided into six broad aims, shown with respect to their relative chapter in this thesis:

- To determine the high resolution structure of MxiG-N and understand the position of the domain within the context of the intact NC – *Chapter 2*;
- To explore potential functions for MxiG-N, potentially based on structural homology, and identify likely interaction partners for this cytoplasmic domain at the base of the NC – *Chapter 3*;
- To establish protocols for purification of the periplasmic domains of MxiG and MxiJ, with a view to high resolution characterisation of their structures and interaction interface – *Chapter 4*;
- To establish a protocol for purification of the T3SS basal body, with a view to high resolution structural characterisation of the IMR and its interaction with the OMR – *Chapter 5*;
- To determine the high resolution structure of Spa33-C and characterise interactions with its putative binding partners – *Chapter 6*;
- To establish a protocol for purification of full-length Spa33, with a view to understanding its intermolecular interactions within the context of the *S. flexneri* C-ring – *Chapter 7*.

2. Structural characterisation of MxiG-N

As outlined previously, multiple copies of MxiG are present within the IMR of the *Shigella* basal body. The protein is predicted to have two soluble domains separated by a transmembrane region, with MxiG-N residing in the cytosol (Allaoui *et al.*, 1995, Schraidt *et al.*, 2010). As the domain exhibits conformational flexibility and is the predominant basal body structure in the cytoplasm, a role for MxiG-N in communicating signals to the rest of the T3SS was postulated (Marlovits *et al.*, 2004, Hodgkinson *et al.*, 2009, Schraidt *et al.*, 2010). Similarly, MxiG-N could be acting to structurally stabilise the IMR or mediate T3SS assembly. As there were no published structures for any domain homologous to MxiG-N, a high resolution structure had the potential to provide a general insight into its role in the T3SS from structural homology. Therefore, the solution structure of MxiG-N was determined, as outlined in this chapter.

2.1 Expression and Purification of MxiG-N

Expression of MxiG-N_{1-126}

In order to achieve the high yield required for structural characterisation of MxiG-N from *S. flexneri*, residues 1-126 of MxiG-N with a MGSSHHHHHHSSGLVPRGSH N-terminal tag (MxiG-N_{1-126}) were recombinantly overexpressed in and subsequently purified from *E. coli*, using protocols established by Dr Steven Johnson. Firstly, pET14b-*mxiG*N_{1-126} (donated by S. Johnson) was transformed into *E. coli* BL21 (DE3) pLysS by the heat-shock protocol (**Appendix 8.2.1**). A resuspension of BL21 (DE3) pLysS pET14b-*mxiG*N_{1-126} colonies was used to inoculate 2 x 1 l Luria Bertani (LB) medium (Fisher) supplemented with 100 µg/ml ampicillin, and the resulting cultures were grown in a shaking incubator at 37^{o}C until an optical density at 600 nm (OD_{600}) of 0.6 was reached. Protein expression was then induced overnight at 21^{o}C with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested by

centrifugation at 4000 x g for 30 min at 4°C, washed once with phosphate buffered saline (PBS; Oxoid) and stored at -20°C.

Overexpression of MxiG-N_{1-126} was verified by non-reducing sodium dodecyl sulphate (SDS) - PAGE (**Appendix 8.2.2**); intense bands around the expected molecular weight for MxiG-N_{1-126} of 16.5 kDa, clearly demonstrated overexpression of the protein after induction with IPTG, although some leaky expression had also occurred prior to induction (**Figure 2.1**).

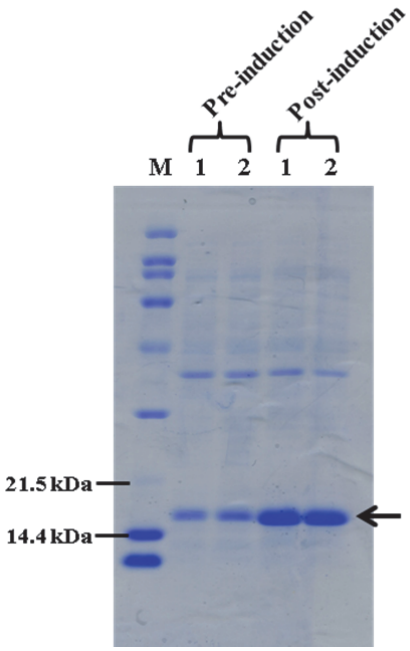


Figure 2.1 *15% SDS-PA gel showing over-expression of MxiG-N_{1-126} in BL21 (DE3) pLysS.*

Samples taken from repeat cultures 1 and 2 pre- and post-induction with 0.4 mM IPTG are labelled. The arrow indicates the band corresponding to MxiG-N_{1-126}. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

Purification of MxiG-N_{1-126}

BL21 (DE3) pLysS pET14b-*mxiGN*$_{1-126}$ cell pellets were resuspended in 40 ml lysis buffer (50 mM Tris(hydroxymethyl)aminomethane (Tris)-HCl (pH 7.5), 500 mM NaCl) with an ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche) added to limit proteolysis of MxiG-N_{1-126} during the initial stages of purification. Cells were lysed by passing them through an Emulsiflex-C5 Homogeniser (GC Technologies) five times at 15,000 p.s.i.. The resulting lysate was centrifuged at 26,000 x g for 30 min at 4°C to pellet unbroken cells,

intact organelles, inclusion bodies and other cell debris. Samples of both the supernatant and pellet (resuspended in an equivalent volume of lysis buffer using a hand-held homogeniser) were analysed by SDS-PAGE (**Figure 2.2A**), confirming that the majority of MxiG-N_{1-126} was present in the soluble cell fraction.

The N-terminal His-tag of MxiG-N_{1-126} allowed purification to proceed via Ni-affinity chromatography. MxiG-N_{1-126} was loaded on to a 5 ml Ni-NTA Superflow Cartridge (Qiagen) by flowing the supernatant from the preceding centrifugation step through at 2.5 ml/min with a peristaltic pump (GE Healthcare). The protein was subsequently washed with 10 column volumes (CV) lysis buffer supplemented with 50 mM imidazole pH 7.5, before elution of 5 x 5 ml fractions with 300 mM imidazole. The 15 % SDS-PA gel in **Figure 2.2A** demonstrates that the majority of the protein eluted in fractions 1 & 2, with only a small amount lost during cartridge loading and washing.

Further purification and buffer exchange was performed by size exclusion chromatography (SEC). A 10 ml sample comprising fractions 1 & 2 was injected onto a HiLoad 26/60 Superdex 75 pg (GE Healthcare) column attached to an ÄKTApurifier (GE Healthcare) system and passed through the matrix at 3 ml/min with 25 mM Na_2HPO_4 (pH 6.8). Although SDS-PAGE analysis (**Figure 2.2A**) shows that the fractions comprising the major elution peak (**Figure 2.2B**) contained MxiG-N_{1-126}, there also appeared to be a significant amount of a 28-36 kDa contaminant present. However comparison of samples on a 15 % SDS-PA gel with and without 10 mM dithiothreitol (DTT) added to the 1 x loading dye, showed that this contaminant could be reduced to MxiG-N_{1-126} (**Figure 2.2C**). Therefore, the contaminant was likely to be a disulphide-bonded dimer of MxiG-N_{1-126} formed via the single cysteine in the protein sequence and could be removed by including a reducing agent such as DTT in the purification buffers.

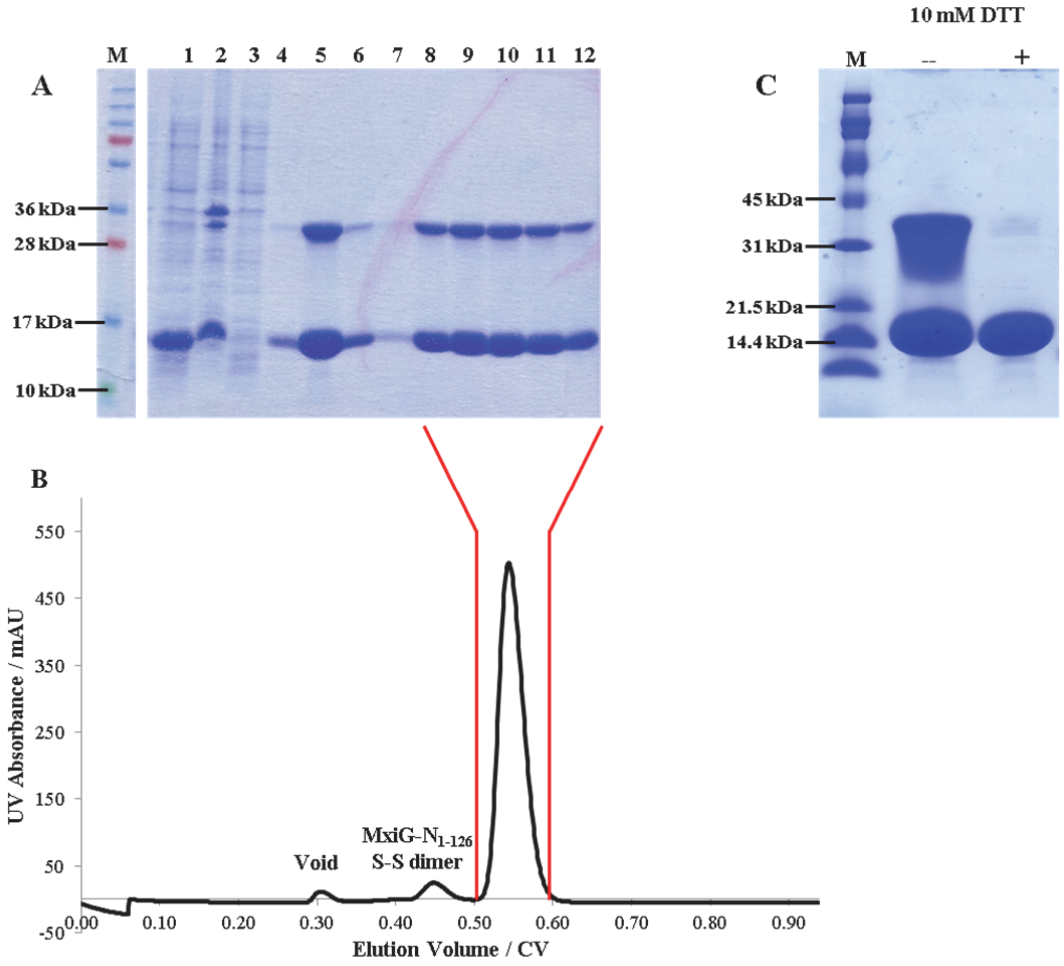


Figure 2.2 *Purification of MxiG-N_{1-126}.*

A 15% SDS-PA gel following the purification of MxiG-N_{1-126}. M: PageRuler Plus Prestained Protein Ladder (Fermentas), 1: Supernatant after 26,000 x g centrifugation, 2: Pellet after 26,000 x g centrifugation, 3: Flow-through from loading Ni-NTA cartridge, 4: Flow-through from 50 mM wash of Ni-cartridge, 5-7: Fractions 1-3 from 300 mM elution of Ni-cartridge and 8-12: Fractions encompassing the major elution peak.

B Elution profile for MxiG-N_{1-126} from a HiLoad 26/60 prep grade column. Elution with 25 mM Na_2HPO_4 pH 6.8. Major elution peak after 0.55 CV. The identities of other minor peaks are shown.

C 15% SDS-PA gel comparing non-reduced and reduced samples of MxiG-N_{1-126}. Either no DTT (--) or 10 mM DTT (+) added to the 1 x loading dye prior to sample boiling. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

Although preliminary crystallisation trials carried out by Dr Steven Johnson had been unsuccessful, a natural abundance $^{1}H,^{15}N$-heteronuclear single quantum coherence (HSQC) spectrum collected for 3 mM MxiG-N_{1-126} (McDowell *et al.*, 2011) (**Figure 2.3**) showed good peak dispersion in both dimensions, indicating the protein was not aggregated and had well-defined secondary and tertiary structure amenable to characterisation with NMR spectroscopy. However, the spectrum was not ideal for assignment due to peak overlap (**Figure 2.3, dashed box**), so further construct optimisation was attempted.

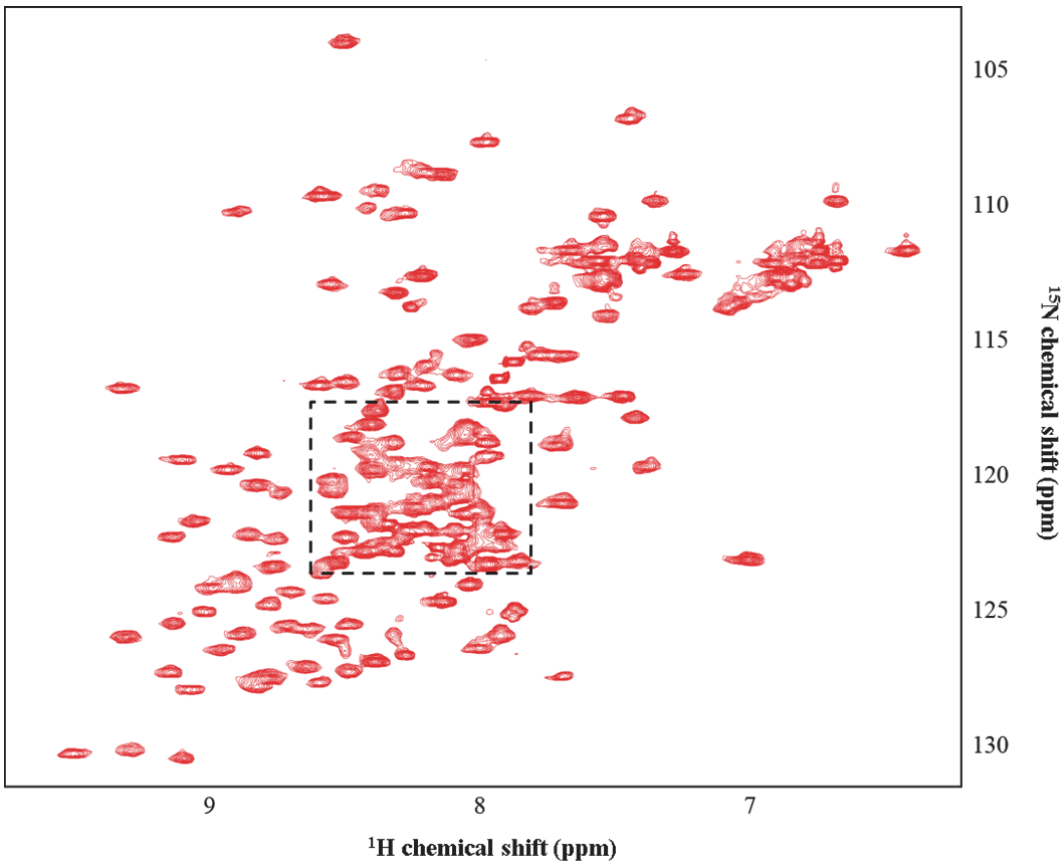


Figure 2.3 *Natural abundance $^{1}H,^{15}N$-HSQC spectrum of 3 mM MxiG-N_{1-126}.*

The dashed box indicates the region of most significant peak overlap. Taken from McDowell et al., 2011.

Limited Proteolysis of MxiG-N_{1-126}

In order to obtain a core structure for the protein, MxiG-N_{1-126} was subjected to limited proteolysis with trypsin (from porcine pancreas, Sigma). After Ni-affinity purification of MxiG-N_{1-126}, fractions 1 & 2 were dialysed against 4 x 1 l 50 mM Tris-HCl (pH 7.5), 150 mM NaCl in 3500 MWCO SnakeSkin Dialysis Tubing (Thermo scientific). Protein concentration was measured with a NanoDrop ND1000 Spectrophotometer using the absorbance at 280 nm (A_{280}). The enzyme was then added at a w/w protein:trypsin ratio of 250:1 for 2 h at room temperature, a protocol previously optimised by Dr Steven Johnson. **Figure 2.4** indicates that the trypsin digest ran to completion and led to the formation of a shorter construct. SEC was then carried out as previously, with 1 mM DTT added to the running buffer. There were no changes to the elution profile shown in **Figure 2.2B**. Furthermore, the reducing agent prevented the formation of a disulphide-bonded dimer and the construct was purified to homogeneity (**Figure 2.4**).

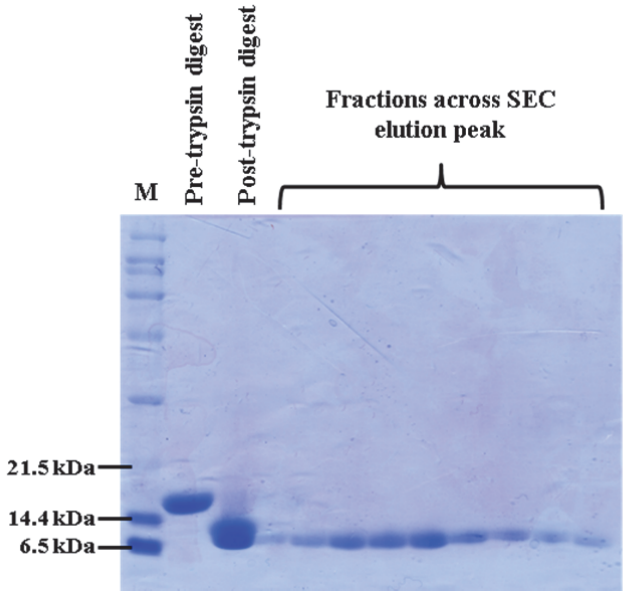


Figure 2.4 *15% SDS-PA gel showing limited proteolysis and purification of MxiG-N_{1-125}.*

Pre-trypsin digest sample obtained from fraction 2 from Ni-affinity purification and post-trypsin digest sample taken 2 h after incubation of MxiG-N_{1-125} with trypsin. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

SEC fractions from across the major elution peak were pooled and concentrated using a Vivaspin 20 5 kDa molecular mass cut-off centrifugal device (GE Healthcare). A 20-100 µM sample was submitted to the Biophysical Instrument Facility (Department of Biochemistry, University of Oxford) for mass spectrometry using an Electrospray Time Of Flight Mass

Spectrometer (ES TOF-MS). The resulting mass of 12.17 kDa indicated that both the N-terminal His-tag and C-terminus had been removed to leave a construct comprising residues 6 to 112 of MxiG-N (MxiG-N_{6-112}; **Figure 2.5**). The yield of MxiG-N_{6-112} was 18.5 mg per 1 l BL21 (DE3) pLysS pET14b-*mxiGN*$_{1-126}$, assuming a molar absorptivity coefficient of 5960 $M^{-1}cm^{-1}$ as predicted by the ExPASy ProtParam server.

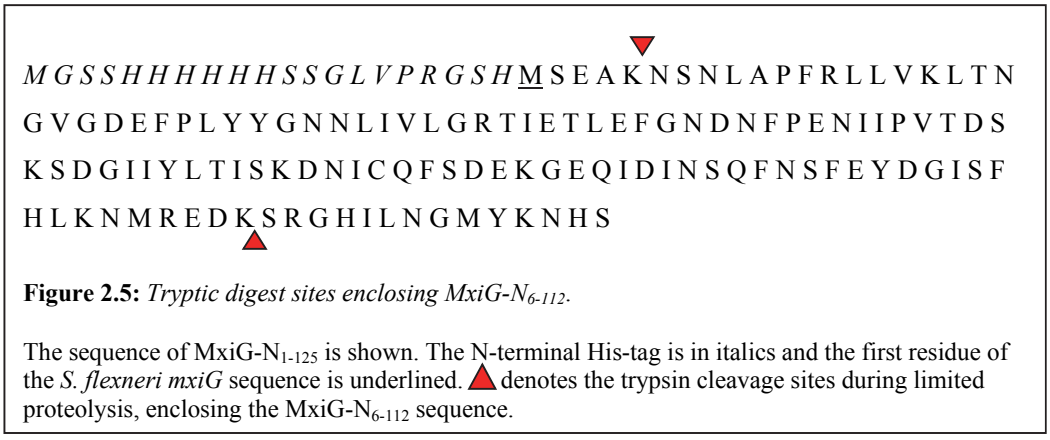


Figure 2.5: *Tryptic digest sites enclosing MxiG-N*$_{6-112}$.

The sequence of MxiG-N_{1-125} is shown. The N-terminal His-tag is in italics and the first residue of the *S. flexneri mxiG* sequence is underlined. ▲ denotes the trypsin cleavage sites during limited proteolysis, enclosing the MxiG-N_{6-112} sequence.

Expression and Purification of $^{13}C/^{15}N$-MxiG-N_{6-112}

Natural abundance levels of ^{13}C and ^{15}N are insufficient for the majority of NMR experiments utilised in triple resonance assignment and structure determination, therefore isotopic enrichment of MxiG-N_{6-112} was required to provide an appropriate sample for NMR spectroscopy. This was achieved by ensuring ^{13}C-glucose and $^{15}NH_4Cl$ were the only carbon and nitrogen sources available during expression of MxiG-N_{1-126}. Subsequent to the transformation step, a resuspension of BL21 (DE3) pLysS pET14b-*mxiGN*$_{1-126}$ colonies was used to inoculate 2 x 100 ml LB with 100 µg/ml ampicillin added. These cultures were grown at 37°C until an OD_{600} of ~0.6, harvested by centrifugation and washed once with unlabelled minimal medium (**Table 2.1**). Each cell pellet was then used to inoculate 2 x 1 l $^{13}C/^{15}N$-labelled minimal medium (**Table 2.1**) and growth resumed at 37°C until an OD_{600} of ~0.5 was reached. Protein expression was then induced as for unlabelled MxiG-N_{1-125}.

Table 2.1 Composition of minimal media

Medium	Composition	Concentration
Unlabelled minimal medium	1x salt solution: Na_2HPO_4	42 mM
	KH_2PO_4	22 mM
	NaCl	8 mM
	Trace elements (pH 7): EDTA	270 µM
	$FeCl_3$	48 µM
	$ZnCl_2$	3 µM
	$CuCl_2$	1 µM
	$CoCl_2$	1 µM
	H_3BO_3	2.5 µM
	Yeast Nitrogen Base (without amino acids and ammonium sulphate)	1.5 g/l
	Glucose	0.5%
	Thiamine	1 mM
	NH_4Cl	1 µg/ml
	$MgSO_4$	1 mM
	CaCl	0.1 mM
	Ampicillin	100 µg/ml
$^{13}C/^{15}N$-labelled minimal medium	As for normal minimal medium except NH_4Cl and glucose replaced by:	
	$^{15}NH_4Cl$	1 µg/ml
	^{13}C-glucose	0.2%

Purification of $^{13}C/^{15}N$-MxiG-N_{6-112} then proceeded as for the unlabelled protein. Mass spectrometry analysis indicated a 12.78 kDa species had been produced. Given that the molecular mass would be 12.85 kDa for a 100% efficiently labelled sample, combined labelling with ^{15}N and ^{13}C was 89% efficient, as would be expected given that leaky expression (**Figure 2.1**) occurred in unlabelled LB medium. Furthermore, yields of between 11.5 and 20.4 mg per 1 l BL21 (DE3) pLysS pET14b-*mxiG*N_{1-126} were achieved for a series of separate purifications, indicating protein expression was not compromised by growth in minimal medium. Typically a 500 µl sample of $^{13}C/^{15}N$-MxiG-N supplemented with 5% (v/v) D_2O could be produced at concentrations up to 6 mM for NMR experiments.

2.2 Assignment of MxiG-N_{6-112}

Unless otherwise stated, all NMR experiments were performed on a Bruker Avance II 500 MHz spectrometer at 25^{o}C under the supervision of Prof. James McDonnell. Details of the sample, experimental parameters and nuclear correlations shown are summarised in **Appendix 8.3** for each spectrum collected in this thesis. All spectra were processed within the Bruker TopSpin software and subsequently analysed with Sparky (Goddard & Kneller), as presented here.

The ^{1}H,^{15}N-HSQC for ^{13}C/^{15}N-MxiG-N_{6-112} showed good peak dispersion in both dimensions and agreed well with peak positions from the MxiG-N_{1-125} spectrum, indicating removal of the termini did not affect the core structure of the protein. Furthermore, the construct formed by limited proteolysis had significantly reduced peak overlap in the central region, resulting in a spectrum with clearly resolved and assignable peaks. (**Figure 2.6A**) Initially 101 of the expected 102 non-proline backbone amide peaks and 29 of the expected 30 asparagine/glutamine side-chain amide peaks could be picked. In addition, a ^{1}H,^{15}N-HSQC with a sweepwidth in the ^{15}N dimension of 60 ppm instead of 32 ppm allowed peaks which were previously folded into the spectrum to be identified, notably the side-chain N^{ε}-H^{ε} for 2/3 arginine residues. (**Figure 2.6B**)

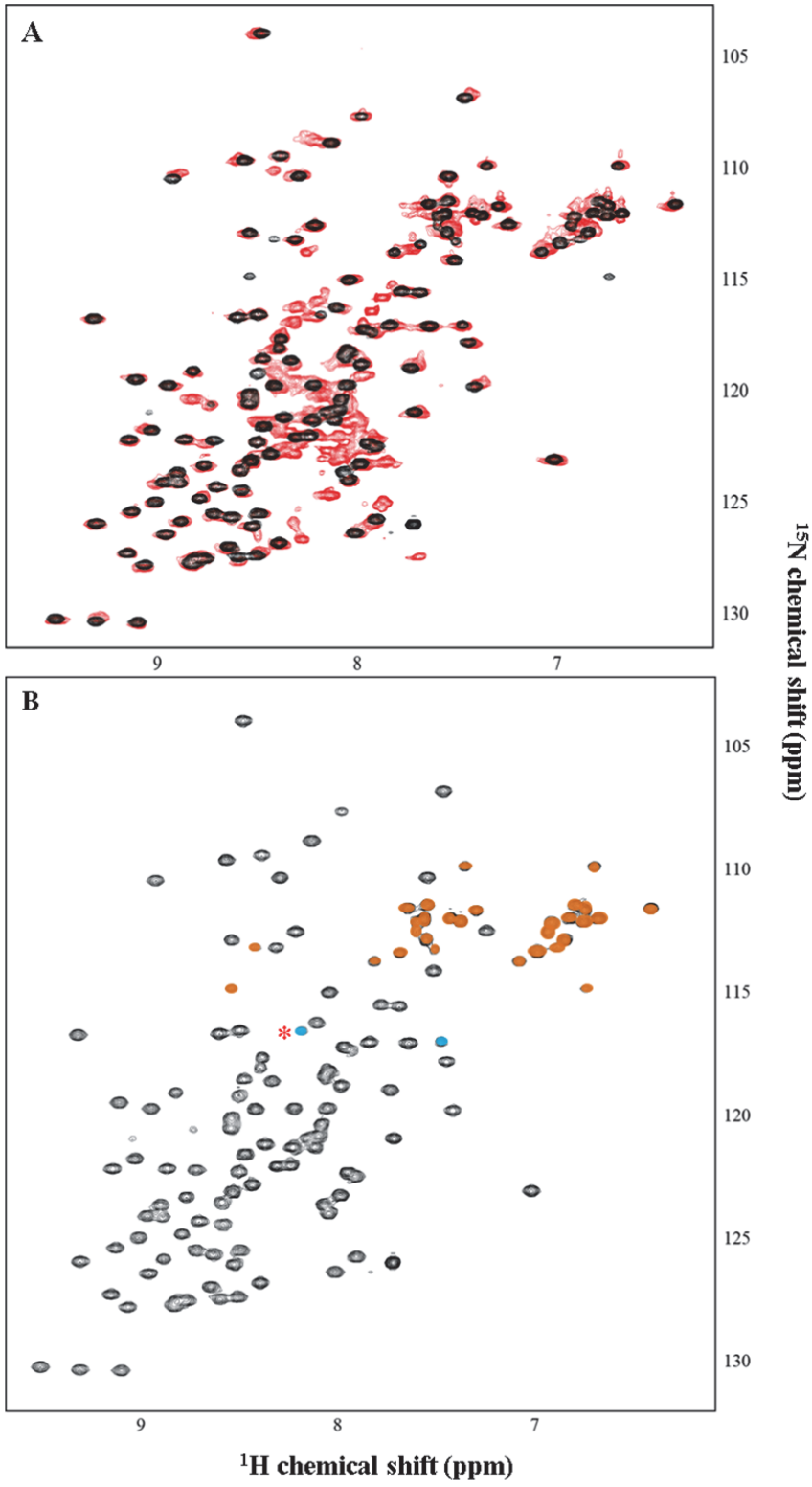


Figure 2.6 *$^{1}H,^{15}N$-HSQC spectra for $^{13}C/^{15}N$-MxiG-N_{6-112}*

A Overlay of a natural abundance $^{1}H,^{15}N$-HSQC spectrum for 3 mM MxiG-N_{1-125} (red) and a spectrum for 2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} (black).

B The $^{1}H,^{15}N$-HSQC spectrum for 2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} discriminating peaks for backbone amides (black), side-chain N^{ε}-H^{ε} arginine amides (blue) and side-chain N^{δ}-H^{δ} or N^{ε}-H^{ε} asparagine or glutamine amides (orange). The N^{ε}-H^{ε} arginine amide peak marked with a red asterisk coincides exactly with a backbone amide peak.

Backbone Assignment by the Triple Resonance Approach

Triple resonance experiments were carried out that separate each amide signal from the 2D $^{1}H,^{15}N$-HSQC spectrum into a third ^{13}C dimension. Comparison of spectra of increasing complexity allowed peaks corresponding to the backbone C^{α} and side-chain carbons of residue i-1 to be identified and distinguished for each backbone amide (**Figure 2.7**). The HN(CO)CA experiment (Bax & Ikura, 1991) correlates N-H^{N} with C^{α}(i-1) only, whilst a CBCA(CO)NH experiment (Grzesiek & Bax, 1992a) gives peaks for both C^{α}(i-1) and C^{β}(i-1). Comparison of equivalent strips from these spectra allow C^{α}(i-1) and C^{β}(i-1) resonances to be distinguished. Similarly, these peaks can be assigned in the CBCANH spectrum (Grzesiek & Bax, 1992b). The CC(CO)NH experiment (Farmer & Venters, 1995) allows magnetisation transfer from N-H^{N} to all side-chain carbons of residue i-1. By reference to standard ^{13}C chemical shifts of amino acids (Wuthrich, 1986) and comparison of equivalent strips in the CBCA(CO)NH and CC(CO)NH spectra, the remaining aliphatic side-chain carbons could be identified and assigned to possible amino acid types. For example, the assignment of resonances in this manner for the

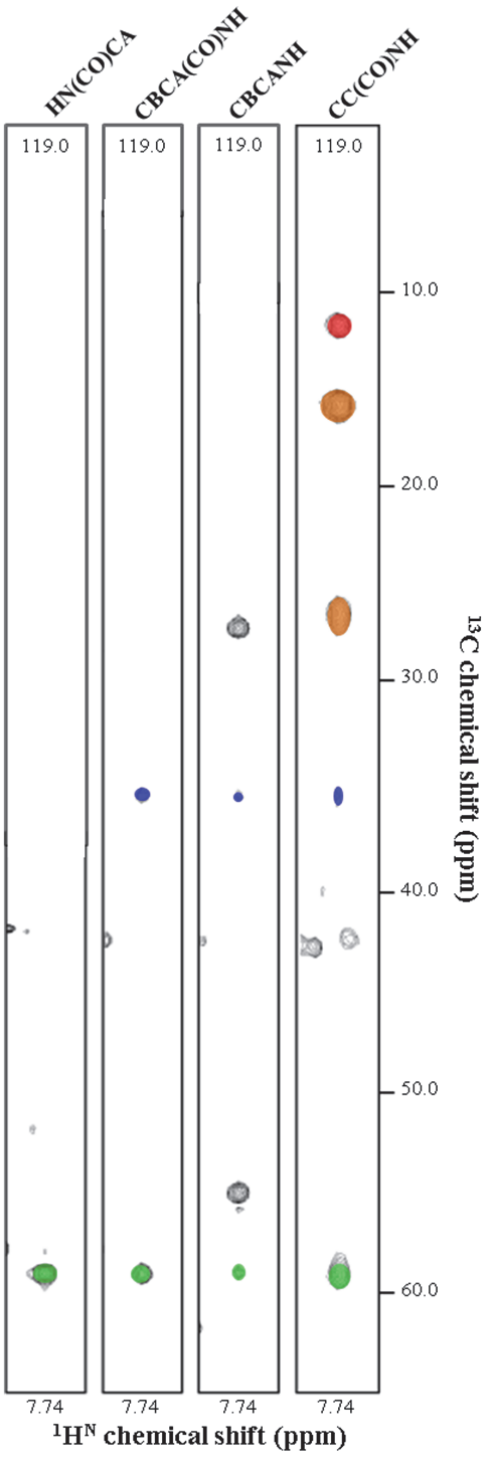


Figure 2.7 *Strips of the HN(CO)CA, CBCA(CO)NH, CBCANH and CC(CO)NH spectra corresponding to an isoleucine residue from the $^{1}H,^{15}N$-HSQC.*

C^{α}, C^{β}, C^{γ} and C^{δ} peaks from residue i-1 are highlighted in green, blue, orange and red respectively. ^{15}N chemical shifts (ppm) are given at the top of each strip.

C^{α}, C^{β}, C^{γ} and C^{δ} atoms of an isoleucine side-chain are shown in **Figure 2.7**.

Sequential connectivities between spin systems were determined using the CBCA(CO)NH and CBCANH spectra in conjunction. The CBCANH experiment correlates N-H^N resonances with C^{α} and C^{β} resonances of both residues i and i-1. These can be easily distinguished via comparison with strips from the CBCA(CO)NH spectrum, which shows i-1 peaks only. The chemical shifts of residue i can then be compared to all other CBCANH strips to identify the partner i-1 peaks, allowing linkages between spin systems to be identified along the protein backbone. Specific assignment of the connected spin systems to the MxiG-$N_{6\text{-}112}$ sequence was achieved by identifying spin systems with side-chain carbon chemical shifts distinctive for a particular residue type (Wuthrich, 1986) and then finding which of the possible neighbouring residues from the sequence would have chemical shifts matching those of the connected spin systems. **Figure 2.8** demonstrates sequential connectivity between strips from the CBCANH spectrum. The two threonine residues were easily identifiable from their ^{13}C chemical shifts, allowing unambiguous sequence-specific assignment of these spin systems to residues I35-L44. In this manner, the backbone amides in the ^{1}H,^{15}N-HSQC of MxiG-$N_{6\text{-}112}$ were assigned (**Figure 2.9**). Assignment showed that the peaks for the backbone amides of F51 and E110 overlapped in the ^{1}H,^{15}N-HSQC, allowing all 102 non-proline backbone amides to be picked in the spectrum. >90% assignment was achieved for side-chain carbon atoms.

The CBCA(CO)NH experiment also allowed magnetisation transfer from N^{δ}-H^{δ} and N^{ε}-H^{ε} to C^{α}/C^{β} and C^{β}/C^{γ} within asparagine and glutamine side-chains respectively. Using the side-chain carbon assignments, resonances for the side-chain amides of asparagine and glutamine residues were able to be distinguished and assigned (**Figure 2.9**). Similarly, the CBCANH spectrum correlates a N^{ε}-H^{ε} resonance to C^{γ} and C^{δ} resonances for an arginine side-chain. These amides were able to be assigned for R13 and R39, although the former overlaps completely with the backbone amide of S7 as highlighted in **Figure 2.6B**. No peak for the side-chain N^{ε}-H^{ε} of R109 was visible.

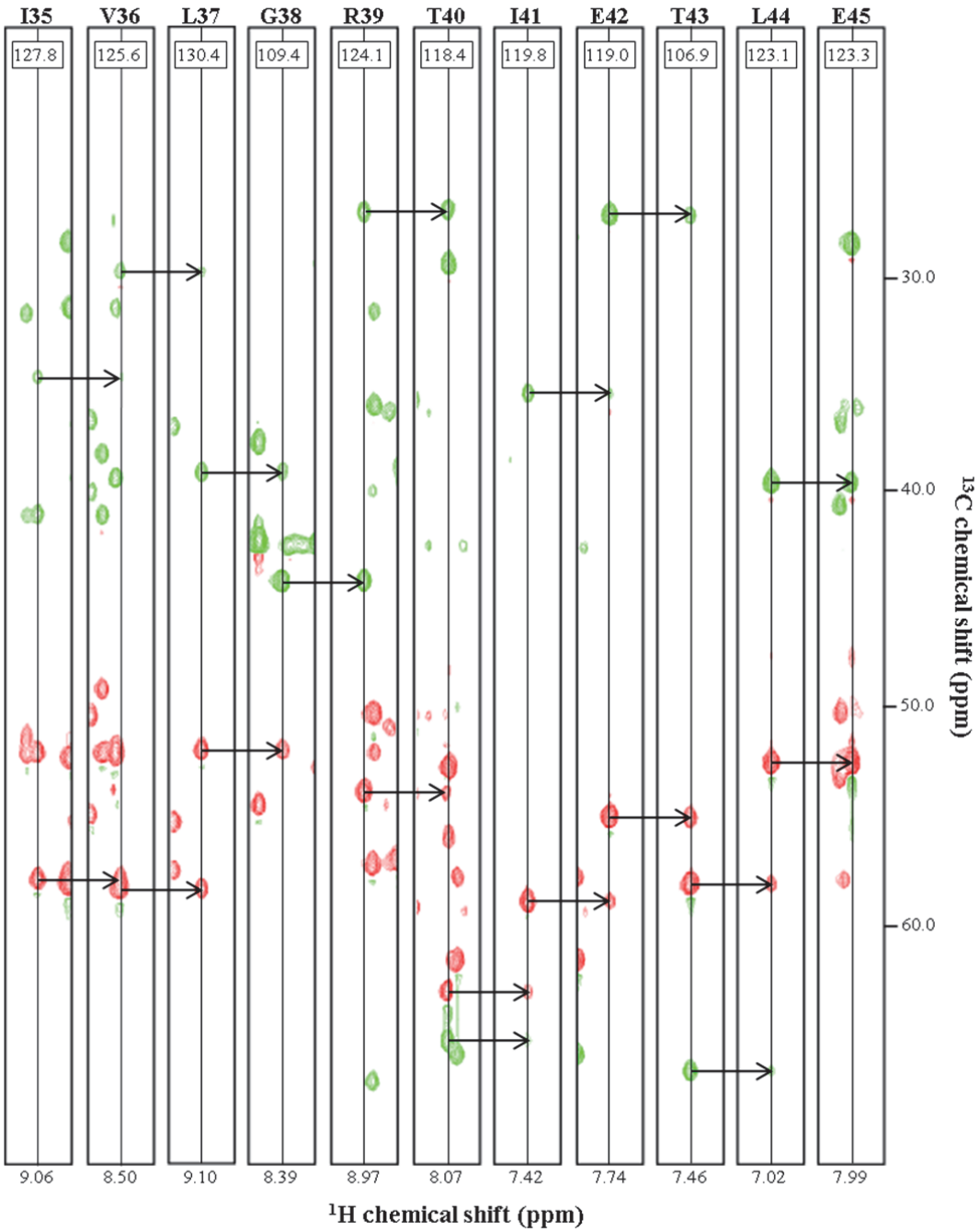


Figure 2.8 *Strips from the CBCANH spectrum with sequence-specific assignment of residues I35-L44 by the triple resonance approach.*

Residue i is indicated above each strip. Red and green contour lines represent positive and negative intensity peaks respectively. Positive peaks are obtained for all non-glycine C^{α} atoms, whilst negative peaks are obtained for C^{β} atoms and the C^{α} of G38. Arrows indicated residue i-1 to i linkages. ^{15}N chemical shifts (ppm) are given at the top of each strip.

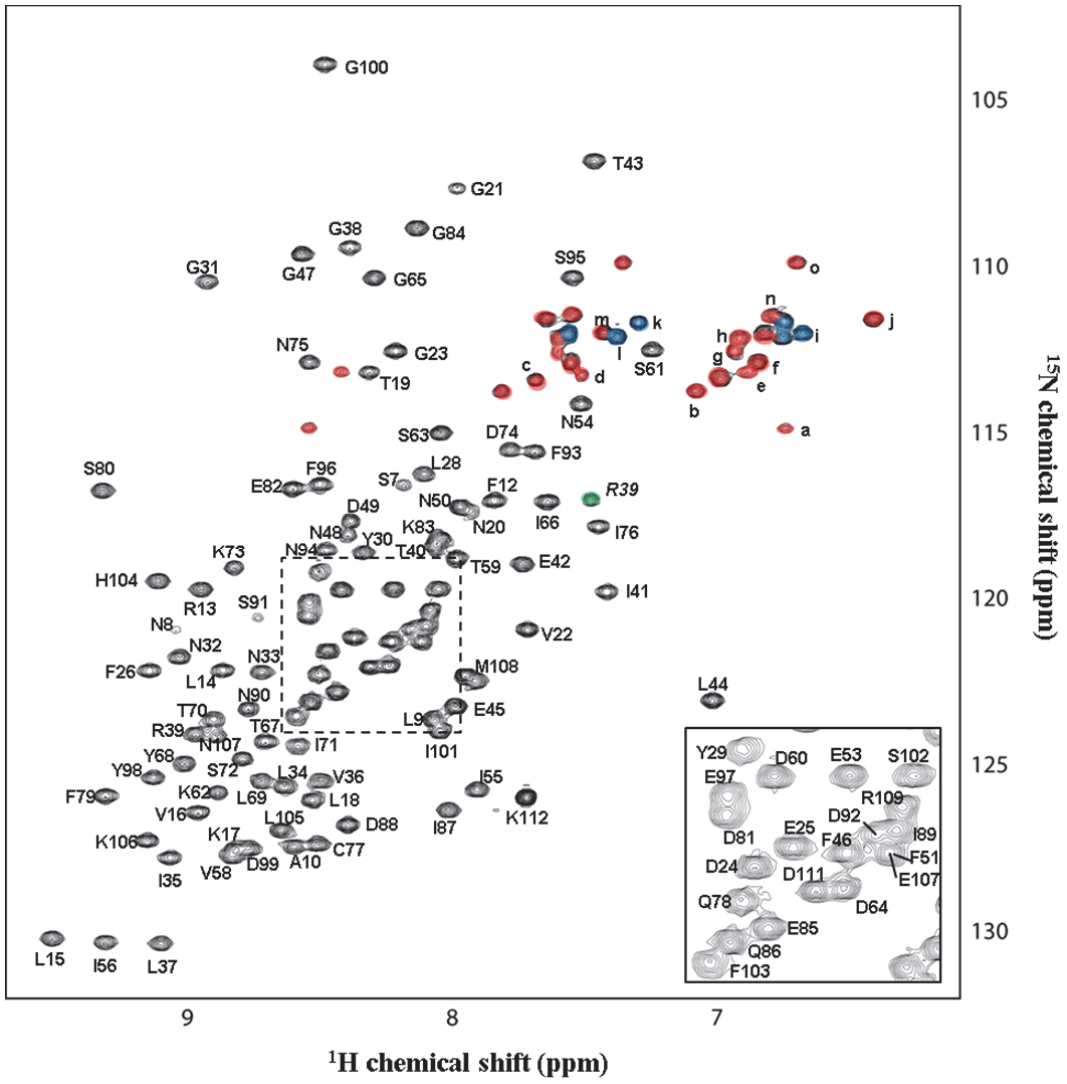


Figure 2.9 *Assignment of the $^{1}H,^{15}N$-HSQC spectrum of MxiG-$N_{6\text{-}112}$.*

Assignments of backbone amides are shown. Peaks corresponding to backbone amides, asparagine side-chains, glutamine side-chains and arginine side-chains are black, red, blue and green respectively. The zoomed inset corresponds to the region of the spectrum indicated by the dashed box. Side-chain amides are assigned as **a** N33, **b** N90, **c** N8, **d** N75, **e** N54, **f** N48, **g** N6, **h** N50, **i** Q86, **j** N32, **k** Q92, **l** Q78, **m** N107, **n** N94 and **o** N20.

A further triple resonance experiment correlated N-H^N resonances to C'(i-1) resonances; C' peaks in the HNCO (Kay *et al.*, 1990) were fully assigned based on the sequence-specific assignment of the $^{1}H,^{15}N$-HSQC of MxiG-$N_{6\text{-}112}$.

Backbone Assignment by the Sequential Approach

Experiments were also carried out that separate each signal from the 2D $^{1}H,^{15}N$-HSQC spectrum into a third ^{1}H dimension. The HNHA experiment (Vuister & Bax, 1993) allows through-bond coherence transfer from N-H^N to H^{α}, allowing unambiguous assignment of H^{α} for each residue based on the backbone amide assignment of the $^{1}H,^{15}N$-HSQC spectrum of MxiG-N_{6-112} (**Figure 2.9**). The $^{1}H,^{15}N$ nuclear Overhauser effect spectroscopy (NOESY)-HSQC experiment (Marion *et al.*, 1989) achieves through-space coherence transfer from N-H^N to all protons within <7 Å, giving rise to cross-peaks that show inter-residue connections between pairs of protons in addition to the intra-residue connections that through-bond experiments are limited to (Wuthrich *et al.*, 1982). The presence of nuclear Overhauser effects (NOEs) between adjacent residues was used to check the sequence-specific backbone assignment of MxiG-N_{6-112} as determined by the triple resonance approach. H^N of residue i should have NOEs with H^N of residues i-1 and i+1 and H^{α} of residue i-1. The $^{1}H,^{15}N$ NOESY-HSQC generally showed the expected connectivities to give confidence in the backbone assignment, as demonstrated by the NOE connections between the strips for residues I35-L44 (**Figure 2.10**).

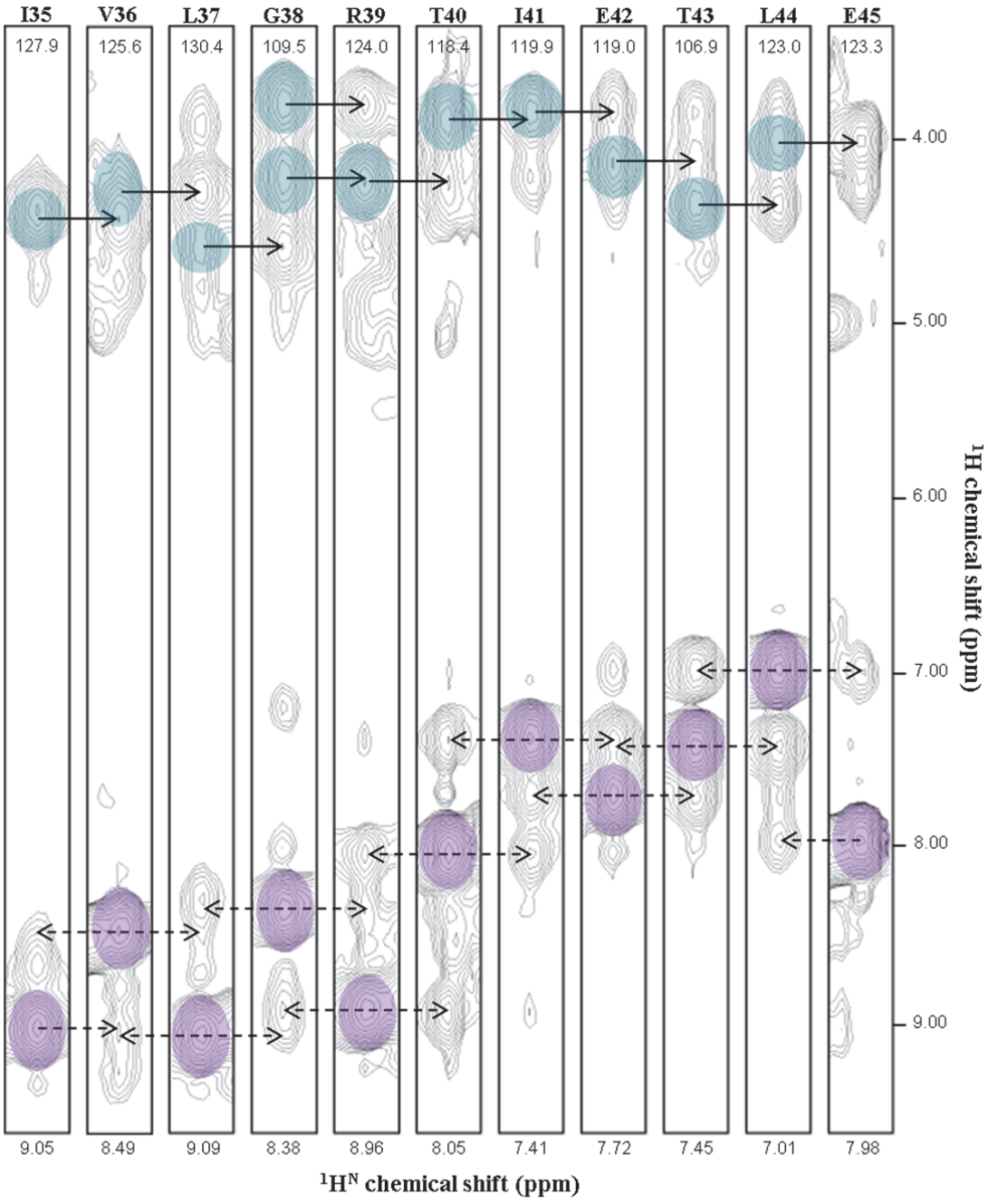


Figure 2.10 *Strips from the $^1H,^{15}N$ NOESY-HSQC spectrum with sequence-specific assignment of residues I35-L44 by the sequential approach.*

Residue i is indicated above each strip. Peaks corresponding to the H^N resonance of i and the intra-residue H^N-H^α NOE are highlighted in purple and turquoise respectively. Forwards and backwards facing dashed arrows indicate inter-residue $H^N(i)$-$H^N(i-1)$ and $H^N(i)$-$H^N(i+1)$ NOEs respectively. Solid arrows indicate the inter-residue $H^N(i)$ -$H^\alpha(i-1)$ NOE. ^{15}N chemical shifts (ppm) are given at the top of each strip.

Assignment of Side-Chain Protons

A series of further 3D spectra that showed through-bond coherence transfer from N-H^N to side-chain protons were used for their assignment (**Figure 2.11**). Previous assignments of H^α atoms from the HNHA experiment allowed H^α and H^β peaks to be distinguished in N-H^N(i+1) strips from the HBHA(CO)NH spectrum (Muhandiram & Kay, 1994). The remaining side-chain protons could then be identified in the corresponding strip from a H(CCO)NH experiment (Grzesiek *et al.*, 1993), for example the H^γ protons of the E30 side-chain (**Figure 2.11**). Pseudo-atoms were assigned for protons that sample the same chemical environment and hence yield a single peak, most commonly those within side-chain –CH_3 groups.

Total correlation spectroscopy (TOCSY) pulse sequences allow through-bond coherence transfer across a whole spin-coupled system, such as an amino acid side-chain. The HCCH-

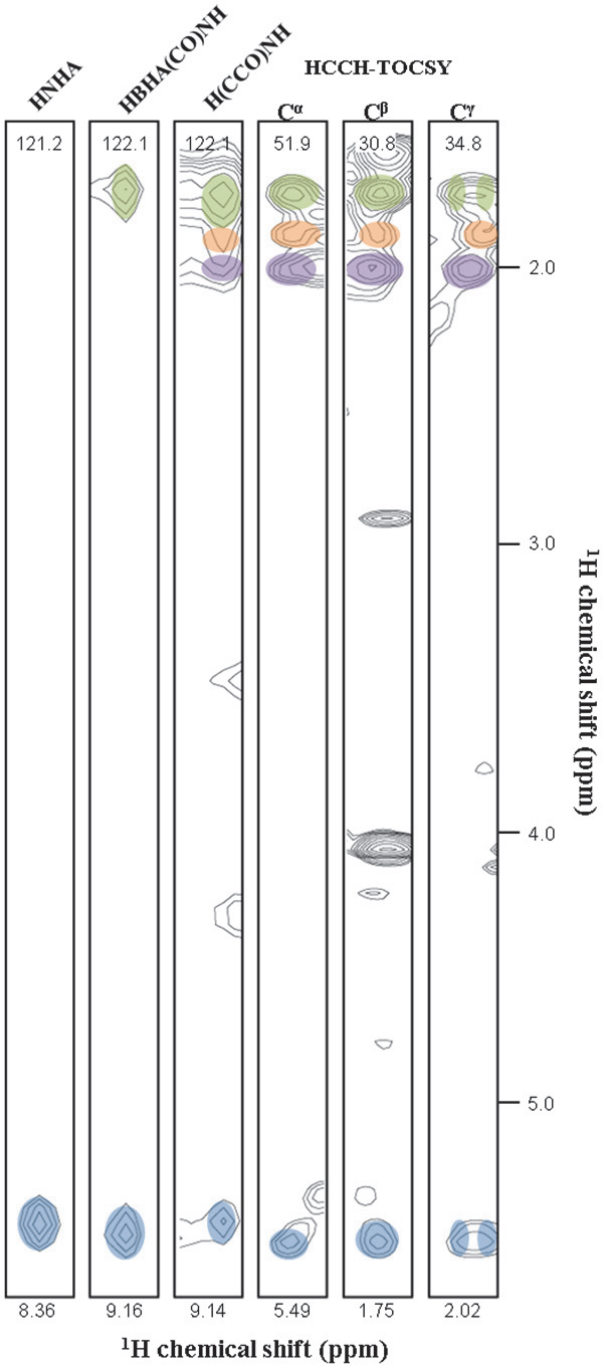


Figure 2.11 *Strips of the HNHA, HBHA(CO)NH, H(CCCO)NH and HCCH-TOCSY spectra corresponding to the side-chain of E30.*

H^α, H^β, $H^{\gamma 2}$ and $H^{\gamma 3}$ peaks are highlighted in blue, green, purple and orange respectively. Chemical shifts (ppm) at the top of each strip correspond to ^{15}N for HNHA, HBHA(CO)NH and H(CCCO)NH strips and ^{13}C for HCCH-TOCSY strips.

TOCSY experiment (Kay *et al.*, 1993) splits a 2D $^{1}H,^{13}C$-HSQC into a third ^{1}H dimension. Although peaks in this spectrum often overlap, it is highly redundant, with each C-H strip showing cross-peaks to the other side-chain protons. This experiment can help distinguish side-chain protons in the H(CCO)NH spectrum to provide unambiguous assignments. For example, strips corresponding to the side-chain of E30 were identified in the HCCH-TOCSY spectrum by searching at each side-chain carbon resonance for the spin system identified in the H(CCO)NH strip (**Figure 2.11**). Two peaks present in all strips were diagonal peaks in the C^{γ} strip, allowing them to be unambiguously assigned as the two H^{γ} protons of E30. MxiG-$N_{6\text{-}112}$ was transferred into >98% D_2O by three cycles of concentration in a 0.5 ml 3 kDa centrifugal filter (Amicon Ultra) and re-dilution with 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT dissolved in D_2O to acquire the spectrum.

2.3 Secondary Structure Prediction for MxiG-$N_{6\text{-}112}$

Predictions from H^{α} Secondary Chemical Shift and $^{3}J_{HNH\alpha}$ Values

Several measurable NMR parameters are strongly dependent on the character and nature of protein secondary structure, therefore can be used as a predictive tool. For all amino acids, the H^{α} chemical shift tends towards upfield values when in an α-helix and downfield values when in an extended β-strand. The deviance of H^{α} chemical shifts from random coil values (Wishart *et al.*, 1992) for MxiG-$N_{6\text{-}112}$ is shown in **Figure 2.12A**. If the random coil range of H^{α} resonances is taken as ±0.1 ppm, a significant downshift in H^{α} chemical shifts was observed for eight regions: residues F12-L18, E25-Y29, N33-L37, I55-V58, D64-I67, L69-S72, I76-D81 and S102-K106. These regions could be predicted to be β-strands if the chemical shift index (CSI) definition is taken of four or more uninterrupted positive deviations from the random coil range (Wishart *et al.*, 1992). Conversely, residues R39-E41 showed a significant negative deviation from the random coil range, predicting an α-helical region.

The relative intensity of the H^{α} cross-peak and H^{N} diagonal peak in the HNHA spectrum provides a sensitive measurement of the scalar three-bond coupling constant between H^{N} and H^{α} ($^{3}J_{HNH\alpha}$; **Figure 2.12B**). The Karplus equation (Karplus, 1959, Pardi *et al.*, 1984) relates $^{3}J_{HNH\alpha}$ to the backbone torsion angle Φ, with both parameters having characteristic values within α-helices and β-strands. Therefore the deviation of $^{3}J_{HNH\alpha}$ from 7 Hz, the representative value for an unstructured peptide, can be used as an indicator of secondary structure. However intramolecular motions lead to averaging over distributions of dihedral angles, so values <6 Hz and >8 Hz are required to classify α-helices and β-sheets respectively. **Figure 2.12B** suggests

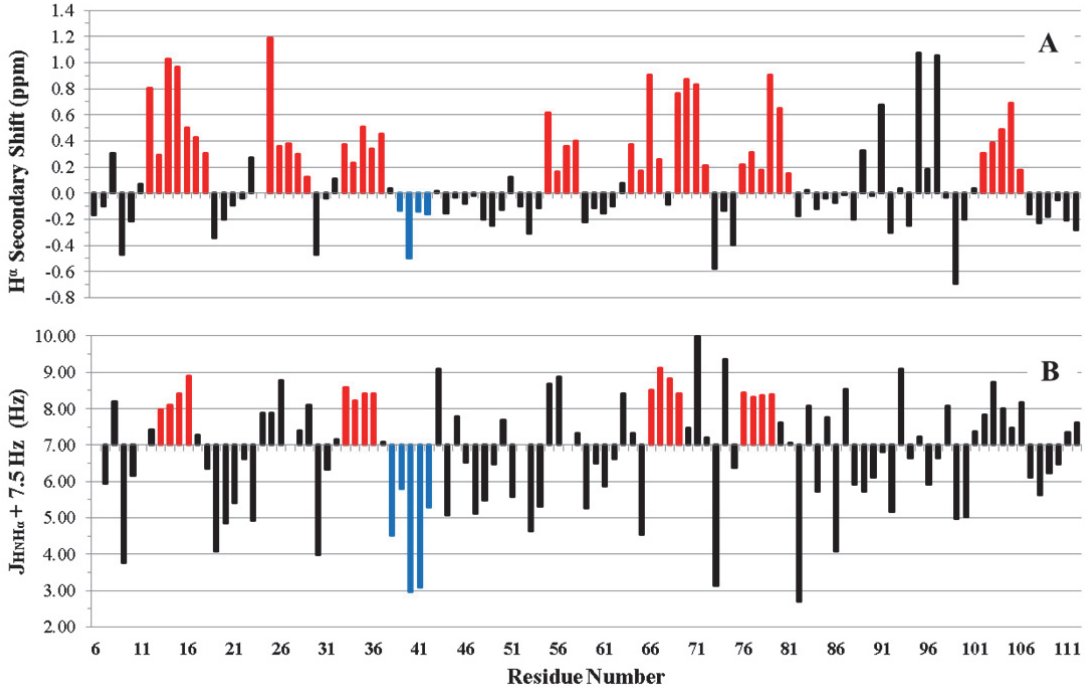


Figure 2.12 *Secondary structure prediction for MxiG-$N_{6\text{-}112}$ from H^{α} secondary chemical shift and $^{3}J_{HNH\alpha}$ values.*

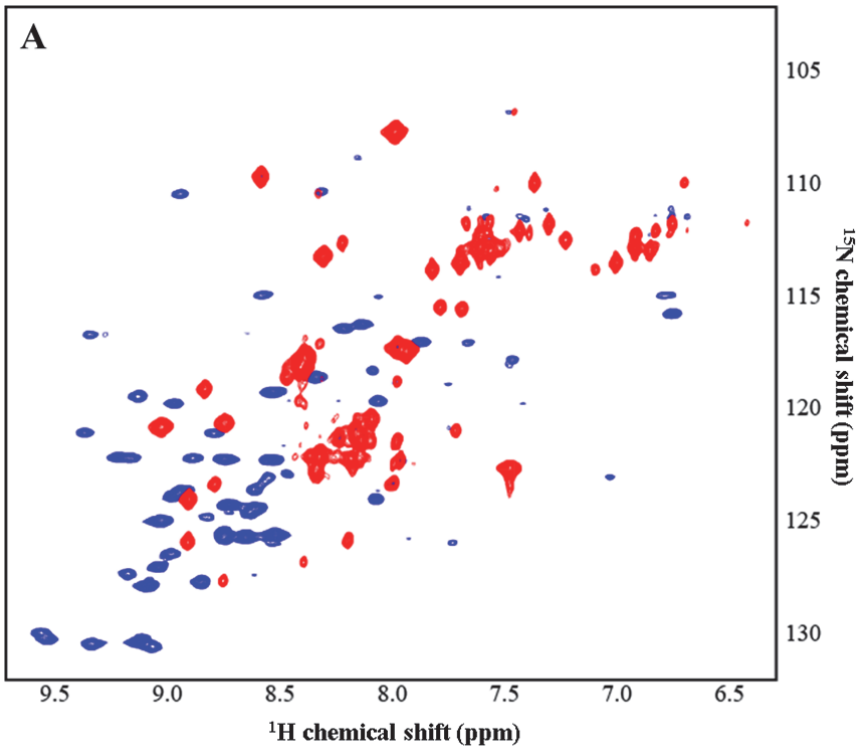
A Secondary chemical shifts of assigned H^{α} atoms in MxiG-$N_{6\text{-}112}$ relative to random coil values (Wishart *et al.*, 1992). Zero values indicate atoms for which assignments are not available. For glycine residues, the mean average of the two H^{α} chemical shifts is taken. Regions which have ≥4 consecutive residues with H^{α} chemical shifts >0.1 ppm above the random coil value are shown in red, whilst those with ≥4 consecutive residues >0.1 ppm below the random coil value are shown in blue.

B $^{3}J_{HNH\alpha}$ values for MxiG-$N_{6\text{-}112}$ calculated from intensities in the HNHA spectrum; $I_{cross}/I_{diagonal} = -\tan^{2}(2\Pi^{3}J_{HNH\alpha}\delta_{2})$ was used (Cavanagh, 2007), where I_{cross} and $I_{diagonal}$ are the peak heights of the H^{α} cross-peak and H^{N} diagonal peak respectively and $\delta_{2} = 0.01305$s. Raw $^{3}J_{HNH\alpha}$ values have been increased by 7.5% to take the differential relaxation between H^{α} and H^{N} into account. Zero values indicate the N-terminal and proline residues not present in the spectrum. Regions which have ≥4 consecutive residues with $^{3}J_{HNH\alpha}$ values >8.0 Hz are shown in red, whilst those with ≥4 consecutive residues <6.0 Hz are shown in blue.

MxiG-N_{6-112} had β-strands at residues R13-V16, N33-V36, I66-L69 and I76-F79, regions also indicated by the H^{α} secondary chemical shift data. There was also a correlation between high $^{3}J_{HNH\alpha}$ values and H^{α} chemical shift deviations for residues D24-Y29 and I101-K106, although the data fluctuated within these regions so some $^{3}J_{HNH\alpha}$ values were below 8 Hz. Residues G38-E42 had $^{3}J_{HNH\alpha}$ values consistently <6 Hz, indicating an α-helical region, in agreement with the H^{α} secondary chemical shift data.

Measurement of Backbone Amide Proton Exchange

Backbone amide protons exchange rapidly with the solvent when exposed, therefore identification of slowly-exchanging protons indicates amide groups that are involved in hydrogen bonding or are shielded from the solvent by the rest of the protein structure. As deuterons are not detected by the $^{1}H,^{15}N$-HSQC pulse sequence, the exchange of MxiG-N_{6-112} backbone amide protons with deuterons was probed. After exchange of MxiG-N_{6-112} into >98% D_2O, a series of $^{1}H,^{15}N$-HSQC spectra were collected over 20 h. After 1.5 h, a significant amount of exchange had already taken place, shown by the disappearance or decrease in intensity of many backbone amide peaks in the $^{1}H,^{15}N$-HSQC spectrum (**Figure 2.13A, blue**). When intense peaks were mapped on to the sequence of MxiG-N_{6-112} (**Figure 2.13B, blue**), these residues with slowly-exchanging backbone amides clustered: F12-L18, F26-L37, I67-S72 and I101-K106. There was a striking correlation between these regions and those identified from the secondary chemical shift and $^{3}J_{HNH\alpha}$ data as possible β-strands (**Figure 2.14**), suggesting the backbone amides of these residues were involved in hydrogen bonds within β-sheets. In addition, the β-strand previously suggested between I76 and D81 was supported by this exchange data, with the alternating pattern of fast and slow exchanging amides suggesting this strand was on the periphery of a β-sheet. R39 and T40 within the predicted α-helical region also had protected amides, suggesting they were involved in hydrogen bonding.



B

```
        10                  20                  30                  40
N S N L A P F R L L V K L T N G V G D E F P L Y Y G N N L I V L G R T I E T L

        50                  60                  70                  80
E F G N D N F P E N I I P V T D S K S D G I I Y L T I S K D N I C Q F S D E K G

        90                  100                 110
E Q I D I N S Q F N S F E Y D G I S F H L K N M R E D K
```

Figure 2.13 *Measurement of backbone amide proton exchange*

A Overlay of a $^{1}H,^{15}N$-HSQC spectrum obtained 1.5 h after transfer of MxiG-$N_{6\text{-}112}$ into >98% D_2O (blue) and a $^{1}H,^{15}N$-HSQC spectrum obtained using the CLEANEX-PM mixing scheme (red).

B Mapping of intense peaks from both spectra on to the sequence of MxiG-$N_{6\text{-}112}$. Peaks with a data height >2 x 10^6 from the D_2O exchange spectrum were selected and the corresponding residues shown in blue, whilst peaks with a data height >1 x 10^7 from the CLEANEX-PM spectrum are show in red. Residues shown in black have peaks with a data height below the threshold in both spectra.

In a converse experiment, the saturation transfer from water to amide protons can be probed using the phase-modulated CLEAN chemical exchange (CLEANEX-PM) pulse sequence, which allows selective excitation of water and observation of magnetisation transfer to other sites (Hwang *et al.*, 1997, Hwang *et al.*, 1998). NOEs only arise for those amide protons that are exposed to the solvent, providing an indication of residues that are present on the surface of the protein. The backbone amides in MxiG-$N_{6\text{-}112}$ with intense peaks in the $^{1}H,^{15}N$-HSQC from the CLEANEX-PM experiment were distinct from those with intense peaks in the same spectrum from the D_2O exchange experiment (**Figure 2.13A**). When residues with solvent-exposed amides were shown with respect to the sequence of MxiG-$N_{6\text{-}112}$, they clustered between regions with slowly-exchanging amides, presumably in putative loop regions (**Figure 2.13B, red**). In addition the CLEANEX-PM experiment indicated all asparagine and glutamine side-chain amides were on the surface of the protein with the exception of N33 and N54, suggesting these were plausible candidates for side-chain to backbone hydrogen bonding within the core of the protein.

Summary of Secondary Structure Prediction for MxiG-$N_{6\text{-}112}$

Regions of secondary structure independently suggested by H^{α} secondary chemical shifts, $^{3}J_{HNH\alpha}$ values and slowly-exchanging amides clearly correlated, allowing seven β-strands and one α-helix to be predicted within the structure of MxiG-$N_{6\text{-}112}$ (**Figure 2.14**).

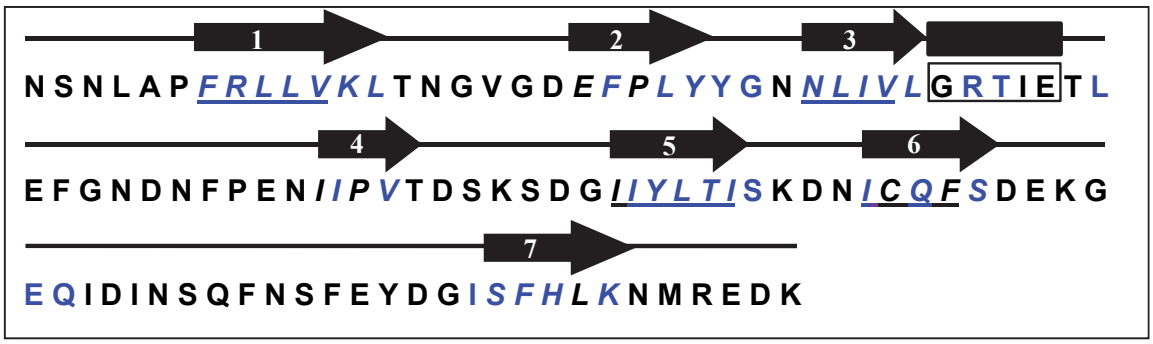


Figure 2.14 *Summary of secondary structure data for MxiG-$N_{6\text{-}112}$.*

Regions which have ≥4 consecutive residues with H^{α} chemical shifts >0.1 ppm above the random coil value and $^{3}J_{HNH\alpha}$ >8.0 Hz are in italics and underlined respectively. Regions which have ≥4 consecutive residues with H^{α} chemical shifts >0.1 ppm below the random coil value and $^{3}J_{HNH\alpha}$ <6.0 Hz are boxed. Residues with slowly-exchanging amide protons are in blue. Predicted β-strands are shown by arrows, whilst the predicted α-helix is represented by a cylinder.

2.4 Determination of Structural Restraints for MxiG-N_{6-112}

Identification of Secondary Structure via Medium and Long-Range NOEs

Within the $^{1}H,^{15}N$ NOESY-HSQC spectrum, medium or long-range NOEs can be defined depending on whether through-space transfer occurs to a residue less or more respectively than four residues distant in the sequence. Secondary structure elements are characterised by signature medium and long-range NOEs between backbone protons. Antiparallel (**Figure 2.15A**) and parallel β-strands have distinctive repeating patterns of long-range H^N-H^N and H^N-H^α NOEs that can be traced their entire length (Wuthrich *et al.*, 1984). Two β-sheets were identified for MxiG-N_{6-112} (**Figure 2.16**), encompassing the seven β-strands predicted by the secondary structure analysis (**Figure 2.14**) and including six antiparallel and one parallel arrangement.

α-helices show a series of medium-range NOEs between backbone H^N and H^α protons up to four residues apart in the sequence, most notably strong $H^N(i)$-$H^N(i\pm1)$, intermediate $H^N(i)$-$H^\alpha(i-1)$, $H^N(i)$-$H^\alpha(i-3)$ and weak $H^N(i)$-$H^N(i\pm2)$ NOEs (Wuthrich *et al.*, 1984). These NOEs were present in the $^{1}H,^{15}N$ NOESY-HSQC spectrum for residues G38-E45 (**Figure 2.15B**), in agreement with the previously predicted α-helical region (**Figure 2.14**).

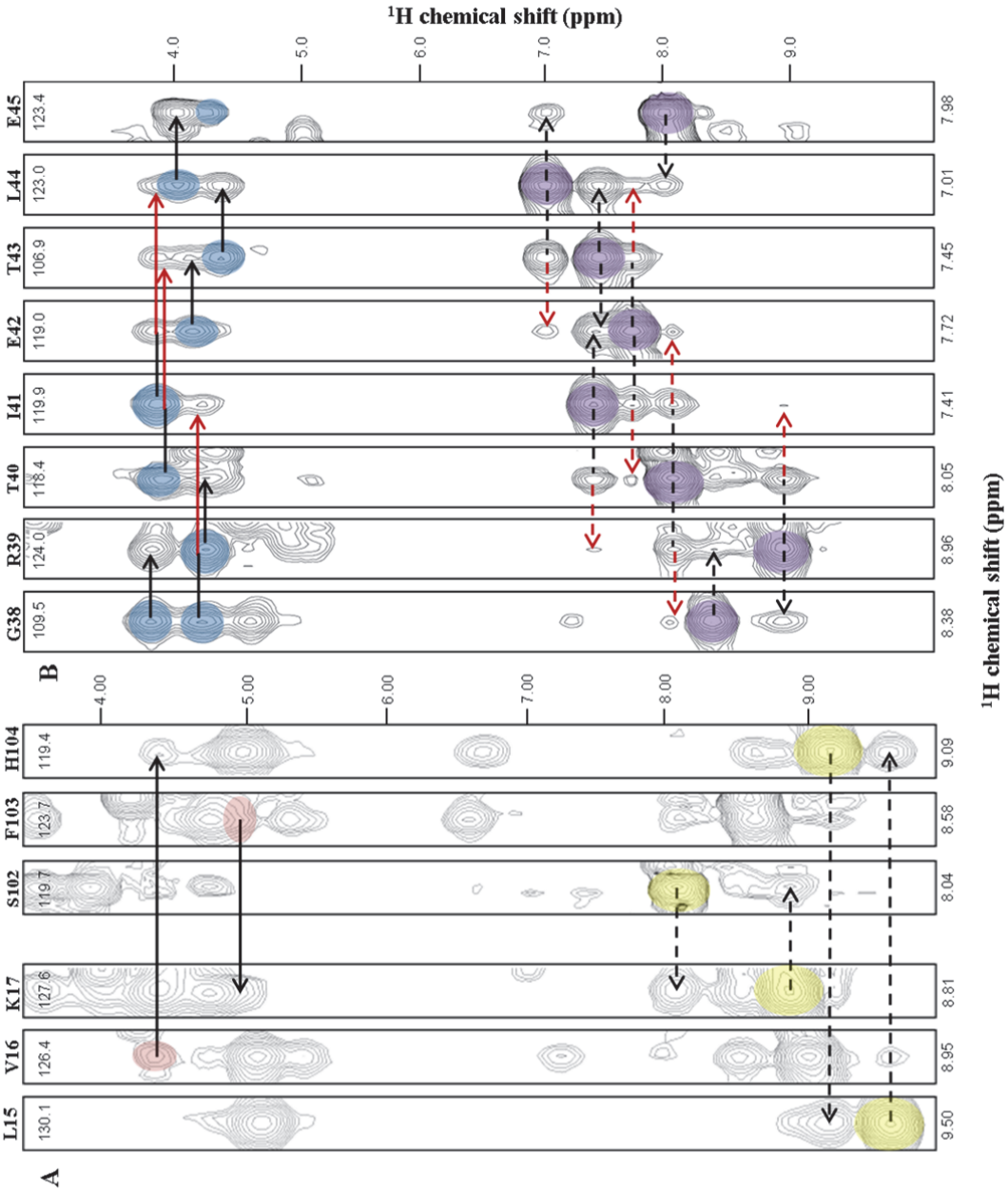


Figure 2.15 *Strips from the $^{1}H,^{15}N$ NOESY-HSQC spectrum identifying NOEs characteristic of secondary structure elements*.

Residue i is shown above and ^{15}N chemical shifts (ppm) are shown at the top of each strip.

A Long-range NOEs between adjacent portions of two antiparallel β-strands. Peaks corresponding to H^N and H^α of residue i are highlighted in yellow and red respectively. Dashed and solid arrows represent long-range H^N-H^N and H^N-H^α NOEs respectively.

B Short and medium-range NOEs indicative of an α-helix from residues G38 to E45. Peaks corresponding to the H^N resonance of i and the intra-residue H^N-H^α NOE are highlighted in purple and blue respectively. Forward and backward facing dashed black arrows indicate inter-residue $H^N(i)$-$H^N(i-1)$ and $H^N(i)$-$H^N(i+1)$ NOEs respectively. Forward and backward facing red arrows indicate inter-residue $H^N(i)$-$H^N(i-2)$ and $H^N(i)$-$H^N(i+2)$ NOEs respectively. Solid black and red arrows indicate inter-residue $H^N(i)$-$H^\alpha(i-1)$ and $H^N(i)$-$H^\alpha(i-3)$ NOEs respectively.

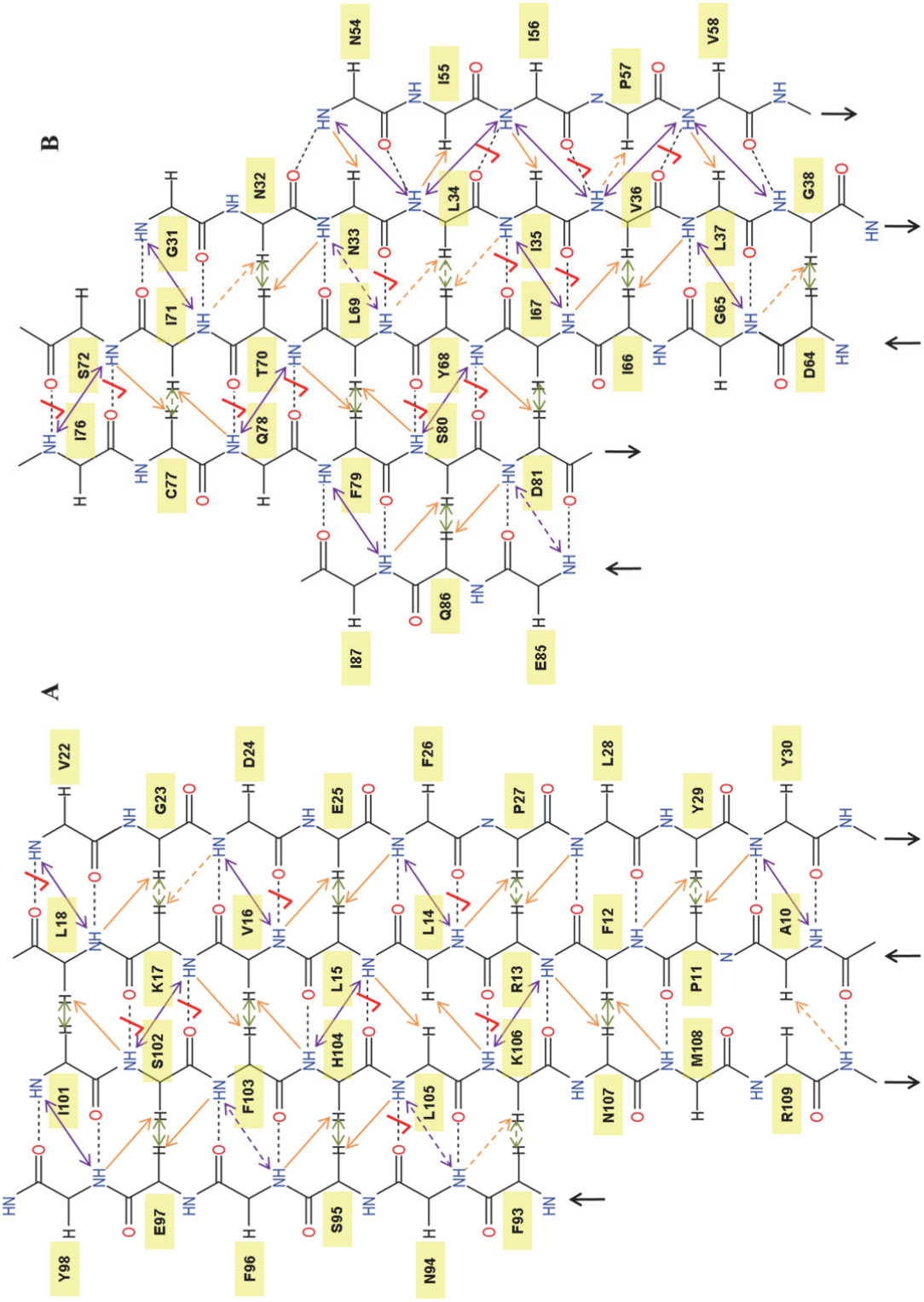


Figure 2.16 *MxiG-$N_{6\text{-}112}$ β-sheets identified from long-range NOEs and hydrogen bond couplings*.

H^N-H^N, H^N-H^α and H^α-H^α NOEs between β-strands are shown by purple, orange and green arrows respectively. Solid arrows represent clear NOEs, whilst dashed arrows represent ambiguous NOEs due to peak overlap. Red ticks show hydrogen bonds confirmed by the long-range H(N)CO experiment. The directionality is shown below each strand. **A** 4-stranded antiparallel β-sheet and **B** 5-stranded β-sheet with one parallel and three antiparallel interactions.

Direct Observation of Hydrogen Bonds

Previous measurements of the rate of backbone amide proton exchange with deuterons indirectly indicated amides that may be involved in hydrogen bonds within MxiG-N_{6-112}. However, hydrogen bonds give rise to a one bond scalar coupling between the donor amide and acceptor carbonyl groups. These can be directly detected by a long-range H(N)CO experiment, which is a modification of the standard H(N)CO experiment used for protein assignment, using a much longer nitrogen to carbon transfer time to achieve transfer of magnetisation across the hydrogen bond instead of along the protein backbone. (Cordier & Grzesiek, 1999, Cordier *et al.*, 2008) However, hydrogen bond couplings are very small, with typical values ranging between 0.1 Hz in α-helical structures to 0.4 Hz in β-sheets (Cornilescu *et al.*, 1999), so the main limitation of the experiment is its sensitivity. Although transverse relaxation optimised spectoscopy (TROSY) techniques have increased the feasible molecular weight for hydrogen bond detection, applications have traditionally been limited to a handful of model proteins (Cordier & Grzesiek, 1999, Wang *et al.*, 1999, Banci *et al.*, 2002).

A 2D long-range H(N)CO-TROSY experiment was successfully performed by Dr Rainer Kümmerle (Bruker BioSpin AG) on a Bruker 700 MHz spectrometer with TCl cryoprobe at 25^{o}C on 5.4 mM $^{13}C/^{15}N$-labelled MxiG-N_{6-112}. Cross-peaks between H^N and C' atoms involved in hydrogen bonding were identified through comparison to their inter-residue backbone correlations in a standard 2D H(N)CO spectrum (**Figure 2.17A**). The experiment allowed the assignment of a large number of hydrogen bonds (**Figure 2.17B**), which were located between the predicted β-strands of MxiG-N_{6-112} (**Figure 2.16**).

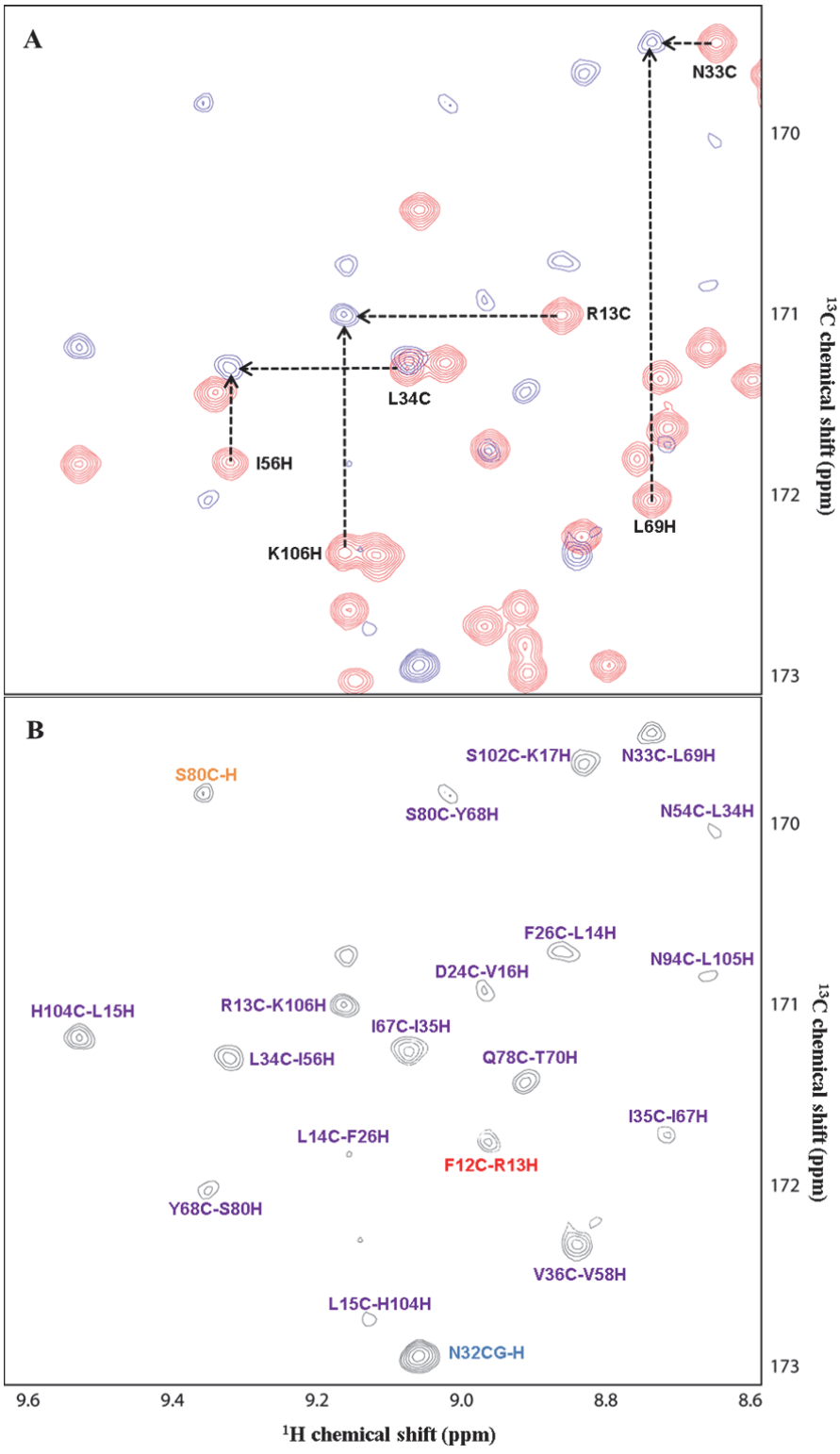


Figure 2.17 *Selected region of the 2D long-range H(N)CO spectrum*.

A Overlay of the standard H(N)CO spectrum (red) with the long-range H(N)CO-TROSY spectrum (blue). Dashed arrows indicate acceptor carbonyl (horizontal) and donor amide (vertical) groups giving rise to cross-peaks representing hydrogen bonds.

B Assignment of the long-range H(N)CO-TROSY spectrum. Purple labels indicate peaks due to hydrogen bond couplings, whilst red labels show inter-residue backbone correlations. Intra-residue backbone and side-chain correlations are shown in orange and blue respectively.

Identification of Long-Range NOEs in the $^{1}H,^{13}C$ NOESY-HSQC spectrum

A $^{1}H,^{13}C$ NOESY-HSQC spectrum (Ikura *et al.*, 1990), obtained after transfer of MxiG-N_{6-112} into >98% D_2O, has the same dimensionality as the HCCH-TOCSY spectrum, except the experiment achieves through-space coherence transfer. Although each C-H strip still shows cross-peaks to all side-chain protons, there are also additional cross-peaks for all other protons within 7 Å that can be distinguished through comparison to the HCCH-TOCSY strips. **Figure 2.18** shows an example of long-range NOEs between residues E97 and S101 in strips taken from the $^{1}H,^{13}C$ NOESY-HSQC spectrum. Assignment of NOEs in this spectrum is less ambiguous than in the $^{1}H,^{15}N$ NOESY-HSQC spectrum due to each NOE appearing reciprocally on two strips in the spectrum. For example, the NOE between E97 H^{α} and S101 H^{α} was present on both the E97 C^{α}-H^{α} strip and the S101 C^{α}-H^{α} strip (**Figure 2.18**). The $^{1}H,^{13}C$ NOESY-HSQC spectrum showed the characteristic pattern of strong H^{α}-H^{α} NOEs between the antiparallel β-strands (**Figure 2.16**) and a large number of other long-range NOEs between β-strands and the two β-sheets.

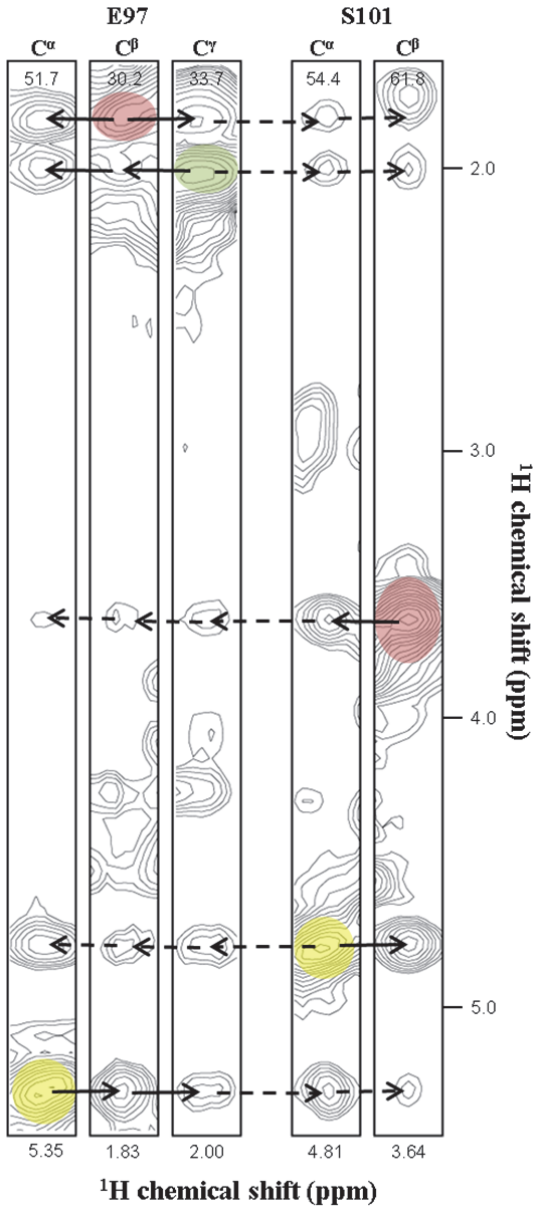


Figure 2.18 *Strips from the $^{1}H,^{13}C$ NOESY-HSQC spectrum showing long-range NOEs between residues E97 and S101.*

Diagonal cross-peaks corresponding to H^{α}, H^{β} and H^{γ} are highlighted in yellow, red and green respectively. Solid and dashed arrows represent intra- and inter-residue NOEs respectively. ^{13}C chemical shifts (ppm) are given at the top of each strip.

Conversion of Experimental Parameters to Distance Restraints

Peak volumes for NOEs from the $^{1}H,^{15}N$ NOESY-HSQC and $^{1}H,^{13}C$ NOESY-HSQC spectra were calculated by Gaussian fitting within Sparky. NOE peaks with no overlap and a fit residual <5% were chosen to calibrate volume with internuclear distance. For the $^{1}H,^{13}C$ NOESY-HSQC spectrum, H^{α}-H^{α} NOEs across antiparallel β-strands were assigned a distance of 2.3 Å, whilst for the $^{1}H,^{15}N$ NOESY-HSQC spectrum, $H^{N}(i)$-$H^{N}(i+1)$ NOEs within antiparallel β-strands were assigned a distance of 4.3 Å (Wuthrich *et al.*, 1984). These calibrations were then used by the CALIBA macro (Güntert *et al.*, 1991) to convert all unambiguous inter-residue NOEs to distance restraints. Stereospecific assignments were obtained where possible using the 'stereoassign' macro within the program DYANA (Güntert *et al.*, 1997). Subsequently distance restraints were modified with the 'distance modify' macro, to remove redundant or meaningless restraints, adjust distance limits for diasterotopic groups and insert pseudo-atoms where stereospecific assignments were unavailable. This produced a final list of distance restraints for use in the structure calculation.

Hydrogen bonds identified by the long-range H(N)CO experiment (**Figure 2.17B**) were assigned upper and lower distances respectively of 3.0 Å and 2.7 Å for N-O and 2.0 Å and 1.8 Å for H^{N}-O atoms (Williamson *et al.*, 1985). However, the β-sheets shown by long-range NOEs (**Figure 2.16**) and the slowly-exchanging amides shown by D_2O exchange data (**Figure 2.13B**) together provide strong evidence for further hydrogen bonds in MxiG-$N_{6\text{-}112}$. This indicates only the strongest hydrogen bonds led to cross-peaks in the long-range H(N)CO experiment, in agreement with the finding that the magnitude of the coupling constant exponentially depends on the length and geometry of the hydrogen bond. Therefore additional hydrogen bonds not present in the spectrum were assigned more relaxed upper distances of 3.5 Å for N-O and 2.5 Å for H^{N}-O atoms, in correlation with $J_{NC'}$ coupling being unobservable for N-O > 3.3 Å. (Cornilescu *et al.*, 1999)

Determination of Dihedral Angle Restraints for MxiG-N_{6-112}

Although values for $^{3}J_{HNH\alpha}$ calculated from the HNHA spectrum can be converted to the backbone torsion angle φ, the error on these restraints is high as peak intensities from a severely overlapped region of the spectrum along the diagonal provide the experimental data. Therefore all dihedral angle restraints were obtained using TALOS+ (Shen *et al.*, 2009), which predicts φ and ψ for each residue from backbone $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, $^{1}H^{\alpha}$ and $^{1}H^{N}$ chemical shifts using a library of 200 crystal structures with complete NMR backbone assignments. The internal check for correct chemical shift referencing (Wang *et al.*, 2005) indicated ^{1}H chemical shift values were correct, whilst a correction of +2.9 ppm was required for ^{13}C. When implemented TALOS+ produced an artificial neural network-predicted secondary structure consistent with **Figure 2.14**, indicating these new ^{13}C chemical shift values were more appropriate for predicting the dihedral angles of MxiG-N_{6-112}. Only φ and ψ predictions which were classified as unambiguous by TALOS+ were used to provide dihedral angle restraints for use in the structure calculation.

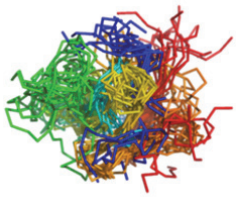
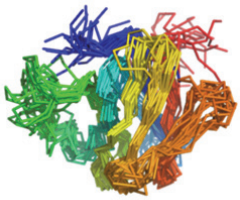
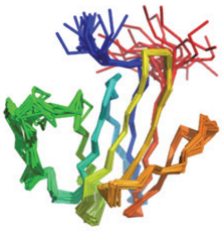
2.5 Determination of the NMR Structure of MxiG-N_{6-112}

Structures for MxiG-N_{6-112} were calculated using default parameters in the 'anneal' macro within DYANA (Güntert *et al.*, 1997), which performs simulated annealing with molecular dynamics in torsion angle space. Experimentally derived values for internuclear distances, hydrogen bond lengths and dihedral angles provided the input conformational restraints. 10000 molecular dynamics steps were performed at the default temperature values, followed by 1500 steps of conjugate gradient minimisation. A family of 350 structures was calculated, from which the 20 structures with the lowest target function were taken as the representative ensemble. These were subsequently visualised using MOLMOL-2.0 (Koradi *et al.*, 1996) and PyMOL (DeLano Scientific).

After each structure calculation, the quality of the MxiG-N_{6-112} ensemble was assessed by three main criteria:

- the root mean squared deviation (RMSD) of the ensemble indicated the similarity between structures and hence the extent to which the input restraints could define the atomic coordinates;
- the number of violations of the input distance and dihedral angle restraints showed the extent to which the NMR data were satisfied by the structures within the ensemble. Given that DYANA cannot take individual errors on peak volumes into account, a distance violation ≤0.5 Å was deemed acceptable given a typical 5% fit residual of peaks calibrated up to distances of 7 Å. Similarly, an angle violation ≤ 5^o corresponded to ~5% of a typical 60^o φ or ψ range predicted by TALOS+;
- the proportion of residues with φ and ψ angles within allowed regions of the Ramachandran plot showed the extent to which the structures satisfied the laws of protein chemistry. Stereochemical quality was assessed using MolProbity (Davis *et al.*, 2007).

Improvements in the statistics associated with the above criteria were achieved in an iterative manner, largely by modifying the input restraints for the structure calculation based on information from the previous MxiG-N_{6-112} ensemble. For example, distance restraints which were violated by the structure ensemble often pointed towards errors in the assignment and integration of NOE peaks that could be corrected. Similarly, dihedral angle restraint violations and residues in disallowed regions of the Ramachandran plot together highlighted some over-stringent TALOS+ predictions, allowing these restraints to be relaxed accordingly. Conversely, improved knowledge of relative atom positions in the MxiG-N_{6-112} tertiary structure removed the ambiguity over the assignment of some NOEs, allowing further distance restraints to be defined. In this manner, 79 rounds of refinement were performed that led to a progressive improvement in structural quality (**Table 2.2**).

Table 2.2 Models from iterative rounds of MxiG-N_{6-112} refinement			
Refinement Round	1	8	62
Final ensemble of 20 structures[1]			
Backbone RMSD (Å)[2]	5.62 ± 3.27	1.93 ± 0.49	0.58 ± 0.12
Residues in allowed regions (%)[3]	73	80	99
Number of violations in >14 structures:			
Distance restraints (≥0.5 Å)	114	107	3
Dihedral angle restraints (≥5°)	100	119	0

[1]Superimposition by residues 13-38, 54-81 & 93-106.
[2]Values reported as the average over 20 structures. RMSD calculated by pairwise superimposition of residues 13-38, 54-81 & 93-106.
[3]Determined using MolProbity (Davis *et al.*, 2007) for residues 11-107 of 20 structures.

The final ensemble of MxiG-N_{6-112} structures (**Figure 2.19B&C**) had no significant violations, a low RMSD and the majority of residues with favourable stereochemistry (**Table 2.3**), indicating the structures satisfy the input NMR restraints, are highly convergent and have good covalent geometry. In addition, the inter-residue NOEs that restrain the final structure were evenly distributed throughout the molecule (**Figure 2.19D**), indicating an unbiased model. The MxiG-N_{6-112} ensemble was deposited in the Protein Data Bank (PDB) with pdb id 2xxs and the model with the lowest target function (1.69 $Å^2$) was chosen as the top model for further analysis of the

MxiG-N_{6-112} structure. The calculated tertiary structure of MxiG-N_{6-112} confirmed the predicted secondary structure (**Figure 2.14**), with a topology of an antiparallel four-stranded β-sheet and a five-stranded β-sheet with a β-α-β motif (**Figure 2.19A**). A ribbon representation of the top model showed these β-sheets sit on top of each other in the protein fold to form a globular structure (**Figure 2.19E**). The MxiG-N_{6-112} ensemble is particularly well-defined within secondary structure elements, whilst the structures converge poorly at the termini and in a large loop region (**Figure 2.19C&F**).

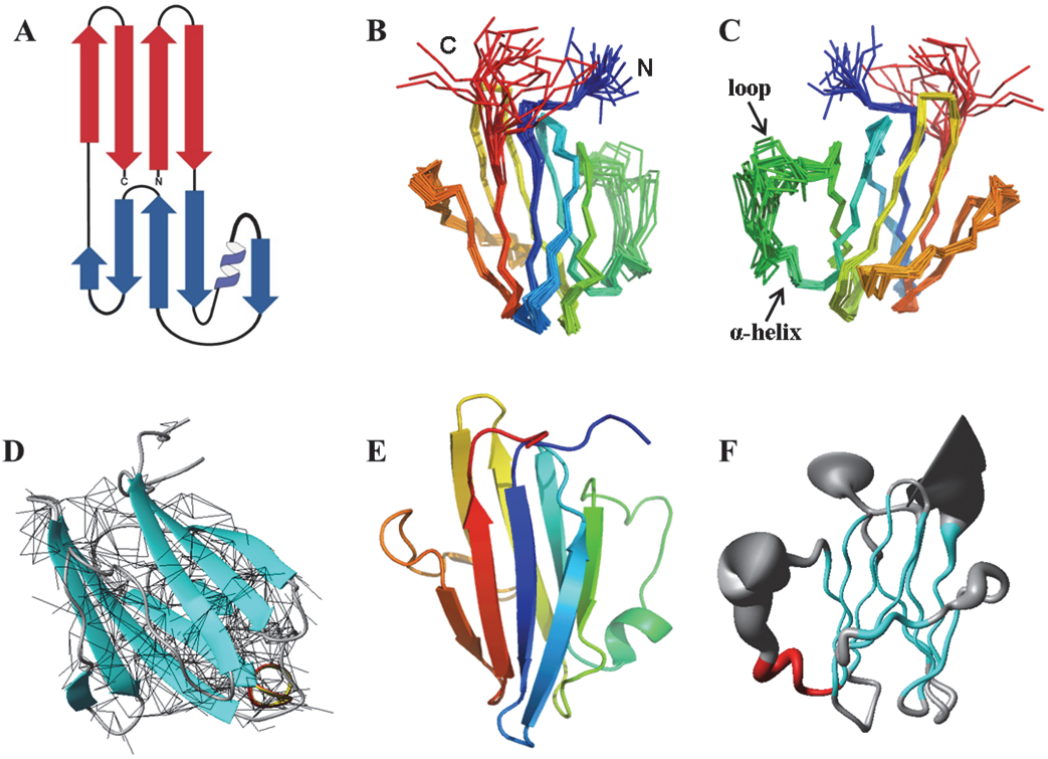


Figure 2.19 *The NMR structure of MxiG-N_{6-112}.*

A Topology of MxiG-N_{6-112} showing β-sheet 1 (red) and β-sheet 2 (blue).

B & C Backbone superimposition of the final 20 structures by residues 13-38, 54-81 and 93-106. Ribbon structures are coloured from the N-terminus (blue) to the C-terminus (red) and the termini, α-helix and a large loop are indicated. **B** View from β-sheet 1 and **C** view from β-sheet 2.

D Cartoon representation of the top NMR model, showing distance restraints derived from NOEs as black lines. β-strands are shown in cyan and the α-helix in red.

E Cartoon representation of the top NMR model viewed from β-sheet 1, coloured from the N-terminus (blue) to the C-terminus (red) and viewed from β-sheet 1.

F Sausage representation of the final NMR ensemble viewed from β-sheet 2 calculated using MOLMOL-2.0. Cylinder thickness is proportional to the backbone RMSD. β-sheets are shown in cyan and the α-helix in red.

Table 2.3 Summary of restraints and structural statistics for MxiG-N_{6-112}	
Construct Information	
Residues	107
Atoms	1694
Number of Restraints	
NOEs used in structure calculation:	
Sequential	355
Medium Range (2-4 residues apart)	75
Long range (≥5 residues apart)	408
Total inter-residue NOEs	838
Dihedral angles from TALOS+ (Φ and Ψ)	183
Hydrogen bonds	37
Total number of restraints	1058
Violations	
Number of violations:	
Distance restraints (≥0.5 Å)	0
Dihedral angle restraints (≥5°)	0
Violations (average over 20 structures):	
Distance restraints (Å)	0.010 ± 0.001
Dihedral angle restraints (degrees)	0.003 ± 0.001
Target function (Å^2)	2.06 ± 0.16
Precision[1]	
Backbone heavy atoms (Å)	0.52 ± 0.07
All heavy atoms (Å)	1.28 ± 0.11
Structure Quality[2]	
Residues in favoured regions	94.0%
Residues in allowed regions	99.9%

[1]Values reported as the average over 20 structures. RMSD calculated by pairwise superimposition of residues 13-38, 54-81 & 93-106.
[2]Determined using MolProbity (Davis *et al.*, 2007) for residues 11-107 of 20 structures.

Solvent-exposed and slowly-exchanging amides previously identified by CLEANEX-PM and D_2O exchange $^{1}H,^{15}N$-HSQC spectra respectively (**Figure 2.13**) were mapped on to the structure of MxiG-$N_{6\text{-}112}$ (**Figure 2.20**). As predicted, residues with solvent-exposed amides were at the termini or in loops and turns, whilst slowly-exchanging backbone amides were generally involved in hydrogen bonding within secondary structure elements. Furthermore, β-strands with protected amides along their length were generally found within the centre of a β-sheet, whilst those with alternating fast and slow exchanging amides were found at the edge. The slowly-exchanging backbone amides of N33 and N54 were seen to be within hydrogen bonding distance of each other in the tertiary structure, allowing this restraint to be imposed in later rounds of structure refinement. In addition, as the CLEANEX-PM spectrum in particular was not used to derive any input restraints for the structure calculation, the correlation between predicted solvent-exposed amides and the structure of MxiG-$N_{6\text{-}112}$ served to validate the final model.

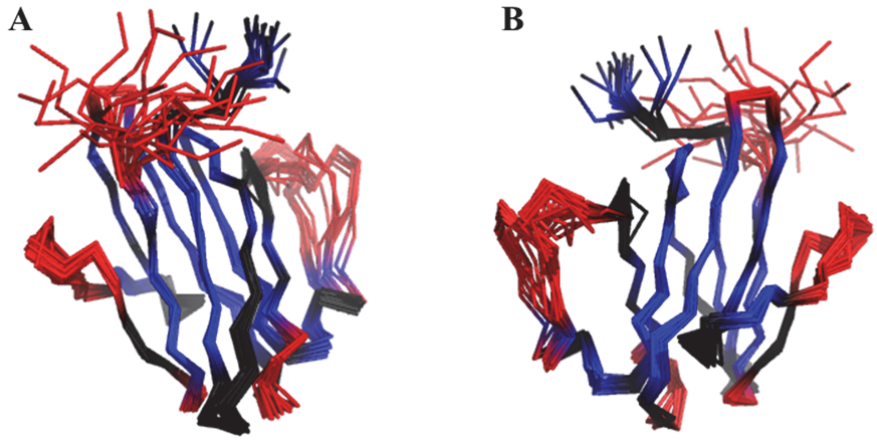


Figure 2.20 *Solvent-exposed and slowly-exchanging amides mapped onto the MxiG-$N_{6\text{-}112}$ structure*.

Slowly-exchanging residues corresponding to peaks with a data height >2 x 10^6 in the D_2O exchange spectrum are shown in blue. Solvent-exposed residues corresponding to peaks with a data height >1 x 10^7 in the CLEANEX-PM spectrum are shown in red (see **Figure 2.13**). Residues shown in black have peaks with a data height below the threshold in both spectra. **A** view from β-sheet 1 and **B** view from β-sheet 2.

2.6 Backbone Dynamics of MxiG-N_{6-112}

In order to fully understand the solution behaviour of MxiG-N_{6-112}, the internal motions of the backbone on the ps-ns timescale were probed. Firstly, the steady-state ^{1}H-^{15}N NOE was measured for each residue by recording ^{1}H,^{15}N-HSQC spectra with and without saturation of ^{1}H magnetisation and taking the ratio of peak intensities (Cavanagh, 2007) (**Figure 2.21A**). Although a large portion of the backbone is rigid with NOE intensities close to 0.82, residues F46-F51 and M108-K112 yield values <0.5, indicating the large loop region (**Figure 2.19C**) and the C-terminus respectively are regions with significant motion. This explained why these regions are poorly convergent in the MxiG-N_{6-112} ensemble (**Figure 2.19F**).

In order to characterise the backbone motions of MxiG-N_{6-112} further, a series of ^{1}H,^{15}N-HSQC spectra were measured using the inversion-recovery technique (Vold *et al.*, 1968) with relaxation periods of 10, 50, 100, 200, 300, 600 and 900 ms. The exponential decay of peak heights was fitted using RELAXFIT (Fushman *et al.*, 1997) to determine the spin-lattice relaxation rate constant for longitudinal relaxation (R1). Similarly a second series of ^{1}H,^{15}N-HSQC spectra were measured using the Carr-Purcell-Meiboom-Gill (CPMG) technique (Carr & Purcell, 1954, Meiboom & Gill, 1958) with relaxation periods of 60, 90, 120, 200 and 300 ms to extract the spin-spin relaxation rate constant for transverse relaxation (R2). The overall diffusion tensor for MxiG-N_{6-112} was determined from the R1, R2 and ^{1}H-^{15}N NOE values and the structure coordinates of the NMR model using ROTDIF (Walker *et al.*, 2004). Flexible regions identified from ^{1}H-^{15}N NOE values < 0.65 were omitted for this purpose. An axially symmetric model for the diffusion tensor (τ_c = 6.21 ns, D/D = 1.18, $\alpha = 130^{o}$, $\beta = 54^{o}$) was deemed most appropriate from the NH vector distribution and F-test results, so was subsequently used to determine values for the residue-specific order parameter (S^2) with DYNAMICS (Fushman *et al.*, 1997). **Figure 2.21B** shows the values of S^2 for the 90 residues that had good fits for R1, R2 and were fit with spectral density models A (17%), B (77%) or C (6%) (Fushman *et al.*, 1997).

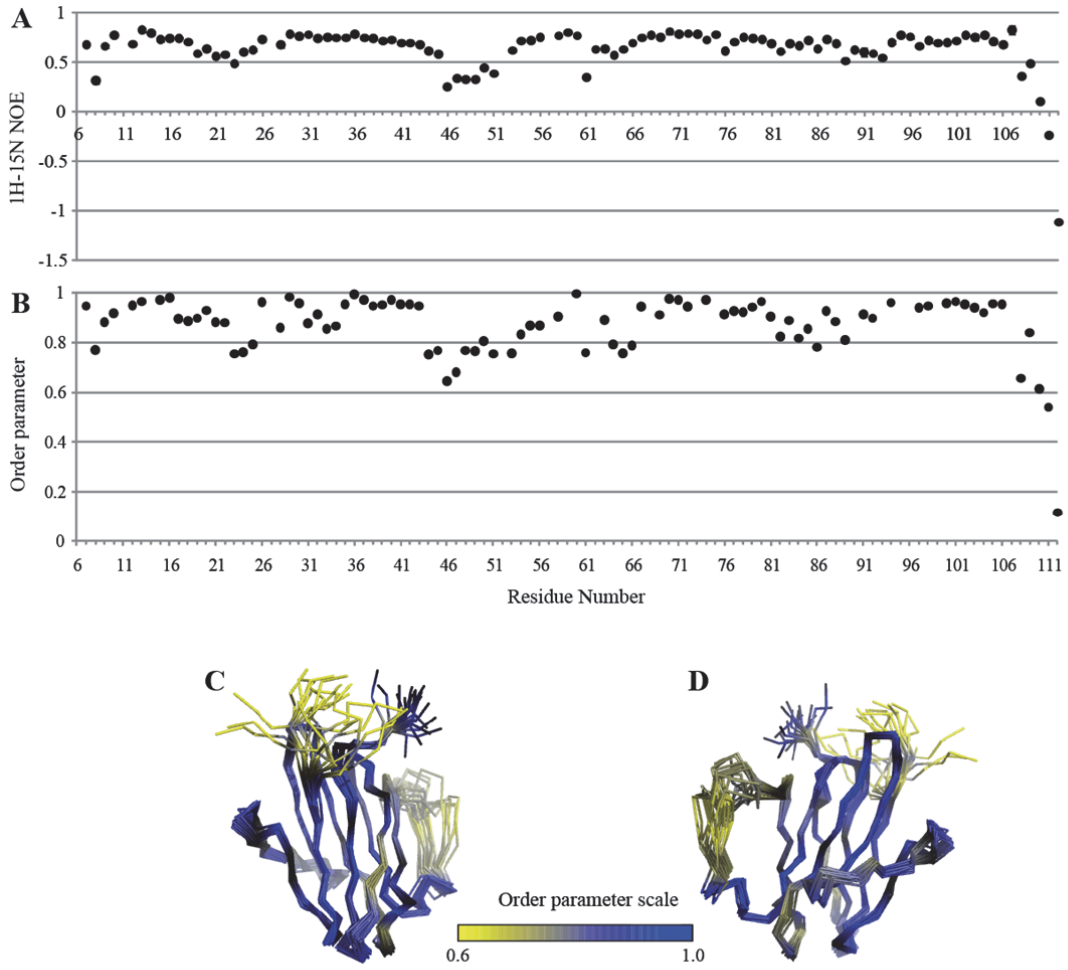


Figure 2.21 *The backbone dynamics of MxiG-N_{6-112}.*

A ^{1}H-^{15}N-NOE measurements for backbone amides. The steady-state NOE was calculated for each residue by taking the ratio of peak intensities from ^{1}H,^{15}N-HSQC spectra recorded with (NOE) and without (noNOE) saturation of ^{1}H magnetization.

B Calculated order parameter (S^2) for 84% backbone amides.

C & D Mapping of the order parameter on to the NMR structure for MxiG-N_{6-112}. The backbone is coloured in a continuous spectrum according to the scale bar shown, with the most flexible regions with the lowest order parameters in yellow and the most rigid regions with the highest order parameters in blue. Those residues for which no order parameter was fitted are shown in black. **C** view from β-sheet 1 and **D** view from β-sheet 2.

Mapping S^2 values on to the structure of MxiG-N_{6-112} (**Figure 2.21C&D**) confirms that the large loop region encompassing residues L44-E53 and residues M108-K112 at the C-terminus are highly mobile, in agreement with ^{1}H,^{15}N NOE values (**Figure 2.21A**). Additionally, residues G23-E25 and S61-I66 appear to show flexibility. These regions all include backbone amides in

fast exchange with the solvent (**Figure 2.13, 2.20**), indicating they also undergo conformational fluctuations on a much slower timescale. However, residues D88-S95 appear to be rigid yet solvent-exposed, suggesting additional intramolecular interactions and conformational restrictions aside from those imposed by hydrogen bonding are holding this loop in a fixed position.

2.7 Structural Analysis of MxiG-N_{6-112}

MxiG-N_{6-112} – A New Model System for NMR?

Further advances in NMR rely on evermore sophisticated spectrometers, pulse sequences and methods for isotopic labelling. Lysozyme (14.7 kDa), protein G (6.5 kDa) and ubiquitin (8.5 kDa) are favoured proteins for testing new experimental techniques, due to their optimal size, shape and spectral quality. The structure determination of MxiG-N_{6-112} has highlighted the suitability of the domain for NMR experiments. MxiG-N_{6-112} has a molecular weight of 12.2 kDa, forms a compact globular structure (**Figure 2.19**) and can be purified at concentrations up to 6 mM. Furthermore, NMR spectra for the protein show excellent peak dispersion in all dimensions, allowing almost complete assignment of ^{1}H, ^{15}N and ^{13}C backbone and side-chain atoms. Such qualities are reminiscent of proteins commonly used as model systems in the development of NMR methodology, therefore it was questioned whether MxiG-N_{6-112} has the potential to be harnessed as a new model system.

For MxiG-N_{6-112} to be useful as a model protein, it must be amenable to those NMR experiments where sensitivity is a major limitation for many proteins. The direct detection of 20 hydrogen bonds for MxiG-N_{6-112} in a 2D long-range H(N)CO spectrum (**Figure 2.17**) indicated its properties facilitate use in such experiments.

Proteins are not static molecules, therefore a picture of their mobility is required to complement structure determination in order to fully understand their behaviour. As the most flexible regions

of a protein often correlate with the most functionally important regions, a key area of research is the development of NMR experiments to probe dynamics. Ironically the favoured model proteins often have a rigid backbone and relatively few flexible regions, with only 14% of residues in ubiquitin having S^2 <0.8 for example (Simorellis & Flynn, 2006). This suggests the current model systems are sub-optimal for enhancing methods that probe the full dynamic range along a protein backbone. It was found that the backbone of MxiG-N_{6-112} has a wide range of motions on the ps-ns timescale and 21% of residues have S^2 <0.8 (**Figure 2.21**), indicating the protein dynamics are comparable to, if not better than, current model systems. This in combination with its other desirable properties indicates MxiG-N_{6-112} could be used as a tool for the illustration and development of NMR experiments.

Comparison with the Crystal Structures of MxiG-N_{1-126} and YscD-N_{2-108}

After determination of the NMR structure of MxiG-N_{6-112}, a 1.6 Å crystal structure of residues 1-126 of MxiG-N (MxiG-N_{1-126}; pdb id 4a4y) was published (Barison *et al.*, 2012). The crystal structure has an additional five residues at the N-terminus and 14 residues at the C-terminus relative to MxiG-N_{6-112} (**Figure 2.22A, arrow *a***), with the latter forming an additional two-stranded antiparallel β-sheet and presumably acting as the linker to the transmembrane portion of MxiG (Allaoui *et al.*, 1995). Despite these addendums, an overlay of the structures shows close agreement with the nine β-strands of MxiG-N_{6-112}, reinforced by a low C^{α} RMSD value of 1.45 Å over 85 atoms (**Figure 2.22A**). The MxiG-N_{1-125} model could not be traced inside the electron density between residues F46-F51 (**Figure 2.22A, arrow *b***), which can be explained by the high backbone mobility in this region of MxiG-N_{6-112} (**Figure 2.21**). There is a minor structural discrepancy between the structures in the D88-S95 loop region (**Figure 2.22A, arrow *c***) that is accounted for by the formation of crystal contacts between MxiG-N_{1-125} molecules and not by backbone dynamics, as discussed in the next chapter. Therefore the crystal structure of MxiG-N_{1-126} generally serves to validate the solution structure of MxiG-N_{6-112}.

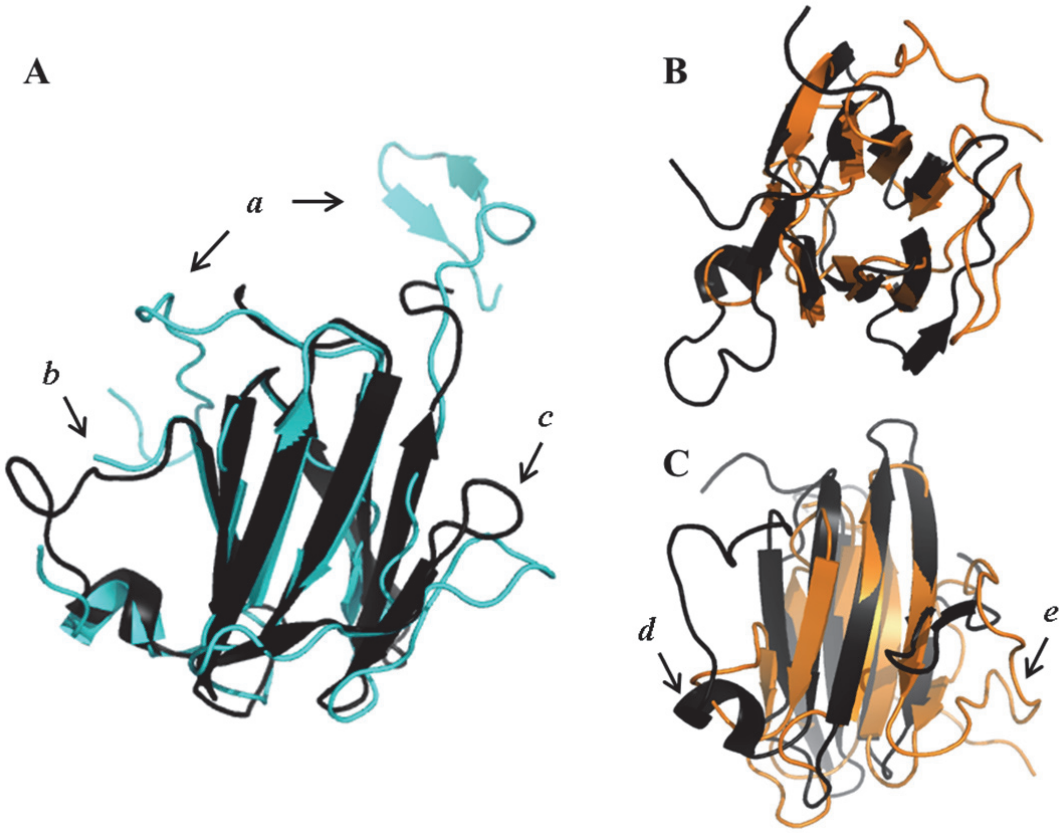


Figure 2.22 *Comparison of the NMR structure of MxiG-N_{6-112} with high resolution structures of homologous T3SS proteins*.

A Cartoon representation of the overlay of MxiG-N_{6-112} (black) with the crystal structure of MxiG-N_{1-126} (cyan; pdb id 4a4y)(Barison *et al.*, 2012) viewed from β-sheet 2 and calculated using PyMol. The C^{α} RMSD over 85 atoms is 1.45 Å, calculated using Superpose (Krissinel & Henrick, 2004). ***a*** shows additional structure at the termini of MxiG-N_{1-126}, ***b*** shows the loop region for which there is no electron density in MxiG-N_{1-126} and ***c*** shows the structural difference in the loop between β7 and β8.

B & C Cartoon representation of the overlay of MxiG-N_{6-112} (black) with the crystal structure of YscD-N_{2-108} (orange; pdb id 4a0e)(Lountos *et al.*, 2012) calculated using SSM within Superpose. The C^{α} RMSD over 74 atoms is 2.83 Å.
B View from the termini showing the superimposition of β-strands.
C View from β-sheet 2 showing ***d*** poor overlay in the region between β3 and β4 and ***e*** the additional β-hairpin in YscD-N_{2-108} between β7 and β8 of MxiG-N_{6-112}.

In addition, a 2.6 Å crystal structure was subsequently determined for residues 2-108 of the N-terminal domain of YscD (YscD-N_{2-108}; pdb id 4a0e), the equivalent domain to MxiG-N in the *Y. pestis* T3SS (Lountos *et al.*, 2012). Although MxiG-N_{6-112} and YscD-N_{2-108} only have a sequence identity of 9% **(Lountos *et al.*, 2012)**, superimposition via secondary structure matching (SSM) shows a very similar β-strand arrangement (**Figure 2.22B**), with a C^{α} RMSD of 2.83 Å over 74 atoms. However, there are some significant structural differences, particularly in loop regions. For example, between β3 and β4 YscD-N_{2-108} has no α-helix, a much smaller

loop with a different conformation and a much smaller β3 strand (**Figure 2.22C, arrow *d***). In addition, YscD-N_{2-108} has a much larger insertion after β7 that includes two further small β-strands (**Figure 2.22C, arrow *e***). Therefore, the core structure of MxiG-N appears to be conserved across homologous domains in other T3SSs, with some minor variation in intervening loop regions.

Positioning of MxiG-N_{6-112} within the T3SS

As outlined in the introduction, the stoichiometry of the IMR is a controversial topic within the field, with current models favouring 12 or 24 subunit rings (Yip *et al.*, 2005, Spreter *et al.*, 2009, Hodgkinson *et al.*, 2009, Schraidt *et al.*, 2010). All previously determined high-resolution structures of basal body components are of periplasmic domains (Spreter *et al.*, 2009, Yip *et al.*, 2005, Crepin *et al.*, 2005). Therefore, all docking into available EM maps has been focused on a region of density that corresponds to a hetero-oligomeric protein complex (Hodgkinson *et al.*, 2009, Spreter *et al.*, 2009, Schraidt *et al.*, 2010, Sanowar *et al.*, 2010). The relatively low resolution of available EM density maps and a lack of understanding of intermolecular interactions make placing these periplasmic domains within the context of the intact NC ambiguous. Additionally as there are only high resolution structures for MxiM (Lario *et al.*, 2005, Okon *et al.*, 2008) available for the *S. flexneri* basal body, docking has had to rely on ring models from *Salmonella* and *Escherichia* (Hodgkinson *et al.*, 2009), ignoring any species-specific variation in T3SS organisation. Therefore, previous estimates of the stoichiometry of the basal body are understandably ambiguous.

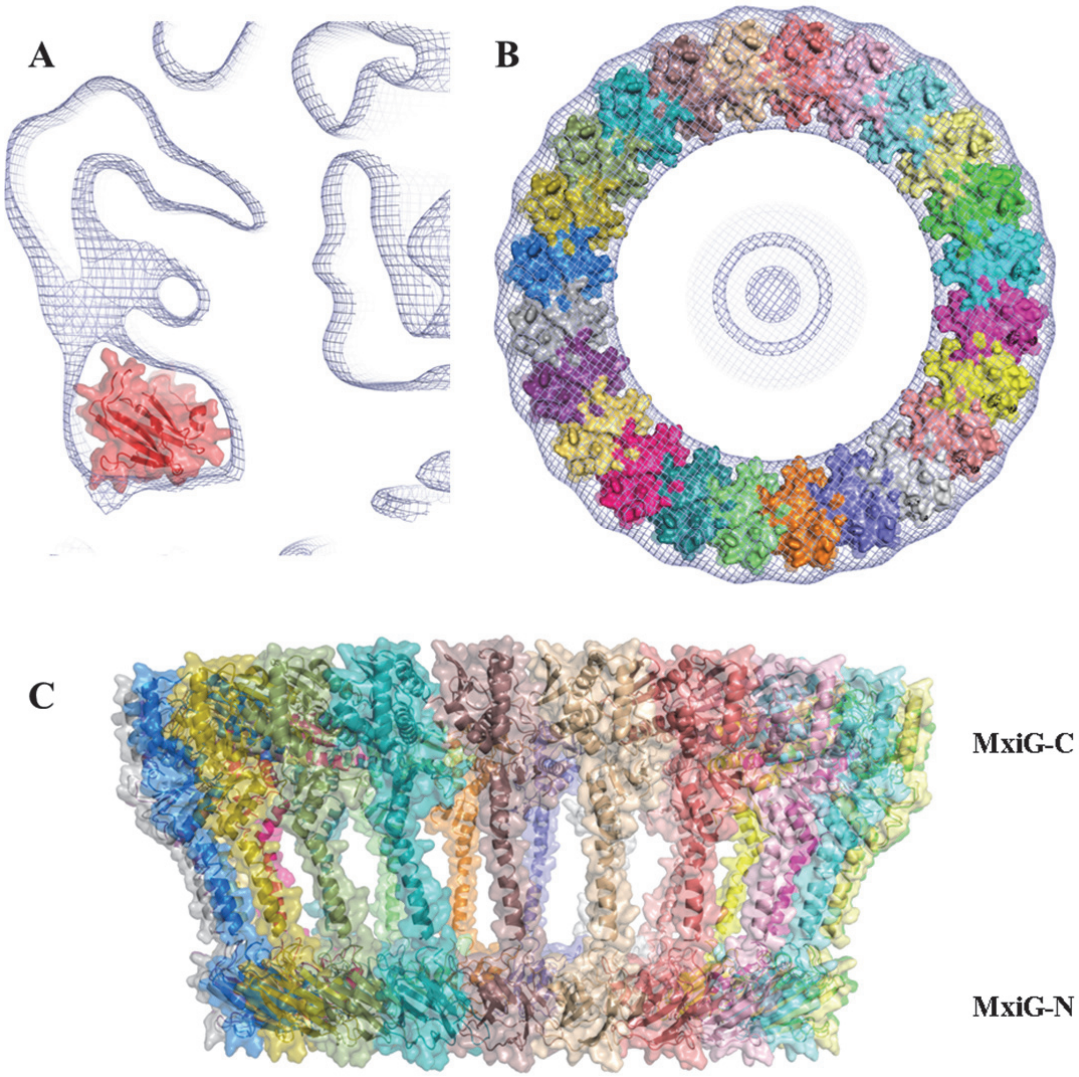


Figure 2.23 *Positioning of MxiG-N_{6-112} in the C24 S. flexneri NC EM map*.

A A position of MxiG-N_{6-112} (red) in the density for the leg domains contoured to 1σ (Hodgkinson *et al.*, 2009), restraining the C-terminus with respect to the likely density for the transmembrane region.

B View of the position of MxiG-N_{6-112} as described in **A** symmetrised to give 24 evenly arranged copies within the likely density for the MxiG-N ring. The NC base is viewed from the cytoplasmic side.

C Model for a 24 subunit MxiG ring. MxiG-N is represented by MxiG-N_{6-112}, whilst MxiG-C is a SCWRL model (Altschul *et al.*, 1990, Canutescu *et al.*, 2003) based on the crystal structure of $PrgH_{170-362}$ (pdb id 3gr0). The transmembrane region is depicted as an arbitrary 20 amino acid helix. Whilst the model shows the likely volume occupied by a 24-member ring, the positioning of the MxiG subunits with respect to each other represents one of many possible conformations given the current data.

MxiG-N is the predominant basal body structure in the cytoplasm, so positioning of MxiG-N_{6-112} had the potential to provide a relatively unambiguous estimate of the maximum number of copies of MxiG within the IMR. Indeed the *S. flexneri* EM map shows clear density for leg domains at the base of the NC (**Figure 2.24A**) (Hodgkinson *et al.*, 2009) and there is good evidence this region corresponds to the N-terminal domain of PrgH in *S. typhimurium* (Schraidt *et al.*, 2010). Positioning of MxiG-N_{6-112} within this density in the C24 *S. flexneri* NC EM map (EMDB 6391)(Hodgkinson *et al.*, 2009) was performed by Dr Steven Johnson. Due to the lack of asymmetry in the domain and the low resolution of the map (25 Å), the precise orientation of the domain could not be distinguished, so the position was defined by restraining the C-terminus with respect to the likely density for the transmembrane region. There is good agreement between the size and shape of MxiG-N_{6-112} and the density envelope (**Figure 2.23A**), supporting evidence that this region corresponds to MxiG-N. When expanded to 24 copies using the symmetry operators of the map (**Figure 2.23B**), this model for the cytoplasmic ring agreed well with the experimental map, having a correlation coefficient of 78% and only minor structural clashes between subunits. This is in agreement with proposals of a 24-fold symmetry in the IMR (Yip *et al.*, 2005, Hodgkinson *et al.*, 2009, Schraidt *et al.*, 2010), allowing a model for the MxiG ring to be constructed (**Figure 2.23C**). Subsequently a precise docking of the periplasmic domains of the basal body components into a 10 Å reconstruction of the *S. typhimurium* NC was published (Schraidt & Marlovits, 2011). This also supported a model with 24 copies of MxiG homologues in the IMR of the T3SS. However higher resolution EM maps are still required to precisely define the orientation of MxiG-N, as this region is blurred in the *Salmonella* reconstruction, presumably due to averaging of the different observed conformational states of MxiG-N (Schraidt & Marlovits, 2011).

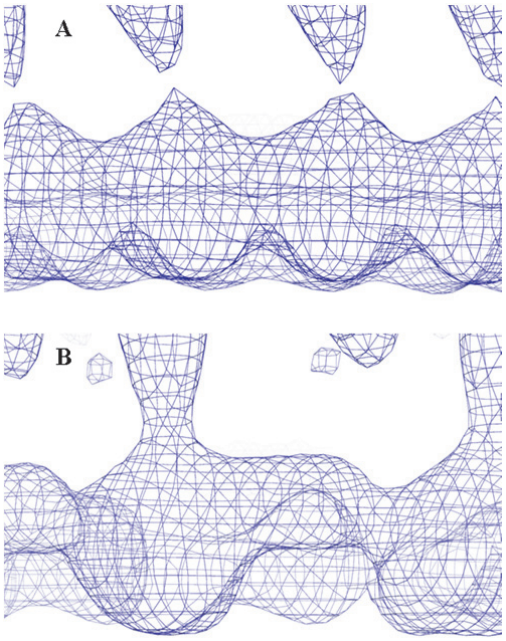


Figure 2.24 *Regions of the S. flexneri NC EM maps corresponding to portions of the MxiG-N ring*.

A A region in the C24 map showing two of the symmetrically arranged leg domains expected to correspond to MxiG-N. Contoured to 4σ.

B An equivalent region in the C12 map highlighting the 12-fold symmetry of the MxiG-N ring.

The best resolution of features within the *S. flexneri* NC has been achieved by applying 12-fold symmetry during the EM reconstruction. This produced a map in which the outer portion of the IMR is seen to consist of 12 asymmetric dimers with pseudo-24-fold symmetry in the periplasmic region, leading to a model for the IMR where 12 copies of an asymmetric MxiG dimer are present (Hodgkinson *et al.*, 2009). However, the density corresponding to the cytoplasmic (leg) portion only displays 12-fold symmetry (**Figure 2.24B**). It was therefore unclear how to position 24 copies of MxiG-N_{6-112} into the C12 map. Whether this is a genuine feature of the *S. flexneri* T3SS, or an artefact of the current C12 map will require higher resolution reconstructions.

2.8 Perspectives

The NMR structure of MxiG-N_{6-112} was determined and validated by comparison to both experimentally determined NMR parameters and the subsequently determined crystal structure of MxiG-N_{6-112}. Furthermore, MxiG-N_{6-112} displayed a wide variety of properties optimal for NMR experiments, suggesting the domain could be a useful model protein for methodological developments. Comparison with the structure of $YscD_{2-108}$ indicated the core structure of MxiG-N_{6-112} is conserved between homologous T3SS domains. Finally, positioning of MxiG-N_{6-112} into EM density for the *Shigella* NC allowed evaluation of current models for IMR arrangement, supporting a model with 24 copies of MxiG.

3. Functional characterisation of MxiG-N

3.1 MxiG-N is an Essential Component of the T3SS

To ascertain whether MxiG-N is required for secretion by the T3SS, the coding sequence for MxiG-N_{6-112} was deleted from pUC19-*mxiG* (pAB13)(Allaoui *et al.*, 1995), a construct that affords constitutive expression of full-length MxiG. The deletion of residues F12-H104 of MxiG (MxiGΔN) to create pAB13-ΔN was achieved by site-directed mutagenesis (**Appendix 8.2.3**) using an annealing temperature (T_a) of 65°C, an elongation time (t_e) of 5 min and the primers shown in **Table 3.1**.

Table 3.1 Primers used to introduce the MxiGΔN mutation.

Forward	**Reverse**
5'- GTC TGA GGC AAA GAA CTC AAA TCT TGC ACC ACT TAA AAA TAT GAG GGA GGA TAA ATC GAG AGG C -3'	5'- GCC TCT CGA TTT ATC CTC CCT CAT ATT TTT AAG TGG TGC AAG ATT TGA GTT CTT TGC CTC AGA C -3'

S. flexneri strains were transformed with 100 ng of the appropriate plasmid (**Table 3.2**) by electroporation using a Gene Pulsar XCell (BioRad) and propagated overnight at 37°C on Congo red (CR) agar plates (0.1 mg/ml CR, 14 mg/ml agar, 30 mg/ml tryptic casein soy broth (TCSB)) with the indicated antibiotic at the appropriate final concentration (kanamycin 50 µg/ml, ampicillin 100 µg/ml). The ability of *Shigella* colonies to bind the small amphipathic dye CR on an agar plate is a preliminary indicator of T3SS assembly and activity (Meitert *et al.*, 1991, Parsot *et al.*, 1995). The negative control *mxiG*$^-$ gave white colonies (**Figure 3.1A**), indicating a defective T3SS. Expression of wild-type MxiG in *mxiG*$^{-/+}$ complemented this defect, giving red colonies (**Figure 3.1B**). The *mxiG*$^{-/\Delta N}$ strain gave white colonies on the CR agar plate (**Figure 3.1C**), indicating the T3SS was not fully formed in the absence of MxiG-N.

Table 3.2 *S. flexneri* strains used for overnight leakage and CR induction assays

Strain	Genotype (strain, plasmid)	Resistance	Reference
WT	Wild-type M90T, serotype 5a		(Sansonetti *et al.*, 1982)
mxiG^{-}	SF703	Kanamycin	(Allaoui *et al.*, 1995)
mxiG$^{-/+}$	SF703 + pAB13	Kanamycin, Ampicillin	(Allaoui *et al.*, 1995)
mxiG$^{-/\Delta N}$	SF703 + pAB13-ΔN	Kanamycin, Ampicillin	This study
mxiG$^{-/mutant}$	SF703 + pAB13-R39A/S61A/S63A	Kanamycin, Ampicillin	This study

Basal functionality of the T3SS can be tested by an overnight leakage assay (Parsot *et al.*, 1995). This was performed by inoculating 5 ml TCSB and the indicated antibiotic with colonies from WT, *mxiG*$^{-}$, *mxiG*$^{-/+}$ and *mxiG*$^{-/\Delta N}$ CR plates and growing the resulting cultures overnight at 37^{o}C. 1 ml of each culture was then harvested and a sample of the supernatant extracted for SDS-PAGE. The uninduced release of early effectors of *Shigella* virulence (**Table 1.5**) during assembly gives rise to a characteristic gel profile, as exemplified by that of WT and *mxiG*$^{-/+}$ strains (**Figure 3.1D**). In contrast, there was no overnight secretion from the *mxiG*$^{-}$ strain as the T3SS did not assemble correctly. The protein SepA, which is secreted independently of the T3SS and acts as a positive control both for the presence of the virulence plasmid and the functionality of other *Shigella* secretion pathways, was found in the supernatant from all three strains. The secretion profile for *mxiG*$^{-/\Delta N}$ was the same as for *mxiG*$^{-}$ (**Figure 3.1D**), indicating MxiG-N is required for T3SS assembly.

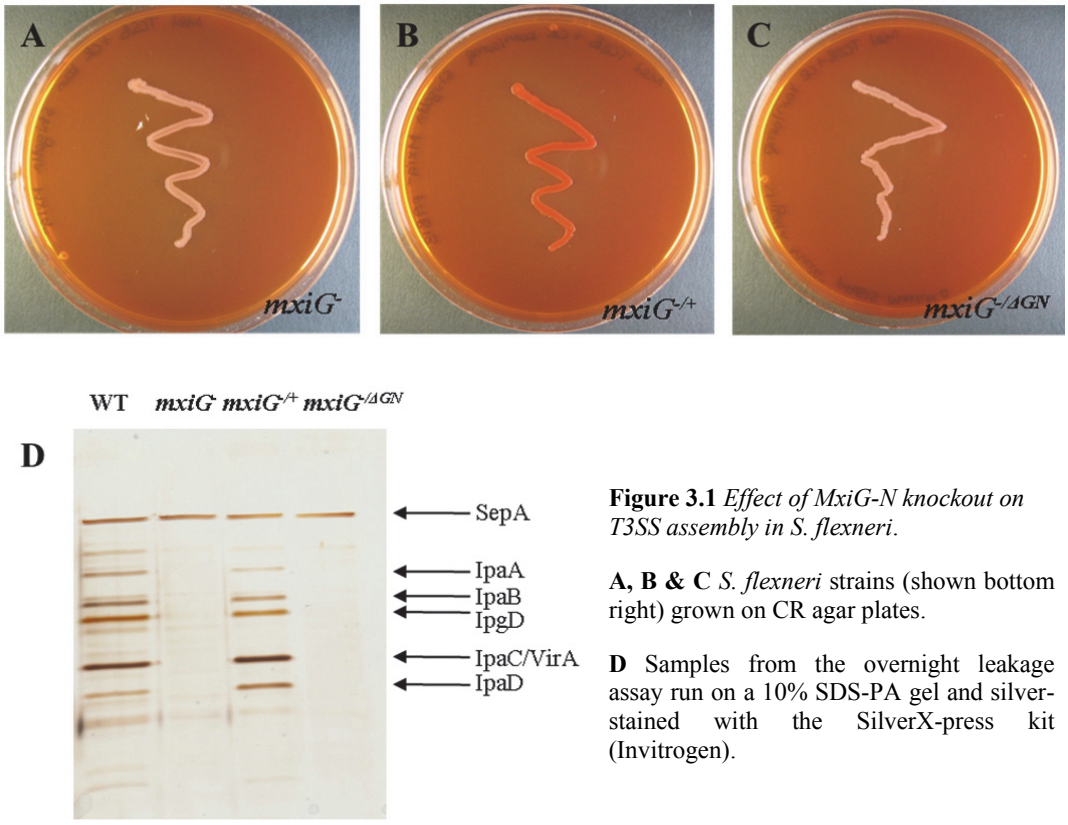


Figure 3.1 *Effect of MxiG-N knockout on T3SS assembly in S. flexneri.*

A, B & C *S. flexneri* strains (shown bottom right) grown on CR agar plates.

D Samples from the overnight leakage assay run on a 10% SDS-PA gel and silver-stained with the SilverX-press kit (Invitrogen).

It could be questioned whether deletion of MxiG-N impairs the membrane insertion signal, leading to mislocalisation of MxiGΔN to the cytosol and ultimately emulation of the *mxiG*- phenotype. Although cellular fractionation and isolation of MxiGΔN in the inner membrane fraction would be required for definitive proof, experiments involving MxiG and its homologues indirectly suggest this is unlikely to be the case. Firstly, intact NCs are able to be purified from *S. flexneri* and *S. typhimurium* strains where MxiG and PrgH respectively are N-terminally His-tagged (Zenk *et al.*, 2007, Schraidt *et al.*, 2010), suggesting that the signal for membrane insertion is more likely to be within the transmembrane region rather than at the N-terminus. Secondly, a YscD construct with the entire N-terminal domain deleted produces a dominant-negative phenotype when expressed in wild-type *Y. pestis* (Ross & Plano, 2011), indicating this truncated protein is membrane inserted and competing with wild-type YscD for incorporation in the T3SS. As YscD-$N_{2\text{-}108}$ was shown in the previous chapter to be structurally similar to MxiG-$N_{6\text{-}112}$ (**Figure 2.22B&C**), these results suggest the core structure of MxiG-N is essential for correct assembly of the T3SS.

3.2 Does MxiG-N Function as a Canonical FHA Domain?

MxiG-N_{6-112} Has the Fold of a FHA Domain

A structural homology search of the PDB with the coordinates of the top model of MxiG-N_{6-112} using the DaliLite algorithm (Holm & Rosenström, 2010), identified structures of forkhead-associated (FHA) domains, with the best 50 matches having a Z-score of 8.3-5.9. The closest structural homologue is the FHA domain of CT664 from *C. trachomatis* (Dali Z-score 8.3; pdb id 3gqs), with a C^{α} RMSD of 2.37 Å over 72 atoms. There is a particularly good alignment of β-strands when superimposed on MxiG-N_{6-112} (**Figure 3.2A**), although some variability exists between the structures in loop regions (**Figure 3.2B**).

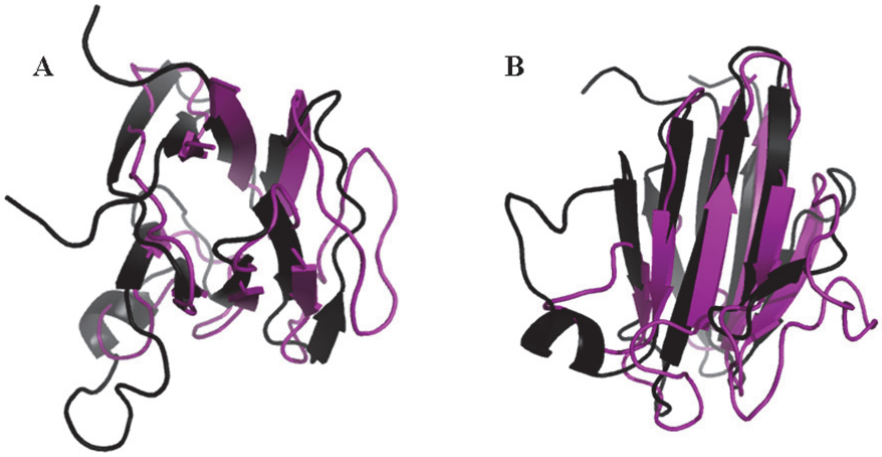


Figure 3.2 *Comparison of the structure of MxiG-N_{6-112} with its closest structural homologue CT664 from C. trachomatis*.

Ribbon diagram overlays of MxiG-N_{6-112} (black) and CT664 (purple; pdb id 3gqs) calculated using DaliLite (Holm & Rosenström, 2010). The C^{α} RMSD is 2.37 Å over 72 atoms, calculated using SSM within Superpose.
A View from the termini showing the superimposition of β-strands.
B View from β-sheet 2 showing variability of loop regions.

FHA domains act as phospho-threonine (P-Thr) binding modules in cell-signalling, regulating a diverse range of functions in both prokaryotic and eukaryotic proteins (Tsai, 2002). The FHA fold is comprised of ~75 amino acids and contains three highly conserved motifs; a GR dipeptide at the end of β3, an SXXH sequence before β5 and an NG motif before β7. These motifs have been shown to be crucial for P-Thr binding and binding site stabilisation. (Pallen *et*

al., 2002, Hammet *et al.*, 2003) The MxiG-N_{6-112} sequence only has 4/6 conserved residues present in the expected positions relative to secondary structure: G38, R39, S61/S63 and G84 (**Figure 3.3A, yellow**), with ambiguity over which of the serine residues corresponds to the conserved residue. To ascertain the structural alignment of R39 and S61/63 with their highly conserved counterparts, the top MxiG-N_{6-112} model was superimposed on functional FHA domains from *Saccharomyces cerevisiae* Rad53 (Dali Z-score 6.0; pdb id 2jqi)(Lee *et al.*, 2008), *Mycobacterium tuberculosis* EmbR (Dali Z-score 7.1; pdb id 2fez)(Alderwick *et al.*, 2006) and human Ki67 (Dali Z-score 7.1; pdb id 2aff)(Byeon *et al.*, 2005) using DaliLite. The RMSDs for the loop regions of interest suggest these are well-ordered portions of the MxiG-N_{6-112} ensemble, so the top model is likely to be representative of the native conformation. Although R39 and S61/63 are in the expected loop regions, alignment of their side-chains with those of the conserved residues is poor, particularly for S61 and S63. (**Figure 3.3B&C**) Therefore it would seem unlikely they could assume their traditional role in direct coordination of phosphate (Hammet *et al.*, 2003, Durocher *et al.*, 2000). G84 is not found within a NG motif and is structurally distant from the equivalent glycine residues of other FHA domains, suggesting G84 may not be a structurally equivalent residue. In addition MxiG-N_{6-112} lacks the conserved histidine and asparagine residues required to stabilise the binding site architecture (Durocher *et al.*, 2000). It is therefore unlikely MxiG-N has the residue arrangement required to form a functional phosphate-binding pocket.

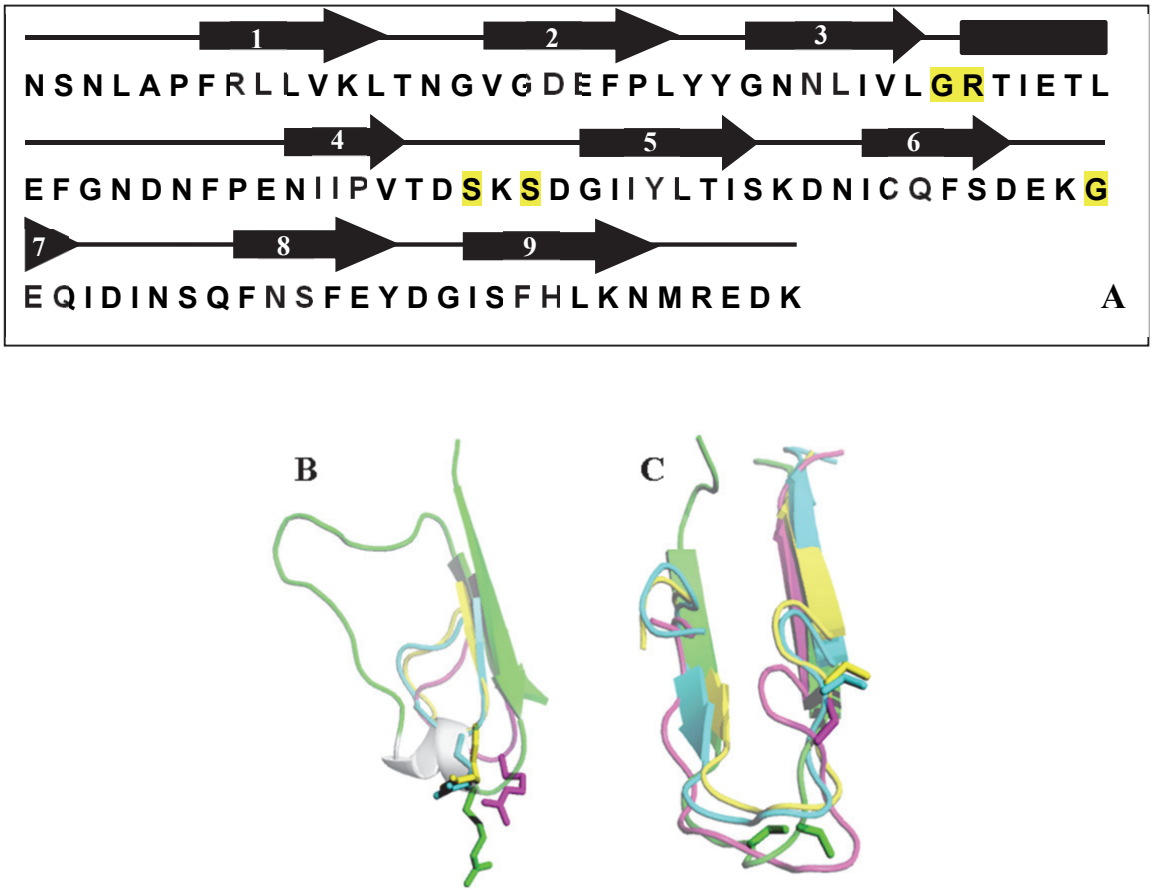


Figure 3.3 *Analysis of conserved FHA domain residues in MxiG-N_{6-112}.*

A Identification of conserved FHA domain motifs in the sequence of MxiG-N_{6-112}. Residues correlating with these motifs are highlighted in yellow and their positions with respect to secondary structure elements are shown.

B & C Overlay of ribbon diagrams of MxiG-N_{6-112} (green), Rad53-FHA1 (magenta; pdb id 2jgi), EmbR (yellow; pdb id 2fez) and Ki67 (cyan; pdb id 2aff).
B Structural comparison of loop regions containing the conserved arginine residues. Residues 31-53, 66-75, 309-318 and 28-37 and the side-chains of R39, R70, R312 and R31 are shown for MxiG-N_{6-112}, Rad53-FHA1, EmbR and Ki67 respectively. Residues 40-43 of MxiG-N_{6-112} are coloured grey for clarity. The heavy atom RMSD for residues 37-40 of the MxiG-N_{6-112} ensemble is 1.29 ± 0.40 Å.
C Structural comparison of loop regions containing the conserved serine residues. Residues 51-71, 73-91, 314-335 and 32-56 and the side-chains of S61 and S63, S85, S326 and S45 are shown for MxiG-N_{6-112}, Rad53-FHA1, EmbR and Ki67 respectively. The heavy atom RMSD for residues 59-63 of the MxiG-N_{6-112} ensemble is 1.08 ± 0.34 Å.

MxiG-N_{6-112} Does Not Interact with P-Thr *in vitro*

To investigate whether MxiG-N_{6-112} possesses a functional phosphate-binding site, ^{15}N-labelled protein was expressed and purified as outlined in the previous chapter. SEC was performed using a non-phosphate containing buffer (100 mM Tris-HCl (pH7.5), 150 mM NaCl, 1 mM DTT). ^{1}H,^{15}N-HSQC spectra were then collected before and after titration of 138 µM MxiG-N_{6-112} with a buffered solution of 50 mM P-Thr (Bachem). There was no change in ^{1}H,^{15}N-HSQC peak positions (**Figure 3.4**), indicating no interaction between MxiG-N_{6-112} and a high concentration of P-Thr. Therefore MxiG-N is unlikely to be involved in cell-signalling via phospho-recognition.

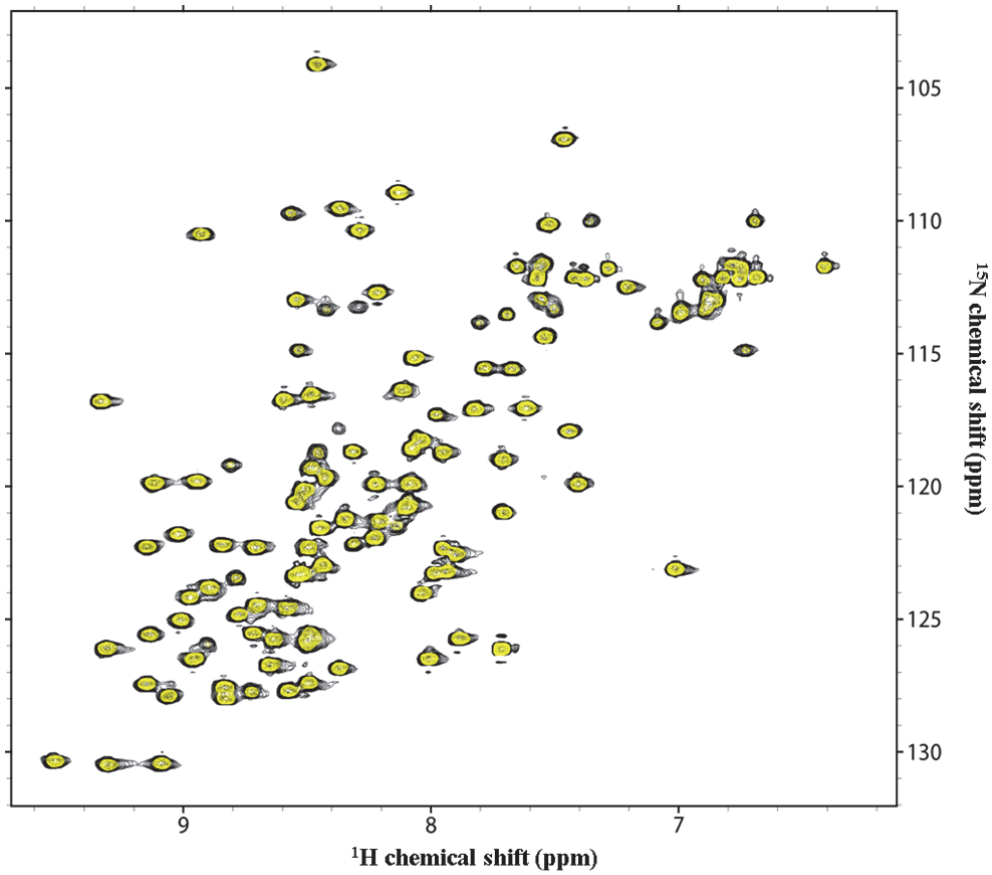


Figure 3.4 *Titration of ^{15}N-MxiG-N_{6-112} with 50 mM P-Thr.*

^{1}H,^{15}N-HSQC spectra of MxiG-N_{6-112} before (black) and after (yellow) addition of 50 mM P-Thr.

Conserved P-Thr Binding Residues Are Not Required *in vivo*

Conserved binding residues are required for the functionality of FHA domain containing proteins (Molle *et al.*, 2003) and would be expected to play a role in the T3SS if MxiG-N has a canonical FHA P-Thr recognition site. To investigate this, the R39A/S61A/S63A triple mutation was introduced into the pAB13 plasmid by two rounds of site-directed mutagenesis (**Appendix 8.2.3**), using T_a of 55^{o}C, t_e of 5 min and the primers outlined in **Table 3.3**. This pAB13-R39A/S61A/S63A construct was subsequently transformed into the *mxiG*$^{-}$ strain of *S. flexneri* to form *mxiG*$^{-/mutant}$ (**Table 3.2**).

Table 3.3 Primers used to create the R39A/S61A/S63A mutation

Mutation	Forward	Reverse
R39A	5'- CTT AAT TGT TCT CGG AGC GAC GAT AGA AAC ACT TGA G -3'	5'- CTC AAG TGT TTC TAT CGT CGC TCC GAG AAC AAT TAA G -3'
S61A/S63A	5'- CCT GTA ACA GAT GCT AAA GCG GAT GGT ATA ATT TAT TTG ACC ATA AG -3'	5'- CTT ATG GTC AAA TAA ATT ATA CCA TCC GCT TTA GCA TCT GTT ACA GG -3'

The *mxiG*$^{-/mutant}$ strain gave red colonies on a CR agar plate, providing a preliminary indication of a functional T3SS (**Figure 3.5A,B&C**). Moreover the *mxiG*$^{-/mutant}$ strain displayed wild-type secretion of early effectors in the overnight leakage assay (**Figure 3.5D**). These results suggested the mutations had not hindered the assembly of the T3SS, although a defect in secretion was still possible.

Addition of CR to the cell medium mimics host-cell contact and induces secretion of early effectors, leading to an assay to probe the basal functionality of the T3SS (Bahrani *et al.*, 1997). Cultures of *S. flexneri* strains from the overnight cultures were used to inoculate 10 ml TCSB with the appropriate antibiotic added (**Table 3.2**). These cultures were grown at 37^{o}C in a

shaking incubator until OD_{600} of ~1.0 was reached. Cells were then harvested and resuspended in 1 x PBS to a final OD_{600} of 5.0. This cell resuspension was supplemented with 0.2 mg/ml CR and incubated at 37°C for 15 min followed by 5 min on ice. As for the overnight leakage assay, cells were then harvested and a sample of the supernatant extracted for SDS-PAGE. The *mxiG*$^{-}$ strain showed no secretion upon CR induction, whilst the secretion profile of the complemented *mxiG*$^{-/+}$ strain was the same as for the WT strain (**Figure 3.5E**). The secretion profile of *mxiG*$^{-/mutant}$ was also indicative of a wild-type phenotype, indicating the mutations did not prevent inducible secretion through the T3SS. Therefore, phospho-recognition at the conserved interaction site does not regulate any major step in T3SS assembly or function, implying that MxiG-N does not function as a canonical FHA domain.

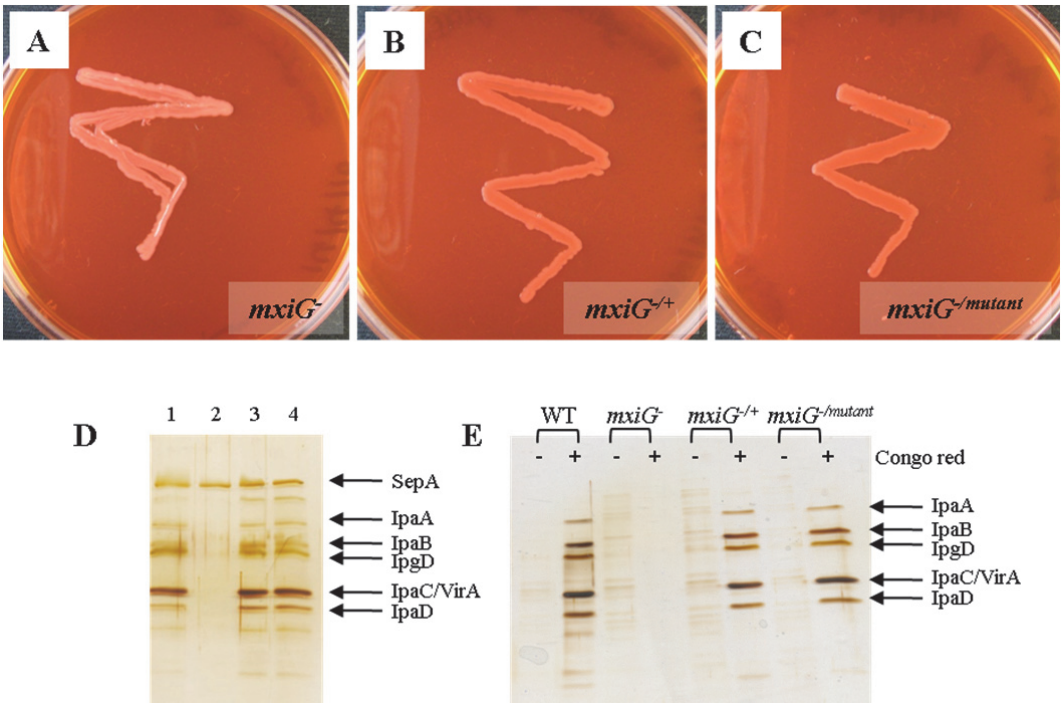


Figure 3.5 *Effect of mutation of conserved FHA domain residues on T3SS assembly and function in S. flexneri*.

A, B & C *S. flexneri* strains (shown bottom right) grown on CR agar plates.
D Samples from the overnight leakage assay run on a 10% SDS-PA gel and silver-stained. 1 WT, 2 *mxiG*$^{-}$, 3 *mxiG*$^{-/+}$ and 4 *mxiG*$^{-/mutant}$.
E Samples from the CR induction assay run on a 10% SDS-PA gel and silver stained. *S. flexneri* strains exposed to either no CR (-) or 0.2 mg/ml CR (+).

MxiG-N Has a Non-Canonical Function for a FHA Domain

The NMR structure of MxiG-N_{6-112} reveals it has the fold of a FHA domain. This protein fold was originally identified as a P-Thr binding module in eukaryotes, for which there is now a good understanding of binding characteristics and its diverse involvement in cell-signalling at the structural level (Hammet *et al.*, 2003, Mahajan *et al.*, 2008). A PSI-BLAST search using the Fha1 domain of the DNA-damage checkpoint kinase Rad53 from *S. cerevisiae* uncovered putative bacterial FHA domains implicated in many processes (Pallen *et al.*, 2002), however their role in bacterial signaling still remains largely uncharacterised. Of particular interest for this study was the classification of YscD-N from *Y. enterocolitica* as a FHA domain (Pallen *et al.*, 2002). This was subsequently verified by the crystal structure of YscD-N_{2-108} from *Y. pestis* (Lountos *et al.*, 2012), which overlays well with the FHA fold of MxiG-N_{6-112} (**Figure 2.22B**). Chlamydial species also have a FHA domain-containing protein within their T3SS gene cluster (Pallen *et al.*, 2002), which appears to be homologous to YscD (Betts *et al.*, 2008). Indeed CT664 from *C. trachomatis* was found to be the closest structural homologue of MxiG-N_{6-112} (**Figure 3.2**). Therefore a cytoplasmic FHA fold appears to be a conserved feature of the MxiG family of proteins, despite a lack of sequence homology.

Given the intimate involvement of a FHA domain in the regulation of the *P. aeruginosa* type VI secretion system (Mougous *et al.*, 2007), we questioned whether the T3SS has conserved a FHA fold for a similar canonical role in phosphorylation-mediated cell-signalling. Widespread sequence comparison of FHA domains has identified three highly conserved motifs, required for direct substrate binding and binding site stabilisation (Durocher *et al.*, 2000). It follows that the presence of these residues could be used to predict whether a FHA domain possesses a canonical binding pocket for P-Thr peptides. As MxiG-N lacks the full repertoire of conserved residues (**Figure 3.3A**), it would not be expected to have a phosphate-binding pocket. We confirmed this by comparing the MxiG-N_{6-112} solution structure with other FHA domains, which showed significant structural deviation in the substrate-binding loops (**Figure 3.3B&C**). A subsequent NMR titration indicated no interaction between MxiG-N_{6-112} and P-Thr (**Figure 3.4**)

and mutagenesis of conserved binding residues had no affect on the *in vivo* function of the T3SS (**Figure 3.5**). Therefore MxiG-N does not appear to function in cell-signalling via phospho-recognition, as would be expected for a canonical FHA domain.

Is this trait shared by the FHA domains of MxiG homologues? The two FHA domains within the cytoplasmic N-terminal region of Cpn0172 from *C. pneumonia* were found to interact with and be phosphorylated by the dual specificity kinase PknD (Johnson & Mahony, 2007). Although the physiological significance, if any, of this observation is unknown, it has been speculated from this that the chlamydial T3SS is regulated by reversible phosphorylation. However FHA-2 of Cpn0172 lacks the highly conserved arginine, indicating a novel interaction with phosphorylated PknD. Indeed mutagenesis of other conserved residues within the domain had no effect on the observed phosphorylation (Johnson & Mahony, 2007), suggesting Cpn0172, much like MxiG-N, does not have the canonical function of a FHA domain.

There is some evidence to suggest that the chlamydial T3SS most closely represents the common ancestor of T3SSs (Kim, 2001, Gophna *et al.*, 2003, Pallen *et al.*, 2005), so it could be speculated that there has been a progressive loss of canonical FHA function during T3SS evolution if it is not required for regulation. Indeed the T3SS gene cluster of chlamydial species encodes a serine/threonine protein kinase (STPK), a prerequisite for signalling via P-Thr recognition. In contrast, the only identified candidate for a STPK in *Shigella* spp. is the effector OspG, which is encoded on a distinct part of the virulence plasmid from *mxiG* and is likely to mediate its kinase activity within the host cell (Kim *et al.*, 2005). However, the FHA fold is found in MxiG homologues across a wide spectrum of T3SSs, suggesting evolutionary conservation for a non-canonical function. Thus, this domain appears to be an example of bacteria using a common structural scaffold for novel functionality. In addition, variation in domain function between T3SSs of different species may still exist, in the same way that regulation of secretion by threonine-phosphorylation does not appear to be a ubiquitous feature of type VI secretion systems (Bingle *et al.*, 2008). Nonetheless this alternative functionality of

MxiG-N is essential for assembly of the T3SS (**Figure 3.1**) and is explored further in the remainder of this chapter.

3.3 Investigation of Alternative Interaction Partners for MxiG-N

Can MxiG-N Pull Down *in vivo* Interaction Partners?

Cloning of pBAD-mxiGN$_{2-125}$ and pGEX-mxiGN$_{2-125}$

In order to be able to pull-down the *in vivo* interaction partners of MxiG-N at a detectable level, constructs were designed that would allow both inducible overexpression of the domain in *S. flexneri* and affinity chromatography with two different N-terminal tags. The DNA sequence corresponding to residues 2-125 of MxiG-N (*mxiGN$_{2-125}$*) was inserted between the *BglII* and *KpnI* restriction sites of pSZ1 (Zenk *et al.*, 2007), a derivative of pBAD/*Myc*-His A (Invitrogen). The resulting pBAD-*mxiGN$_{2-125}$* construct would lead to expression of MxiG-N$_{2-125}$ with an N-terminal MGHHHHHHGSR His-tag. In addition, *mxiGN$_{2-125}$* was also inserted between the *BamHI* and *EcoRI* restriction sites of pGEX-2T (GE Healthcare) to form pGEX-*mxiGN$_{2-125}$*, leading to expression of MxiG-N$_{2-125}$ with an N-terminal glutathione S-transferase (GST)-tag.

mxiGN$_{2-125}$ inserts for both constructs were amplified using the standard polymerase chain reaction (PCR) protocol (**Appendix 8.2.3**) with a T_a of 54^{o}C, pET14b-*mxiGN$_{1-126}$* as the template DNA and the primers outlined in **Table 3.4**. 5 µl of each PCR mixture was separated on a 1% agarose gel (**Appendix 8.2.4**), confirming inserts of the expected size had been produced with no non-specific amplification (**Figure 3.6A**). The remaining 45 µl of each PCR mixture was supplemented with 1 µl FastDigest *DpnI* (Fermentas) and incubated at 37^{o}C for 1 h to selectively digest the template DNA, then 80^{o}C for 20 min to inactivate the enzyme. PCR mixtures for each construct were then pooled, and the inserts purified with the QIAquick PCR Purification Kit (Qiagen) to give ~200 ng/µl DNA.

Table 3.4 Primers used to amplify insert for pBAD-*mxiGN*$_{2\text{-}125}$ and pGEX-*mxiGN*$_{2\text{-}125}$ constructs		
Construct	**Forward**	**Reverse**
pBAD-*mxiGN*$_{2\text{-}125}$	5'– AGC CAT AGA TCT GAG GCA AAG AAC TCA AAT CTT GCA CC -3'	5' – GCA TCA GGT ACC CTA GTG GTT CTT ATA CAT TCC GTT TAG TAT ATG GCC -3'
pGEX-2T-*mxiGN*$_{2\text{-}125}$	5' – GGC AGG GAT CCG TCT GAG GCA AAG AAC TCA AAT CTT GCA -3'	5'- GCA TCA GAA TTC CTA GTG GTT CTT ATA CAT TCC GTT TAG TAT ATG GCC -3'

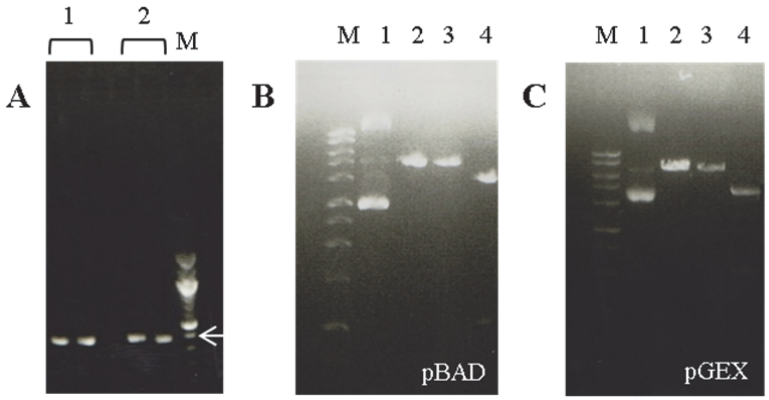


Figure 3.6 *1% agarose gels of DNA fragments from pBAD-mxiGN*$_{2\text{-}125}$ *and pGEX-mxiGN*$_{2\text{-}125}$ *cloning*.

A *mxiGN*$_{2\text{-}125}$ inserts amplified using the standard PCR protocol. M 100 bp DNA ladder (NEB), 1 insert for pBAD-*mxiGN*$_{2\text{-}125}$ (396 bp) and 2 insert for pGEX-*mxiGN*$_{2\text{-}125}$ (396 bp). The white arrow indicates the 400 bp marker.

B & C Samples from vector double digest reactions. M 1 kb DNA ladder (Promega).
B Digestion of pSZ1 construct. 1 undigested vector, 2 +ve control for *KpnI* digestion, 3 +ve control for *BglII* digestion and 4 double digest with *KpnI* and *BglII*.
C Digestion of pGEX-2T-IcsA construct. 1 undigested vector, 2 +ve control for *BamHI* digestion, 3 +ve control for *EcoRI* digestion and 4 double digestion with *BamHI* and *EcoRI*.

Inserts and vectors were sequentially digested with *KpnI* then *BglII* for pBAD-*mxiGN*$_{2\text{-}125}$ and *BamHI* then *EcoRI* for pGEX-*mxiGN*$_{2\text{-}125}$. In each case, a 50 µl reaction containing ~3 µg purified insert or ~2 µg vector, 5 µl 10 x FastDigest buffer (Fermentas) and 3 µl FastDigest restriction enzyme 1 (Fermentas) was incubated at 37^{o}C for the maximum allowed time before star activity occurs, then at an appropriate deactivation temperature, as specified by the enzyme

manufacturer. The reaction volume was then increased to 70 µl by supplementing with 2 µl 10 x FastDigest buffer and 3 µl FastDigest restriction enzyme 2, and the digest repeated with the appropriate parameters for the second enzyme. For each enzyme, a 10 µl control reaction was run in parallel, containing ~0.6 µg vector, 1 µl 10 x FastDigest buffer and 0.6 µl FastDigest enzyme to ensure the same ratio of components as in the large-scale digest. Subsequent separation of a sample of the vector double digest reaction on a 1% agarose gel demonstrated that digestion of pSZ1 (**Figure 3.6B**) and pGEX-2T-IcsA (**Figure 3.6C**; donated by J. Deane) proceeded to completion, producing linearised pBAD and pGEX vectors respectively for insertion of *mxiGN*$_{2\text{-}125}$. To minimise vector self-ligation, 20 µl each vector digest were treated with 1 µl FastAP alkaline phosphatase (Fermentas) and 2 µl 10 x FastAP buffer before incubation at 37^{o}C for 10 min then 75^{o}C for 5 min. Both vectors and inserts were then separated on a 1% agarose gel, with 1:10000 SYBR Safe DNA gel stain (Invitrogen) added to both the gel and running buffer, and visualised with blue light using a Safe Imager (Invitrogen). Linearised vectors were excised and purified using the QIAquick gel extraction kit (Qiagen) to give ~10 ng/µl DNA. Inserts were also re-purified with the QIAquick PCR Purification Kit to give ~100 ng/µl DNA.

10 µl ligation reactions were subsequently set up, comprising 50 ng alkaline phosphatase-treated vector, 5 µl 2 x Rapid Ligation buffer (Promega), 1 µl T4 DNA ligase (Promega) and purified insert at vector:insert molar ratios of 1:20, 1:10 and 1:6. In addition, a control reaction for vector self-ligation was set up with no insert added. The ligation reaction was left to proceed at 25^{o}C for 30 min then ligase was inactivated at 75^{o}C for 10 min. NovaBlue competent cells (Merck) were subsequently transformed with the ligation mix using the 'gentle transformation' protocol (**Appendix 8.2.1**). Selected transformants were propagated in 5 ml overnight cultures and constructs purified using the QIAprep miniprep kit (Qiagen). The presence of the correct insert was confirmed for both pBAD-*mxiGN*$_{2\text{-}125}$ and pGEX-*mxiGN*$_{2\text{-}125}$ by DNA sequencing with pBADf and pGEX5 primers respectively (Source BioScience).

Pull Down Assays With MxiG-N_{2-125}

MxiG-N_{2-125} constructs were overexpressed in the *mxiG*$^-$ strain of *S. flexneri* (**Table 3.2**), so there was no endogenously expressed untagged MxiG-N competing for binding to *in vivo* interaction partners. To overexpress MxiG-N_{2-125} with an N-terminal His-tag (His-MxiG-N_{2-125}), *mxiG*$^-$ was transformed with pBAD-*mxiG*N_{2-125} via the 'gentle transformation' protocol (**Appendix 8.2.1**) and propagated overnight on TCSB agar (14 mg/ml agar and 30 mg/ml TCSB). Colonies were picked to set up 5 ml TCSB overnight cultures, which were then used to inoculate 2 x 1 l TCSB supplemented with 0.2% arabinose to induce His-MxiG-N_{2-125} overexpression. These cultures were grown at 37^{o}C in a shaking incubator until OD_{600} of ~1.0 was reached, at which point cells were harvested and stored at -20^{o}C. 50 μg/ml kanamycin and 100 μg/ml ampicillin were present throughout cell growth. As a negative control to monitor the non-specific binding of proteins to the affinity column, 2 x 1 l cultures of *mxiG*$^-$ were grown in parallel with 50 μg/ml kanamycin added. A low level of His-MxiG-N_{2-125} was overexpressed at the expected molecular weight (15.4 kDa) in *mxiG*$^-$ pBAD-*mxiG*N_{2-125} construct (**Figure 3.7A**).

MxiG-N_{2-125} with an N-terminal GST-tag (GST-MxiG-N_{2-125}) was overexpressed in *mxiG*$^-$ largely as described above for His-MxiG-N_{2-125}. However, when expression from pGEX-*mxiG*N_{2-125} was induced immediately after the inoculation of 2 x 1 l TCSB, cell growth was severely retarded and no overexpression was observed after OD_{600} of ~1.0 was reached, suggesting the construct was no longer maintained due to the toxicity of GST-MxiG-N_{2-125} to *S. flexneri*. Therefore overnight cultures were used to inoculate 2 x 1 l TCSB supplemented with 0.5% glucose to repress GST-MxiG-N_{2-125} expression. Cells were grown until OD_{600} of ~0.3 was reached, then they were harvested, washed once with PBS and resuspended in 2 x 1 l TCSB and 0.4 mM IPTG. Cell growth was resumed until OD_{600} of ~1.0, before harvesting and storage at -20^{o}C. In addition, GST was overexpressed alone from a pGEX-2T construct with a stop codon inserted after the GST-tag (pGEX-stop; donated by J. Deane) in an analogous manner to GST-MxiG-N_{2-125} as a negative control for protein binding to the GST-tag. There was good

overexpression of both GST and GST-MxiG-N_{2-125} at their expected molecular weights of 26.3 kDa and 40.5 kDa respectively (**Figure 3.7B**).

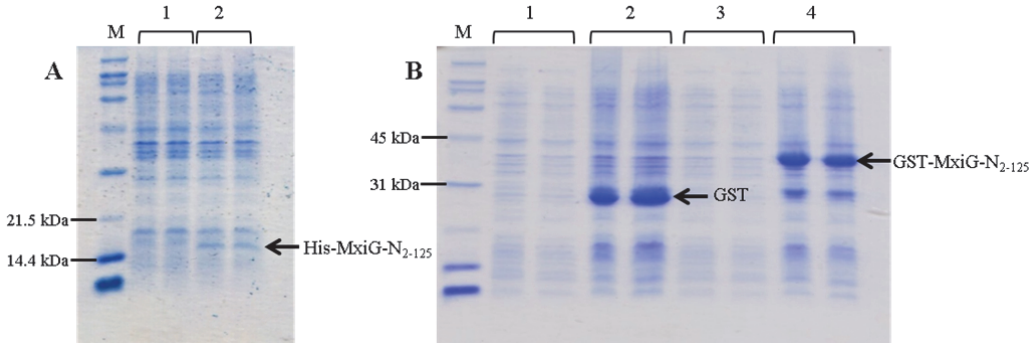


Figure 3.7 *15% SDS-PA gels showing overexpression of MxiG-N_{2-125} constructs in mxiG⁻.*

M SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad). Dual repeats are shown for each sample.

A 1 *mxiG⁻* and 2 *mxiG⁻* pBAD-*mxiGN_{2-125}*. Overexpression of His-MxiG-N_{2-125} is indicated.

B 1 *mxiG⁻* pGEX-stop pre-induction with 0.4 mM IPTG, 2 *mxiG⁻* pGEX-stop post-induction with 0.4 mM IPTG, 3 *mxiG⁻* pGEX-*mxiGN_{2-125}* pre-induction with 0.4 mM IPTG and 4 *mxiG⁻* pGEX-*mxiGN_{2-125}* post-induction with 0.4 mM IPTG. Overexpression of GST and GST-MxiG-N_{2-125} are indicated.

To determine which proteins are pulled-down with His-MxiG-N_{2-125}, *mxiG⁻* and *mxiG⁻* pBAD-*mxiGN_{2-125}* cells were resuspended in PBS in the presence of an EDTA-free protease inhibitor, homogenised and the resulting lysate centrifuged, as for the purification of MxiG-N_{1-126} in the previous chapter. The supernatants from both cell types were flowed over separate 1 ml Ni-NTA Superflow Cartridges (Qiagen) at 1 ml/min to allow both non-specific protein binding and specific binding of His-MxiG-N_{2-125} protein complexes where present. The cartridges were washed with 10 CV PBS and 50 mM imidazole to remove non-specifically bound proteins, before elution of 5 x 1 ml fractions with PBS and 300 mM imidazole. **Figure 3.8** shows samples taken from the first eluted fractions analysed by SDS-PAGE. Whilst His-MxiG-N_{2-125} was mostly eluted in the first fraction, Coomassie staining showed few additional bands and none of them were only present in the soluble fraction of *mxiG⁻* pBAD-*mxiGN_{2-125}* cells (**Figure 3.8A**). Silver-staining of the same gel identified further bands; however these were also present in the absence of His-MxiG-N_{2-125}, indicating that they arose from proteins non-specifically binding the Ni-NTA cartridge as opposed to proteins forming a specific complex with His-MxiG-N_{2-125}

(**Figure 3.8B**). These results suggested no proteins were pulled down in high abundance with His-MxiG-N_{2-125}, with the high background binding to the resin making Ni-affinity chromatography unsuitable for detecting low abundance interaction partners.

Pull-down assays with GST-MxiG-N_{2-125} were performed with the aim of improving the stringency of the affinity chromatography step. The soluble fractions of *mxiG*$^-$, *mxiG*$^-$ pGEX-

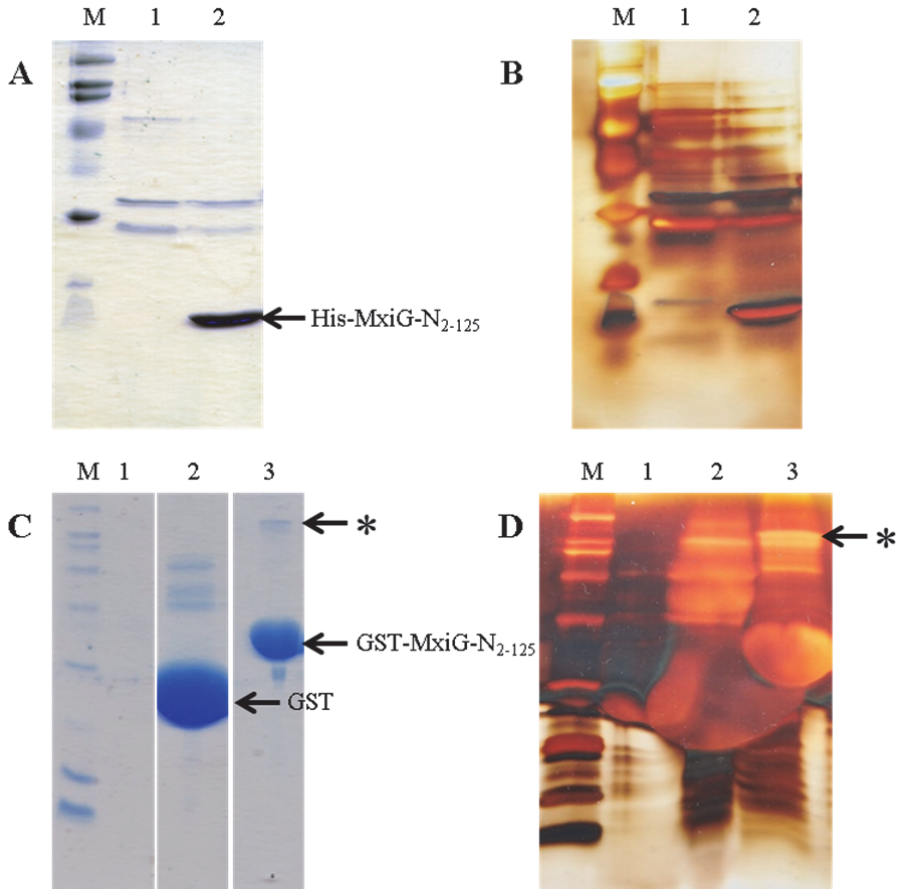


Figure 3.8 *15% SDS-PA gels showing the pull-down assays with MxiG-N_{2-125}.*

M SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

A & B Pull-down assays with His-MxiG-N_{2-125}. Samples taken from the first fraction eluted from a Ni-NTA column which had bound proteins within **1** the soluble fraction from *mxiG*$^-$ cells and **2** the soluble fraction from *mxiG*$^-$ pBAD-*mxiG*N_{2-125} cells. The band corresponding to His-MxiG-N_{2-125} is indicated.
A Gel staining with Coomassie stain (**Appendix 8.2.2**).
B Gel staining with SilverX-press kit (Invitrogen).

C & D Pull-down assays with GST and GST-MxiG-N_{2-125}. Samples taken from the first fraction eluted from a GSTrap FF column which had bound proteins within **1** the soluble fraction from *mxiG*$^-$ cells, **2** the soluble fraction from *mxiG*$^-$ pGEX-stop cells and **3** the soluble fraction from *mxiG*$^-$ pGEX-*mxiG*N_{2-125} cells. The bands corresponding to GST and GST-MxiG-N_{2-125} are indicated. The additional band that is unique to the GST-MxiG-N_{2-125} pull-down assay is indicated by an asterisk.
C Gel staining with Coomassie stain.
D Gel staining with SilverX-press kit.

stop and *mxiG*$^-$ pGEX-*mxiGN*$_{2\text{-}125}$ cells were isolated as described previously and flowed over separate 1 ml GSTrap FF columns (GE Healthcare) at 0.2 ml/min. The flow rate was increased to 1 ml/min for column washing with 10 CV PBS and elution of 5 x 1 ml fractions with 50 mM HEPES (pH 8.0), 150 mM NaCl and 10 mM reduced glutathione. SDS-PAGE analysis of the first eluted fractions indicated that the pull-down assay with GST-MxiG-$N_{2\text{-}125}$ would be more appropriate for the identification of low abundance MxiG-$N_{2\text{-}125}$ binding partners than that with His-MxiG-$N_{2\text{-}125}$. Firstly, relatively few proteins from the *mxiG*$^-$ soluble fraction had non-specifically bound the column (**Figure 3.8C&D, lane 1**), indicating a much clearer background for identifying faint gel bands. Secondly, a much larger amount of GST-MxiG-$N_{2\text{-}125}$ was specifically eluted relative to His-MxiG-$N_{2\text{-}125}$ (**Figure 3.8C, lane 3**), which would likely increase the amount of any interaction partners pulled-down. Despite these factors, gels had to be severely overloaded to allow the detection of faint bands for interaction partners, so bands with molecular weights close to those of GST and GST-MxiG-$N_{2\text{-}125}$ could not be distinguished (**Figure 3.8C&D**). Although, some higher molecular weight bands could be visualised, most proteins pulled-down by GST-MxiG-$N_{2\text{-}125}$ were also pulled-down by GST alone (**Figure 3.8C&D**), indicating an interaction with the GST-tag rather than MxiG-$N_{2\text{-}125}$. There was one protein isolated specifically with GST-MxiG-$N_{2\text{-}125}$ (**Figure 3.8C&D, asterisk**), however when the gel band was subjected to trypsin digestion and mass spectrometry identification (Central Proteomics Facility, University of Oxford) it was shown to be MxiG-N (mascot score = 185), indicating formation of a GST-MxiG-$N_{2\text{-}125}$ disulphide-bonded dimer.

Alternative Approaches to Identifying MxiG-N Interaction Partners

MxiG-$N_{2\text{-}125}$ did not pull-down *in vivo* interaction partners at a level that could be unambiguously detected by either Coomassie or silver-staining, suggesting either MxiG-N complexes are not amenable to isolation by this type of assay or modifications to the current protocol are required. The C-ring component Spa33 was identified as a putative interaction partner for MxiG-N by a pull-down assay which required incubation of purified GST-tagged MxiG with lysate prepared from *E. coli* overexpressing Spa33 (Morita-Ishihara *et al.*, 2006).

Therefore, both interaction partners were present at elevated levels relative to wild-type expression, suggesting native levels of GST-MxiG-N_{2-125} binding partners may have been undetectable. An identical approach was used to show interactions between CdsD and purified GST-tagged CdsQ, CdsL and CdsN, the *C. pneumonia* homologues of MxiG, Spa33, MxiN and Spa47 respectively (Johnson *et al.*, 2008, Stone *et al.*, 2008). However, incubation of His-tagged CdsQ and CdsL with *C. pneumonia* lysate also led to copurification of native CdsD (Johnson *et al.*, 2008), suggesting the levels of MxiG-N_{2-125} binding partners could be detectable using the above approach. In all these experiments the interaction partners of MxiG-N and its homologues were detected with specific antibodies, indicating antibodies towards T3SS components could be used to distinguish and unambiguously identify potential binding partners from within complex gel profiles.

Interestingly, C-terminally truncated CdsD was unable to bind GST-tagged CdsQ and CdsL (Johnson *et al.*, 2008). This corresponds to loss of the putative periplasmic region of CdsD, which is unlikely to be directly mediating the interactions with CdsQ and CdsL. Therefore the C-terminal truncations could have affected membrane insertion or oligomerisation of CdsD, indirectly abrogating interactions with the N-terminus. Similarly, overexpression of MxiG-N_{2-125} constructs in the absence of the rest of the MxiG protein could have been insufficient for isolation of binding partners; the MxiG-N domain may have to be membrane-localised within a ring as in the *Shigella* basal body to reach the high local concentration required for protein complexes to form. Even then, putative interactions with MxiG-N and its homologues have been characterised qualitatively as opposed to quantitatively, so any complexes formed with the domain could either be stable or transient in nature, with the latter making them difficult to isolate.

The identification of Spa33, MxiN and Spa47 as putative binding partners for MxiG-N (Morita-Ishihara *et al.*, 2006, Johnson *et al.*, 2008, Stone *et al.*, 2008), implies the domain is acting as a promiscuous docking platform for cytoplasmic T3SS components. It is equally likely these interactions are indirect or have been forced by artificially high levels of both binding partners

mixed *in vitro* and ultrasensitive detection with specific antibodies. An alternative approach which directly targets likely binding partners and attempts to biophysically characterise their interaction with MxiG-N is required to complement the results from pull-down assays and meet the need for quantitative analysis. An example of this approach for the putative interaction with Spa33 will be discussed in the next section.

MxiG-N_{6-112} Does Not Interact With Phosphorylated Spa33

Recently, expression of residues 1-126 of MxiG-N with an N-terminal strep-tag (Strep-MxiG-N_{1-126}) in *S. flexneri* M90T 5a (**Table 3.2**) facilitated *in vivo* crosslinking and subsequent pull-down of the domain in complex with Spa33 (**Figure 3.9A**)(Barison *et al.*, 2012). As CR was present in the growth medium during Strep-MxiG-N_{1-126} expression in a wild-type strain, the T3SS was likely in a secretion-competent form and MxiG-N may have been exposed to more of its interaction partners than in the pull-down assays attempted in this study. Furthermore, transient complexes would have been permanently trapped by cross-links and therefore much easier to isolate.

However, caution must be taken when analysing the gel profile shown in **Figure 3.9A** as Strep-MxiG-N_{1-126} was extracted under non-reducing conditions and the band denoted by the asterisk is in a similar position as would be expected for a disulphide-linked dimer of the protein (31.5 kDa), shown in the previous chapter to have formed readily when MxiG-N_{2-126} was extracted from *E. coli* under non-reducing conditions (**Figure 3.9B, asterisk**). Furthermore, only cross-linking with ethylene glycol bis(succinimidylsuccinate) (EGS) gave rise to this additional band on the gel. EGS cross-links are broken by elevated pH, whilst dithio-bis(succinimidyl propionate) (DSP) cross-links are broken by reduction. Therefore the treatment of the cross-linked sample for running on a gel would have reduced any Strep-MxiG-N_{1-126} disulphide-bonded dimer in the case of DSP, but not in the case of EGS, indicating why the higher molecular weight band is only present in the case of EGS. The mass spectrometry data identifying the band as Spa33 (33 kDa) was not shown in the paper (Barison *et al.*, 2012).

However, as this result corroborated the finding that GST-tagged MxiG interacts with residues 80-217 of Spa33 (Morita-Ishihara *et al.*, 2006), further investigation was needed.

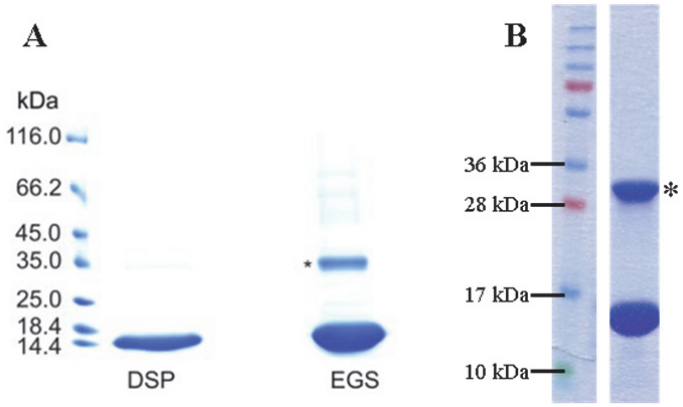


Figure 3.9 *Comparison of SDS-PA gel profiles of purified MxiG-N extracted from S. flexneri cells that were subjected to in vivo crosslinking and from E. coli cells in the absence of reducing agent.*

A Coomassie-stained gel taken directly from Barison *et al.,* 2012. Pull-down fractions obtained after EGS or DSP crosslinking. The asterisk indicates the putative band for Spa33.

B 15% SDS-PA gel showing the first fraction eluted from the Ni-NTA cartridge during the purification of MxiG-N_{1-126} under non-reducing conditions. The asterisk indicates the band shown to be a disulphide linked dimer of MxiG-N_{1-126} through comparison of reducing and non-reducing gels in the previous chapter.

Barison et al., 2012 then questioned whether MxiG-N is functioning as a canonical FHA domain by interacting with Spa33 in a phosphorylation-mediated manner. Without evidence for *in vivo* phosphorylation of Spa33, putative threonine phosphorylation sites were proposed within residues 80-217 based on common FHA domain recognition motifs. A phosphorylated peptide based on one of these Spa33 sequences (KRIRpTGDIIIVQ; peptide P_5) was shown to specifically interact with His-tagged MxiG-N_{1-126} by surface plasmon resonance (SPR), with a reported K_d of 33 ± 9 μM. This contradicted results presented in the previous section (McDowell *et al.*, 2011) that led to the conclusion that MxiG-N is not functioning as a canonical FHA domain, so further investigation was required.

A custom synthesised P_5 peptide (Severn Biotech) was found to have a very low solubility. This could also be inferred to be the case from the SPR experiments, as a maximum concentration of 15 µM was used, despite reporting a K_d of ~33 µM (Barison *et al.*, 2012). A saturated solution of P_5 was prepared by dissolving the amount of lyophilised peptide required for a 1 mM solution in 100 mM Tris-HCl (pH 7.5), 150 mM NaCl overnight at 4°C and subsequently removing undissolved peptide by centrifugation. The concentration of P_5 in solution was estimated to be 80 ± 47 µM by measuring the peptide bond absorbance at 220 nm (A_{220}) five times and assuming A_{220} of a 1 mg/ml solution is 11 (Scopes, 1994). A 1D ^{1}H spectra obtained for this saturated solution of P_5 supplemented with 5% D_2O confirmed the presence of the peptide in solution (**Figure 3.10, blue**). Furthermore, when peak intensities were compared to a spectrum obtained for 250 µM ^{15}N-MxiG-N_{6-112} (**Figure 3.10, black**), transferred into the same buffer using a Slide-A-Lyzer Dialysis Cassette (Thermo Scientific), the concentration estimated from UV absorbance appeared to be of the correct magnitude. Therefore P_5 was soluble at a concentration above the proposed K_d for MxiG-N_{1-126} binding (Barison *et al.*, 2012), allowing an NMR titration to be used to probe the interaction further.

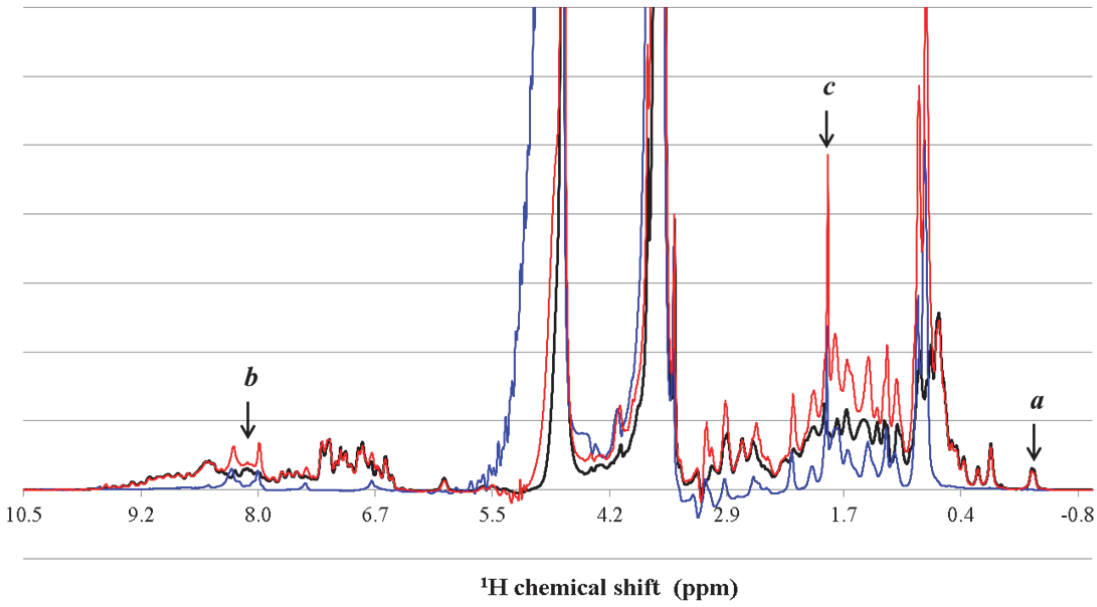


Figure 3.10 *1D ^{1}H spectrum showing titration of ^{15}N-MxiG-N_{6-112} with P_5.*

Overlay of spectra for a ~80 uM solution of P_5 (blue), a 250 µM solution of ^{15}N-MxiG-N_{6-112} (black) and a solution of 250 µM ^{15}N-MxiG-N_{6-112} with a saturating amount of P_5 (red) in 95 mM Tris-HCl (pH 7.5), 143 mM NaCl, 5 % D_2O. The baseline intensity is zeroed at 13.6 ppm and the intensity of all spectra are normalized to the intensity of the peak at 3.6 ppm corresponding to 95 mM Tris. ***a*** denotes a peak unique to ^{15}N-MxiG-N_{6-112}, whilst ***b*** and ***c*** denote peaks diagnostic for P_5.

To maximise the concentration of P_5 present in the titration, the peptide was directly dissolved overnight in a 250 µM solution of ^{15}N-MxiG-N_{6-112}, with subsequent removal of undissolved peptide by centrifugation. Comparison of the 1D ^{1}H spectrum of the resulting solution (**Figure 3.10, red**) with those obtained previously for the components in isolation showed that ^{15}N-MxiG-N_{6-112} was still present at 250 µM (**Figure 3.10, arrow a**). In addition, the intensities of diagnostic peaks for P_5 suggested it was present at a concentration equal to, if not greater than, that of the solution of the peptide in isolation (**Figure 3.10, arrows b&c**). A ^{1}H,^{15}N-HSQC spectrum subsequently obtained for this solution of 250 µM ^{15}N-MxiG-N_{6-112} and a saturating amount of P_5 showed no changes in peak positions relative to an equivalent spectrum obtained for 250 µM ^{15}N-MxiG-N_{6-112} alone, suggesting the peptide was not interacting with MxiG-N_{6-112} at this concentration (**Figure 3.11**).

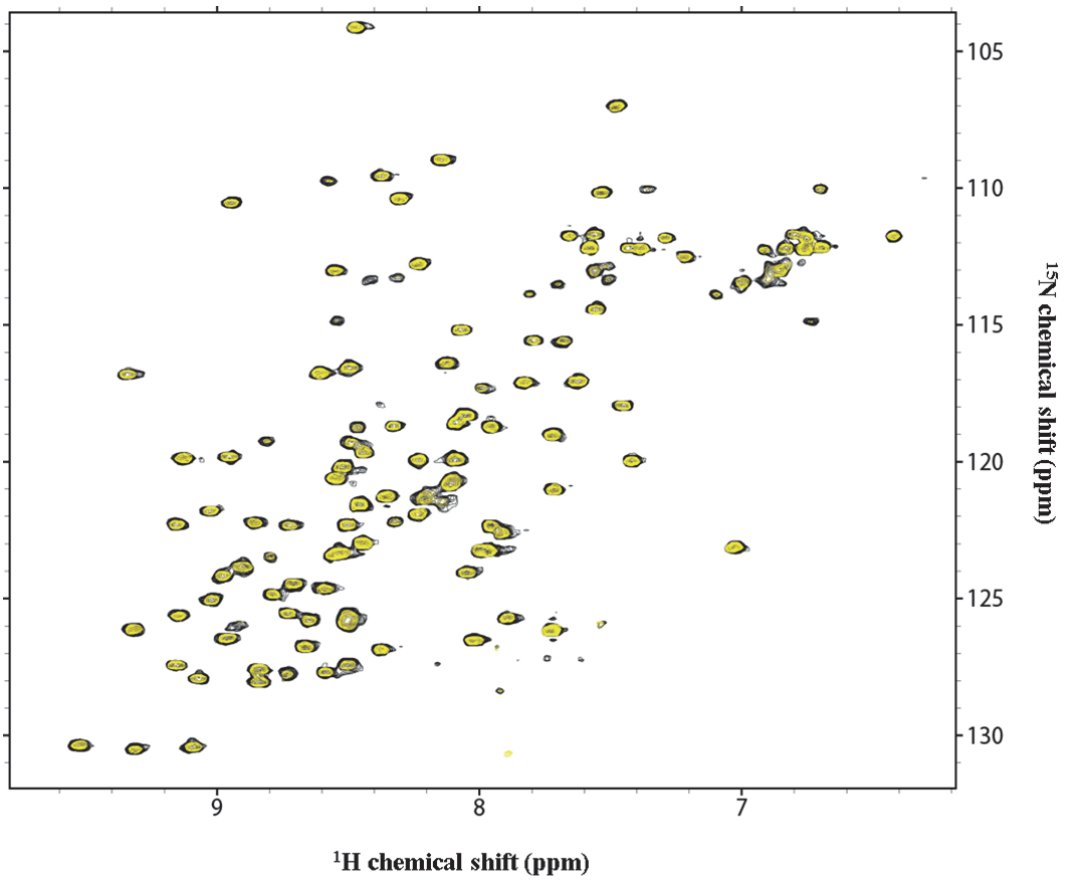


Figure 3.11 *Titration of ^{15}N-MxiG-N_{6-112} with a saturated solution of P_5.*

Overlay of ^{1}H,^{15}N-HSQC spectra obtained from 250 µM ^{15}N-MxiG-N_{6-112} (black) and a solution of 250 µM ^{15}N-MxiG-N_{6-112} and a saturating amount of P_5 (black) in 95 mM Tris-HCl (pH 7.5), 143 mM NaCl, 5 % D_2O.

To obtain a more concentrated solution of this hydrophobic peptide, P_5 was dissolved completely in 100 % dimethyl sulfoxide (DMSO) to obtain a 1 mM solution. Subsequently, a titration of 125 µM ^{15}N-MxiG-N_{6-112} with 100 µM P_5 resulted in a final buffer of 85 mM Tris-HCl (pH 7.5), 128 mM NaCl, 10 % DMSO, 5 % D_2O. A ^{1}H,^{15}N-HSQC spectrum obtained for this sample overlaid precisely with that obtained for 125 µM ^{15}N-MxiG-N_{6-112} in the same buffer (**Figure 3.12**), providing further evidence that P_5 does not interact with MxiG-N_{6-112}.

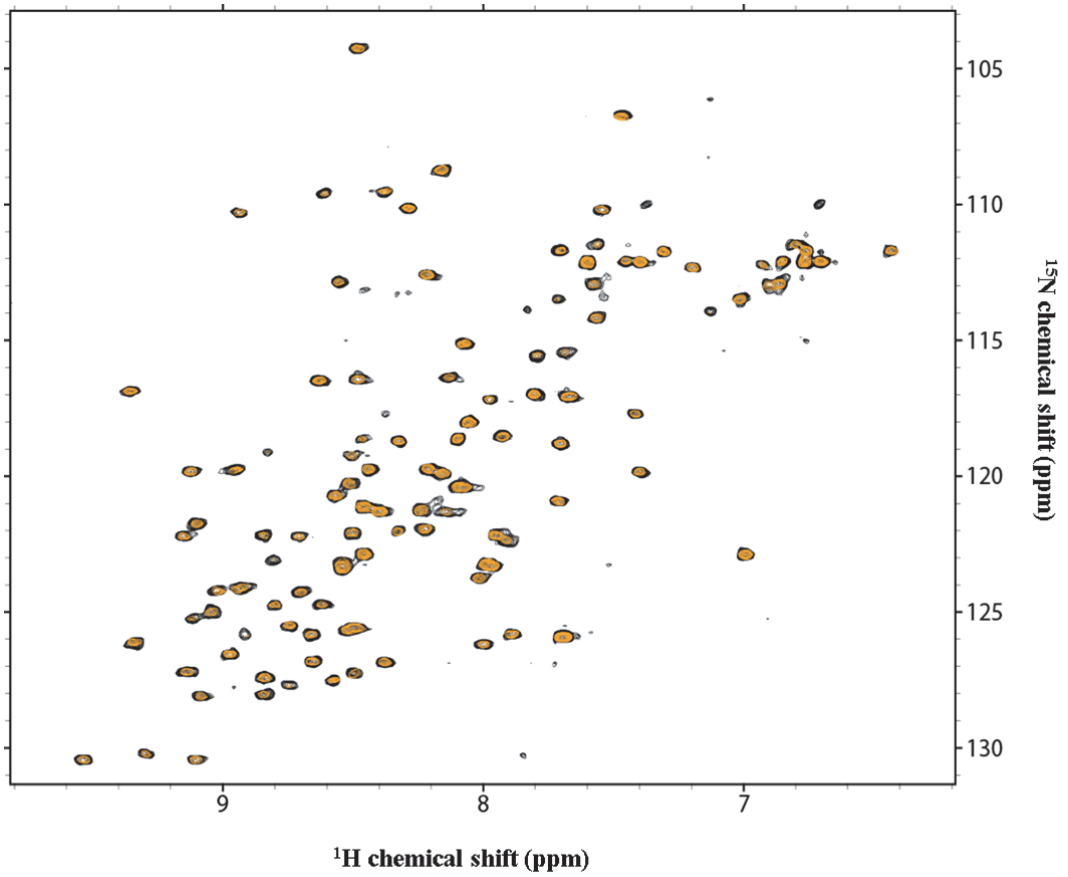


Figure 3.12 *Titration of ^{15}N-MxiG-N_{6-112} with P_5 dissolved in DMSO.*

Overlay of ^{1}H,^{15}N-HSQC spectra obtained from 125 µM ^{15}N-MxiG-N_{6-112} (black) and a solution of 125 µM ^{15}N-MxiG-N_{6-112} and 100 µM of P_5 (black) in 85 mM Tris-HCl (pH 7.5), 128 mM NaCl, 10 % DMSO, 5 % D_2O.

These results are in disagreement with those of Barison et al., 2012, which would predict that chemical shift perturbations would be seen for ^{15}N-MxiG-N_{6-112} at a concentration of P_5 tens of micromolar lower than used in these titrations. The low solubility of P_5 had not been previously discussed, so it is intriguing how peptide concentrations <15 µM were accurately measured for

SPR. Furthermore, the binding curve was only obtained for concentrations of P_5 below the estimated K_d, making quantitative analysis of the interaction by this technique invalid. Therefore much more evidence is required to support the conclusion that phosphorylated Spa33 interacts with MxiG-N, which cannot be provided by NMR titrations with ^{15}N-MxiG-N_{6-112}.

The finding that the conserved FHA residues R39 and S61 are required for efficient secretion of the translocator IpaB and invasion of mammalian cells by *S. flexneri* (Barison *et al.*, 2012) also directly contradicts results from this study. As opposed to using silver-staining to analyse the profile of early effectors secreted during a CR induction assay, western blotting with a specific α-IpaB antibody was used, which affords more indirect detection and therefore arguably more margin for error in interpretation of band intensities. Furthermore, the greater sensitivity of antibody detection to IpaB levels highlights relatively subtle differences that could be overinterpreted. However, the results of the invasion assay are intriguing, possibly highlighting a more subtle phenotype for the R39A and S61A mutants that would have been missed by the crude CR induction assay. Further investigation would be required to ascertain whether the phosphate-binding site of MxiG-N has a real physiological role. However, as stated above, my evidence does not support any role in binding phosphorylated Spa33.

3.4 Does MxiG-N Structurally Stabilise the IMR?

MxiG-N_{6-112} Does Not Interact with $MxiJ_{232-241}$

As outlined in the introduction, the IMR of the T3SS basal body is formed by concentric rings of MxiG and MxiJ, with the MxiG ring likely to be larger and encompassing the MxiJ ring. Although the cytoplasmic portion of MxiJ is only 10 residues long (KTGWFKRNKI; $MxiJ_{232-241}$)(Allaoui *et al.*, 1992), an interaction between MxiG-N and the C-terminus of MxiJ may contribute to the stable association of the two proteins. It was recently demonstrated that the C-terminus of PrgK cross-linked with the N-terminal domain of PrgH, suggesting interactions in

the cytoplasm may contribute to structural stabilisation of the IMR (Schraidt *et al.*, 2010). To further investigate this, $^{1}H,^{15}N$-HSQC spectra were collected before and after addition of 1 mM of a custom synthesised $MxiJ_{232-241}$ peptide (Severn Biotech) to a 220 µM sample of ^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT, 5 % D_2O. There were no changes in ^{15}N-MxiG-N_{6-112} peak positions in the presence of 1 mM $MxiJ_{232-241}$ (**Figure 3.13**), indicating no interaction is occurring at these concentrations.

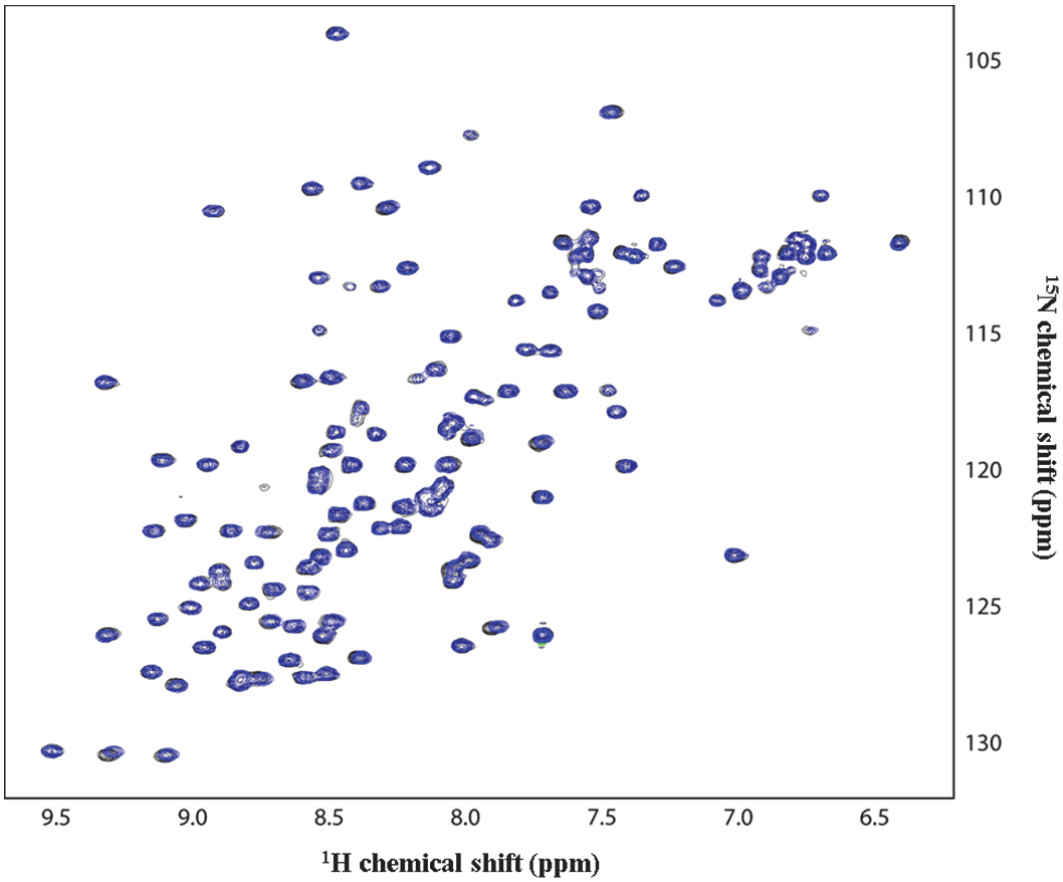


Figure 3.13 *Titration of ^{15}N-MxiG-N_{6-112} with 1 mM $MxiJ_{232-241}$.*

Overlay of $^{1}H,^{15}N$-HSQC spectra of 220 µM ^{15}N-MxiG-N_{6-112} obtained before (black) and after (blue) addition of 1 mM $MxiJ_{232-241}$.

This result indicates the stable association of MxiG and MxiJ in the IMR is unlikely to require a significant interaction between the cytoplasmic regions. This is in contrast to the cross-linking observed between the cytoplasmic portions of *Salmonella* PrgK and PrgH (Schraidt *et al.*,

2010). However these experiments imply proximity of these regions and not necessarily an interaction between them, whilst a more recent study found that cross-links between PrgH and PrgK within the context of the NC are confined exclusively to the periplasm (Sanowar *et al.*, 2010). 22 residues at the C-terminus of PrgK extend into the cytoplasm (Pegues *et al.*, 1995), so the region of PrgK that can potentially interact with the N-terminal domain of PrgH is much longer. In addition the MxiJ homologue in EPEC, EscJ, lacks a C-terminal cytoplasmic region (Yip *et al.*, 2005). Therefore the interactions important for stabilising the IMR may not be the same between the T3SS of different bacterial species.

A 24 Subunit Ring Suggests Close Association of MxiG-N

An alternative function of MxiG-N could be in the structural stabilisation of the IMR via self-association. Some EM reconstructions showed a continuous ring of density in the region corresponding to MxiG-N at the base of the NC (Marlovits *et al.*, 2004, Hodgkinson *et al.*, 2009), whilst others indicated a discontinuous ring (Sani *et al.*, 2007). Therefore it was previously unclear whether MxiG-N domains were in close enough proximity to oligomerise. The positioning of 24 subunits of MxiG-N_{6-112} within its ring of density in the *S. flexneri* NC EM map, as presented in the previous chapter (**Figure 2.23**), suggested that MxiG-N domains were closely associated in the IMR.

The electrostatic surface of MxiG-N_{6-112} calculated using APBS (Baker *et al.*, 2001) shows one face of the molecule to have a large positively charged patch (**Figure 3.14A**), whereas the opposing face is predominantly negatively charged (**Figure 3.14B**). In addition, it can be seen from both views in **Figure 3.14** that surface complementarity also exists between opposite faces of the molecule, suggesting that an interaction between MxiG-N domains in the IMR would be feasible. However, undirected docking of two copies of the MxiG-N_{6-112} top model using Z-Dock (Chen *et al.*, 2003) only resulted in dimer models where the C-termini pointed in opposite directions (not shown), making it unlikely both could cross the inner membrane in this orientation. This is a similar arrangement to the dimer observed in the crystallographic

asymmetric unit of YscD-N_{2-108} (Lountos *et al.*, 2012), suggesting this crystal dimer is also not physiologically relevant.

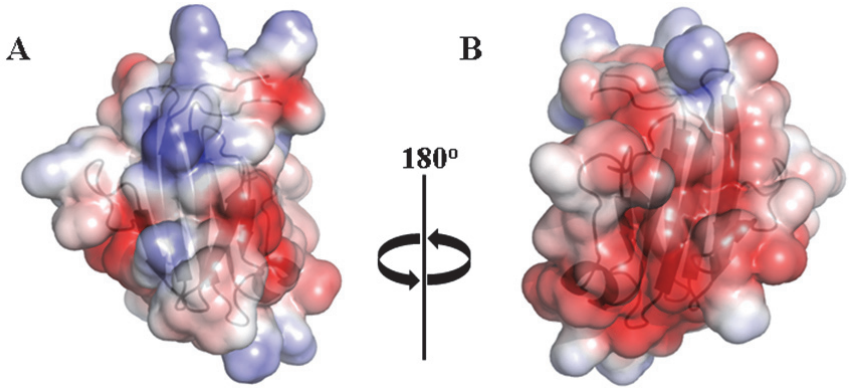


Figure 3.14 *The electrostatic surface of the NMR structure of MxiG-N_{6-112}.*

Positive (blue) and negative (red) regions are mapped on the surface representation, with the cartoon representation shown underneath. Electrostatics were calculated with the APBS plugin for PyMol (Baker *et al.*, 2001) using a physiological ion concentration of 150 mM and surface contouring to ± 3. **A & B** show different views of MxiG-N_{6-112} rotated 180° relative to each other.

Although the crystallographic asymmetric unit for MxiG-N_{1-126} contains only one copy (Barison *et al.*, 2012), analysis of the crystal contacts between symmetry related copies using PISA (Krissinel & Henrick, 2007) indicated one dimer that could be stable in solution, with a buried interface of 6410 Å^2 (**Figure 3.15**). Two copies of MxiG-N_{1-126} are intertwined so that the two-stranded β-sheet at the C-terminus is arranged antiparallel to β6 at the edge of β-sheet 2, providing a possible explanation for why the loop region between β7 and β8 adopts a slightly different conformation compared to the MxiG-N_{6-112} NMR structure (**Figure 2.22A, arrow *c***). Although the C-termini of both copies in this dimer point towards the membrane,

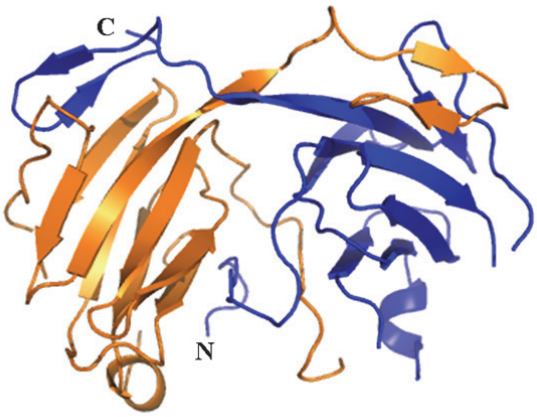


Figure 3.15 *MxiG-N_{1-126} dimer formed by crystal contacts*.

Analysis of the structure coordinates (pdb id 4a4y)(Barison *et al.*, 2012) was performed using PISA (Krissinel & Henrick, 2007). Each copy within the dimer is shown in a different colour, with labeling of the N- and C-termini for one copy.

MxiG-N association in this manner would imply formation of a 24-subunit ring with 12 copies of the dimer, contradicting the ring model proposed in the previous chapter (**Figure 2.23**). As stated previously, although a 12-fold symmetry EM map exists for the *S. flexneri* NC (Veenendaal *et al.*, 2007), it was hard to reconcile the density with the position of an asymmetric dimer of MxiG-$N_{6\text{-}112}$ (**Figure 2.24B**), so the physiological relevance of this crystallographic dimer is questionable. However, unpublished evidence from a LexA-based bacterial one-hybrid system has indicated that isolated YscD-N dimerises or further oligomerises (Ross & Plano, 2011). Therefore it should be questioned whether MxiG-N shows a propensity to oligomerise in solution.

MxiG-N is Monomeric in Solution

The oligomeric state of MxiG-N was initially characterised by parameters derived from NMR spectroscopy. The overall correlation time (τ_c) is defined as the average time taken for a molecule to rotate in solution and is dependent on a number of factors, including molecular weight (Cavanagh, 2007). Based on experimentally derived R1, R2 and ^{1}H-^{15}N NOE values, τ_c was estimated to be 6.21 ns for 1.3 mM ^{15}N-MxiG-$N_{6\text{-}112}$ during determination of the diffusion tensor with an axially symmetric model in ROTDIF (Walker *et al.*, 2004), as presented in the previous chapter. Comparison of this value with a standard curve obtained under similar

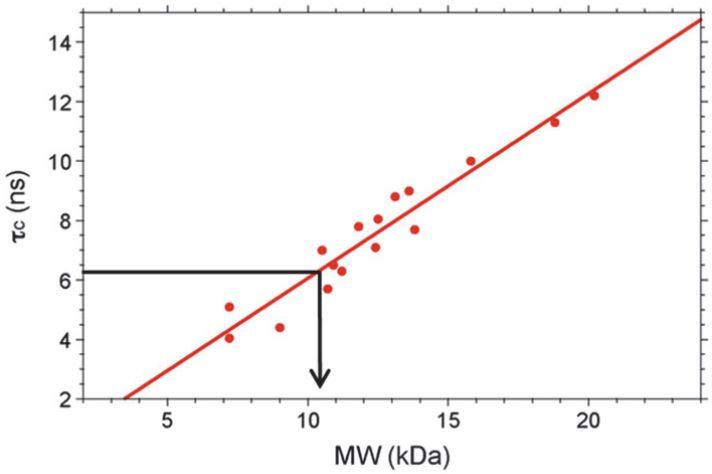


Figure 3.16 *MxiG-$N_{6\text{-}112}$ tumbles as a monomer in solution.*

Standard curve correlating τ_c with molecular weight adapted from the NESG wiki (Aramini, 2010). Measurements were obtained at 25°C on a 600 MHz spectrometer. A τ_c value of 6.21 ns for 1.3 mM ^{15}N-MxiG-$N_{6\text{-}112}$ is shown to correlate to a molecular weight of approximately 10 kDa.

experimental conditions (Aramini, 2010), suggested MxiG-N_{6-112} was tumbling with a molecular weight of ~10 kDa and was therefore likely to be monomeric (12.2 kDa; **Figure 3.16**). Although measurements of ^{15}N relaxation parameters were only performed on a 1.3 mM sample, the 2D $^{1}H,^{15}N$-HSQC and 3D spectra obtained with up to 6 mM $^{13}C/^{15}N$-MxiG-N_{6-112} had similar linewidths, chemical shifts and resolution, suggesting the domain was still predominantly monomeric even at these high concentrations.

To corroborate the above findings with data from the full-length domain, unlabelled MxiG-N_{1-125} was purified in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT and concentrated to 0.45 mM. Dr Steven Johnson then carried out multiangle light scattering (MALS) on a dilution series of this sample, using a multiangle light photometer (Dawn Heleos II, Wyatt) connected to a high pressure liquid chromatography (HPLC) system with a Superdex 200 10/300 GL SEC column attached. The in-line multi-angle scattering across the only elution peak showed a molecular weight consistent with a MxiG-N_{1-126} monomer (16.5 kDa) for all concentrations tested (**Figure 3.17**).

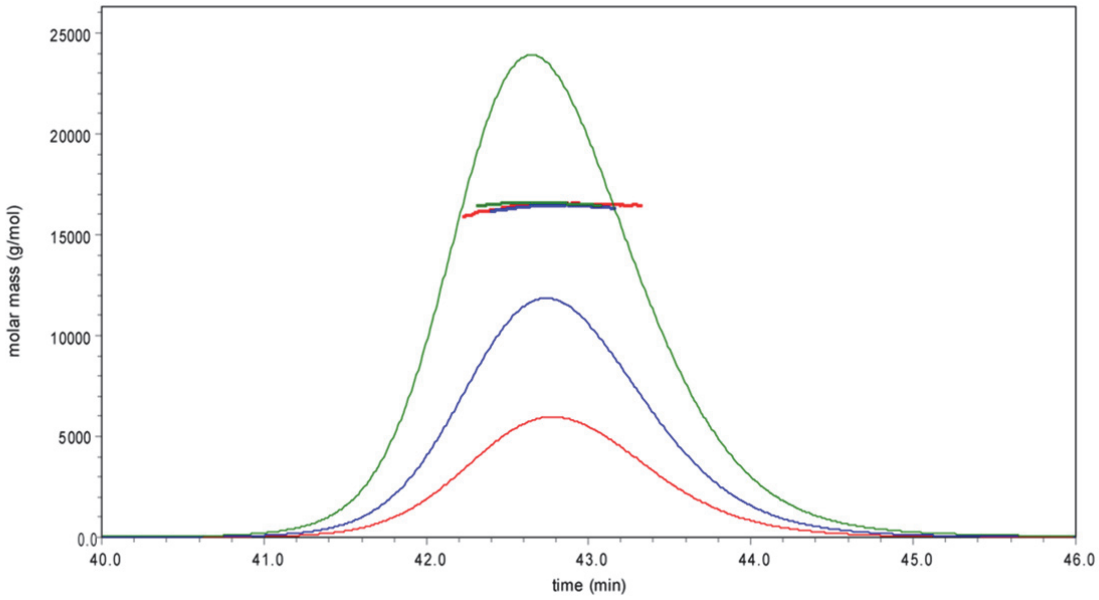


Figure 3.17 *Analysis of the oligomeric state of MxiG-N_{1-126} by SEC and in-line multi-angle light scattering.*

The left axis represents the molecular mass at any given point in the elution profile. Elution profiles are shown for the only elution peak at 0.1 mM (red), 0.2 mM (blue) and 0.45 mM (green) MxiG-N_{1-126} and demonstrate the domain is monomeric at these concentrations.

Together NMR spectroscopy and MALS data indicated that no high affinity interactions existed between MxiG-N domains. This was in agreement with a monomeric position in the SEC profile for YscD-N_{2-108} (Lountos *et al.*, 2012), suggesting this could be a general property of MxiG-N homologues. Furthermore, a stable MxiG-N_{1-126} dimer was not present in solution despite the extremely large buried interface observed in the crystal lattice (**Figure 3.15**), suggesting this crystallographic dimer is likely non-physiological and an artefact of the crystallisation condition (85 mM HEPES (pH 7.4), 1.7 M ammonium sulphate, 1.7% polyethylene glycol (PEG) 400, 15% glycerol)(Barison *et al.*, 2012).

Low Affinity Interactions Could Still Be Important For Ring Assembly or Stabilisation

As previously stated, the 24-subunit ring model presented in the previous chapter would suggest a close association of MxiG-N domains within the IMR, contradicting the finding that MxiG-N_{6-112} was monomeric in a solution of up to 6 mM. However MxiG is effectively localised at a high concentration within the physiological context of the IMR, so lower affinity interactions could play a role in stabilisation in the assembled ring. Indeed, assuming a volume of ~500,000 $Å^3$ and a stoichiometry of 24-subunits for the MxiG-N ring, the effective molarity of MxiG-N_{6-112} within the IMR can be estimated to be 80 mM, suggesting interactions with a K_d > 6 mM could be significant.

Interestingly, recent mathematical modeling of ring assembly by Deeds and coworkers suggests that high affinity interactions are not necessarily optimal for building a homomeric ring from its constituent monomers (Deeds *et al.*, 2012). Using the example of the formation of a three-membered ring, it was found that strong interaction affinities quickly resulted in depletion of monomers from solution, resulting in a delay in assembly whilst stable dimeric intermediates dissociated to release more monomeric building blocks. Therefore low affinity interactions between MxiG-N subunits may be productive for ring formation.

However, interactions which are too weak hinder the formation of rings as the intermediate structures form slowly and readily dissociate. For initial monomer concentrations ranging from

1 nm to 1 μM, optimal K_d values for monomer-monomer interactions in the formation of a three-membered ring were found to be in the same order of magnitude. In addition, larger-membered rings optimally require higher affinity interactions between monomers, in order to stabilise the extra intermediates that must be formed during the pathway to ring formation. (Deeds *et al.*, 2012) Therefore, although interactions between MxiG-N domains with an affinity in the millimolar range may be strong enough to stabilise an assembled ring, they would be unlikely to contribute significantly to the assembly of the IMR.

In EM reconstructions of both *Shigella* and *Salmonella* NCs, the ring of density corresponding to the cytoplasmic domains of MxiG and PrgH respectively has shown positional variation (Hodgkinson *et al.*, 2009, Marlovits *et al.*, 2004, Schraidt & Marlovits, 2011), suggesting MxiG-N is flexible and not held in a rigid ring structure by inter-domain interactions, at least in the absence of the export apparatus and cytoplasmic T3SS components. However, defined multimeric structures can be isolated from an *E. coli* strain overexpressing PrgH (Kimbrough & Miller, 2000), suggesting full-length MxiG does have the propensity to oligomerise in the absence of other NC components. Therefore interactions between the larger periplasmic MxiG-C domains may be more important for assembling and stabilising the MxiG ring, which comprises a rigid region in the NC substructure (Schraidt & Marlovits, 2011). Furthermore, dissociation of the *S. typhimurium* NC yields intact IMRs rather than separate PrgH and PrgK rings (Schraidt *et al.*, 2010, Schraidt & Marlovits, 2011), indicating the interaction with MxiJ must be considered when questioning how the MxiG ring is stabilised. The next chapter will present attempts to characterise the periplasmic domains of MxiG and MxiJ and the interactions between them, with a view to further understanding IMR assembly and stabilisation within the *S. flexneri* T3SS.

3.5 Perspectives

MxiG-N was found to be essential for activity of the T3SS, highlighting an important yet elusive role. MxiG-N_{6-112} was identified as a FHA domain, although a canonical role in cell-signalling via P-Thr binding at the conserved site was shown to be unlikely. Therefore the bacterial T3SS appears to have adopted the FHA fold for a novel function. Pull-down assays were unable to identify possible interaction partners for MxiG-N, which suggested complexes are either in low abundance or transient, requiring further optimisation of the protocol to be amenable to study. Putative interactions with both a phosphorylated peptide of Spa33 and the C-terminus of MxiJ were found not to be significant for the function of MxiG-N. The domain also appears to be predominantly monomeric in concentrated solutions, suggesting any role for MxiG-N in structural stabilisation of the IMR would rely on the low affinity interactions that could exist between the closely associated domains implied by the 24-subunit ring model for MxiG.

4. Purification and Characterisation of the MxiG and MxiJ Periplasmic Domains

As outlined in the introduction, the periplasmic region of the *S. flexneri* T3SS basal body is comprised of multiple copies of soluble domains from MxiG and MxiJ in the IMR and MxiD in the OMR. **Figure 4.1** shows the regions of MxiG, MxiJ and MxiD that are localised in the periplasm, with each having a significant region represented by a homologous structure. However, there are currently no high-resolution structures available for any periplasmic domain from the *S. flexneri* basal body. Furthermore, the interactions between these periplasmic domains are poorly characterised, with current models relying on data from biochemical cross-linking studies in *S. typhimurium* (Schraidt *et al.*, 2010, Sanowar *et al.*, 2010) or docking of the

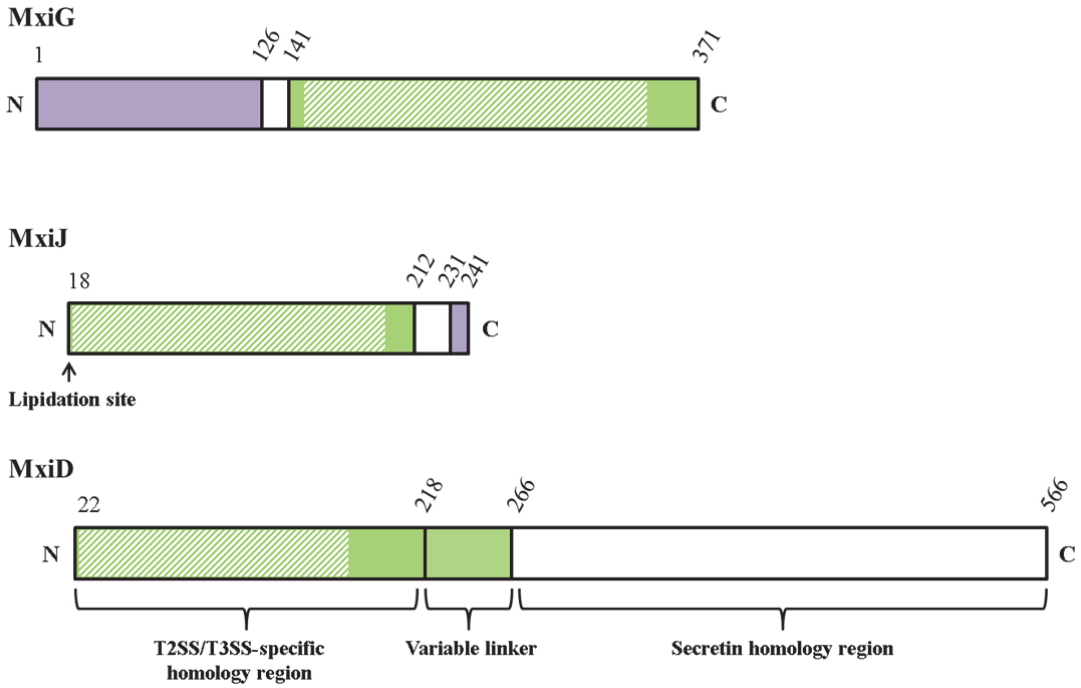


Figure 4.1 *Domain organisation of proteins contributing to the periplasmic region of the S. flexneri basal body.*

Schematics for MxiG, MxiJ and MxiD represent the mature protein, after signal peptide cleavage in the case of MxiJ and MxiD. Periplasmic, transmembrane and cytoplasmic regions are shown in green, white and purple respectively. Residue numbers at domain boundaries for MxiG (Allaoui *et al.*, 1995), MxiJ (Allaoui *et al.*, 1992) and MxiD (Allaoui *et al.*, 1993, Spreter *et al.*, 2009) are indicated with respect to the translation start site. The lipidation site at the N-terminus of MxiJ is shown by an arrow and the regions of MxiD are defined as in Spreter *et al*., 2009. Dashed boxes indicate regions of the periplasmic domains of MxiG (Spreter *et al.*, 2009), MxiJ (Crepin *et al.*, 2005, Yip *et al.*, 2005) and MxiD (Spreter *et al.*, 2009) for which a homologous structure has been determined.

available high resolution structures from different bacterial species into the *S. typhimurium* EM map (Schraidt & Marlovits, 2011). The uncertainty surrounding the arrangement of the IMR is significant, with multiple possible models for the position of the MxiJ periplasmic domain within the EM reconstruction of the *S. flexneri* NC (Hodgkinson *et al.*, 2009). In order to provide these missing insights into the structure and interactions of the periplasmic region of the *S. flexneri* IMR, the expression, purification and characterisation of the MxiG and MxiJ periplasmic domains was attempted, as outlined in this chapter.

4.1 Purification and Characterisation of MxiG-C

The only insight into the structure of MxiG-C homologues has been obtained from the crystal structure of residues 170-362 of PrgH ($PrgH_{170-362}$) from *S. typhimurium*, which showed three similar α/β domains connected by short linkers (Spreter *et al.*, 2009). A global pairwise alignment of the sequences of MxiG-C and $PrgH_{170-362}$ showed significant homology throughout the sequences (**Figure 4.2A**), indicating the protein fold is likely to also be conserved. Therefore this alignment was used in conjunction with the high-resolution structure of $PrgH_{170-362}$ to construct a SCWRL homology model (Altschul *et al.*, 1990) for residues 154-340 of MxiG-C (MxiG-$C_{154-340}$; **Figure 4.2B**). Unlike $PrgH_{170-362}$, MxiG-$C_{154-340}$ has three cysteine residues and interestingly C170 and C196 are in close enough proximity to form a disulphide bond when located within the homology model (**Figure 4.2B**). This indicated the correct folding of MxiG-C could require an oxidising environment reminiscent of the periplasm, providing an important consideration when designing protocols for the expression and purification of the domain. In addition, the last 30 residues of the domain are missing from the MxiG-$C_{154-340}$ model. As the extreme C-terminus of both PrgH (Schraidt *et al.*, 2010) and YscD (Ross & Plano, 2011) has been shown to interact with the OMR secretin, the full-length MxiG-C domain is an attractive target for structural characterisation.

A

```
PrgH 177 LGQEKERFQVLPGRDKMLYVAAQNERDTLWARQVLARGDYDKNARVIN   224
         :..:|....|..|:..  |:.|:.:.|.:||...|.:..:.|...::.
MxiG 156 IIDDKRYGIVNTGQCN--YILAETQNDAVWASVALNKTGFTKCRYILV   201

PrgH 225 ENEENKRISIWLDTYYPQLAYYRIHF--DEPRKPVFWLSRQRNTMSKKEL 272
         .|:|..||..:::..:|.:..|.::.  |:....|| ||::||:....||
MxiG 202 SNKEINRIQQYINQRFPFINLYVLNLVSDKAELLVF-LSKERNSSKDTEL 250

PrgH 273 EVLSQKLRALMPYADSVNITLMDDVTAAGQAEAGLKQQALPYSRRNHKGG 322
         :.|...|....||..::....:.|..|.|.|:....:..:.|........
MxiG 251 DKLKNALIVEFPYIKNIKFNYLSDHNARGDAKGIFTKVNVQYKEICENNK 300

PrgH 323 VTFVIQGALDDVEILRARQFVDSYYRTWGGRYVQFAIELKDDWLKGRSFQ 372
         ||:.::..|.|.::....:.:..:...:|.:|::|::.|.||..||:|:.
MxiG 301 VTYSVREELTDEKLELINRLISEHKNIYGDQYIEFSVLLIDDDFKGKSYL 350

PrgH 373 YGAEGYIKMSPGHWYF
         ...:.|:.::..||:|
MxiG 351 NSKDSYVMLNDKHWFF
```

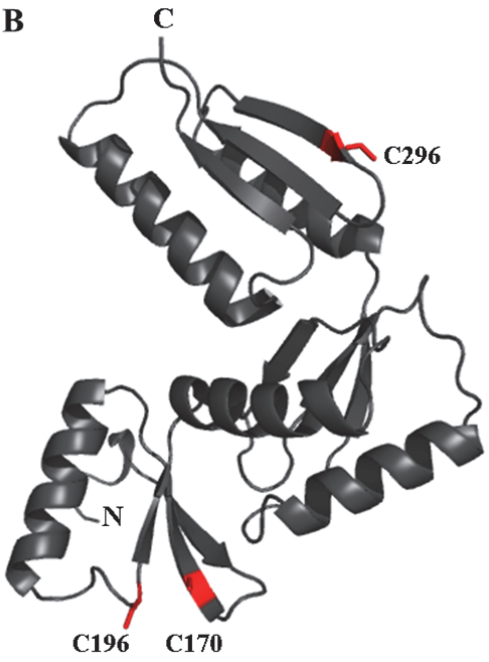


Figure 4.2 *Alignment of MxiG-C with PrgH*$_{170\text{-}362}$.

A Global pairwise alignment of MxiG and PrgH sequences using the EMBOSS-Needle server (EMBL-EBI). The alignment shown has 23.8% identity and 47.7% similarity. Dashes, colons and full-stops represent identical, conserved and semi-conserved residues respectively.

B SCWRL homology model (Altschul *et al.*, 1990) of MxiG-C$_{154\text{-}340}$ constructed using the structure of PrgH$_{170\text{-}362}$ (pdb id 3gr0)(Spreter *et al.*, 2009) and the pairwise sequence alignment shown above. Cysteine residues are shown in red and labelled: C170 and C196 are in close enough proximity to form a disulphide bond, whilst C296 is a free cysteine exposed to solution.

Purification of MxiG-$C_{142-371}$ from the Soluble Fraction

Expression and Purification from pET14b-mxiGC$_{142-371}$

Residues 142-371 of MxiG-C with a MGSSHHHHHHSSGLVPRGSHM N-terminal tag (MxiG-$C_{142-371}$) were expressed from a *pET14b-mxiGC$_{142-371}$* construct (donated by S. Johnson). Previous overexpression in B834 (DE3) cells and subsequent purification had ascertained that MxiG-$C_{142-371}$ was largely insoluble in the cytoplasm (Matthew Cottee, Part II Project). In an attempt to provide conditions that mimic the periplasm for MxiG-C folding, *pET14b-mxiGC$_{142-371}$* was 'gently transformed' (**Appendix 8.2.1**) into Origami 2 (DE3) cells, an *E. coli* strain with two mutations in the cytoplasmic disulphide-reduction pathway. To test the expression characteristics, 2 x 10 ml LB were inoculated with 100 µl of an overnight culture and grown at 37°C in a shaking incubator. When an OD_{600} of 0.6 was reached, expression was induced by the addition of 0.4 mM IPTG and each culture was split into 2 x 5 ml cultures, with one being grown for a further 3 h at 37°C and the other being grown overnight at 21°C. All growth media were supplemented with 100 µg/ml ampicillin and 12.5 µg/ml tetracycline throughout. SDS-PAGE analysis showed a prominent band at the expected molecular weight of 29.2 kDa for MxiG-$C_{142-371}$, with better overexpression after growth at 37°C for 3 h (**Figure 4.3**).

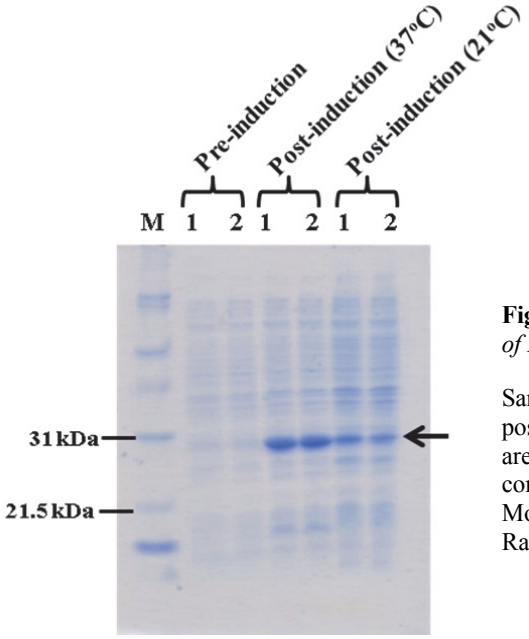


Figure 4.3 *15% SDS-PA gel showing overexpression of MxiG-$C_{142-371}$ in Origami 2 (DE3) cells.*

Samples taken from repeat cultures 1 and 2 pre- and post-induction after growth at both 37°C and 21°C are labelled. The arrow indicates the band corresponding to MxiG-$C_{142-371}$. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

In a large-scale expression, 6 x 1l LB were inoculated with an overnight culture of Origami 2 (DE3) pET14b-*mxiGC*$_{142\text{-}371}$ and grown as for the test cultures, with post-induction growth at 37^{o}C for 3 h. Cells were subsequently harvested, lysed in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl and the homogenate centrifuged as for the purification of MxiG-N$_{1\text{-}126}$ in Chapter 2. It was found that the majority of MxiG-C$_{142\text{-}371}$ expressed in Origami 2 (DE3) cells was found within the insoluble fraction (**Figure 4.4A**). Nonetheless, affinity purification of the soluble fraction was carried out using a 1 ml Ni-NTA Superflow Cartridge (Qiagen), with elution of 5 x 1 ml fractions. A small amount of soluble MxiG-C$_{142\text{-}371}$ mostly eluted in the second fraction (**Figure 4.4B, asterisk**).

Fraction 2 from Ni-affinity chromatography was injected on to an analytical Superdex 200 10/300 GL column (GE Healthcare) and passed through the matrix at 1 ml/min with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. The resulting elution profile had two main overlapping peaks (**Figure 4.4C, *1* & *2***). Analysis of their composition by SDS-PAGE showed the main species in both peaks to be MxiG-C$_{142\text{-}371}$ (**Figure 4.4B**), suggesting heterogeneity either in the folding or oligomeric state of the protein gave rise to separate peaks. In addition, fractions eluted at higher CVs contained an observable level of contamination. Fractions encompassing both elution peaks were pooled and concentrated using an Amicon Ultra centrifugal device with a 10 kDa cut-off (Millipore). Estimation of the concentration of the sample using a molar absorptivity coefficient of 28380 $M^{-1}cm^{-1}$ showed that 35 µg MxiG-C$_{142\text{-}371}$ could be purified from the soluble fraction of 1 l Origami 2 (DE3) pET14b-*mxiGC*$_{142\text{-}371}$. This low yield in combination with the heterogeneity observed above indicated this sample was inadequate for structural studies, suggesting a different protocol needed to be established for the purification of MxiG-C$_{142\text{-}371}$.

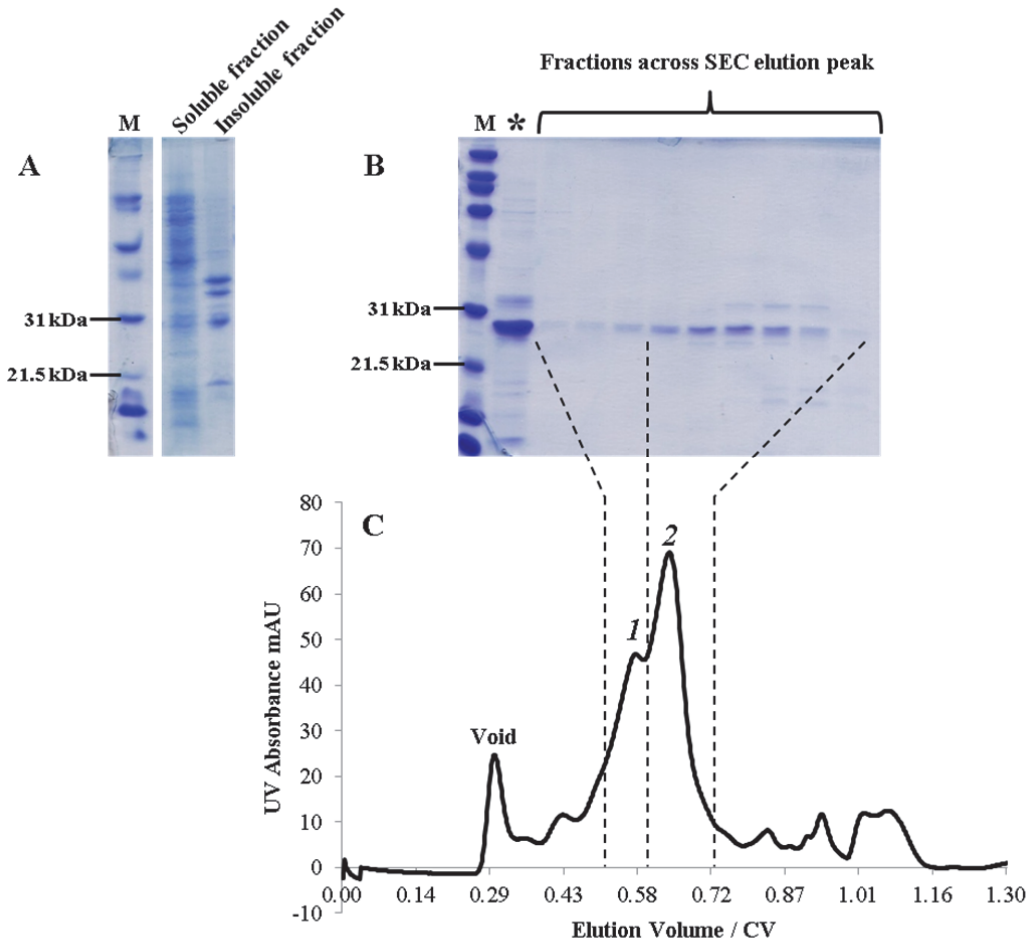


Figure 4.4 *Purification of MxiG-$C_{142-371}$ from Origami 2 (DE3) cells*.

M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

A 15% SDS-PA gel showing the solubility of MxiG-$C_{142-371}$ in Origami 2 (DE3) cells. The soluble fraction is a sample taken from the supernatant after 26,000 x g centrifugation, whilst the insoluble fraction is a sample taken from the pellet.

B 15% SDS-PA gel showing eluted fractions containing MxiG-$C_{142-371}$ from Ni-affinity and SEC. The asterisk shows a sample taken from fraction 2 eluted with 300 mM imidazole from a Ni-NTA cartridge. Samples taken during SEC are shown relative to the position of the fraction in the MxiG-$C_{142-371}$ elution profile.

C Elution profile for MxiG-$C_{142-371}$ from a Superdex 200 10/300 GL column. Elution with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. The two main elution peaks are numbered.

Construction of pET22b-mxiGC$_{142-371}$(C296S)

Although Origami 2 (DE3) cells minimise reduction of disulphide bonds, they do not promote the initial bond formation, suggesting their cytoplasm may not optimally encourage folding of MxiG-$C_{142-371}$. Therefore a construct was designed that would allow targeting of MxiG-$C_{142-371}$

for transport across the *E. coli* inner membrane, enabling folding of the domain in the periplasm, as would occur in *S. flexneri*. The DNA sequence of residues 142-371 of MxiG-C (*mxiGC*$_{142\text{-}371}$) was inserted between the *BamHI* and *XhoI* restriction sites of the pET22b vector (Novagen) to form pET22b-*mxiGC*$_{142\text{-}371}$. This provided MxiG-C$_{142\text{-}371}$ with both a C-terminal LEHHHHHH tag for affinity purification and a N-terminal pelB leader sequence for periplasmic targeting, which would leave a MDIGINSDP extension after its cleavage by signal peptidase.

Amplification of *mxiGC*$_{142\text{-}371}$ was achieved using pET14b-*mxiGC*$_{142\text{-}371}$ as a template and the primers outlined in **Table 4.1**. As the standard PCR protocol led to low yields and non-specific amplification, a touch-down PCR protocol was used instead (**Appendix 8.2.4**). Optimal amplification of *mxiGC*$_{142\text{-}371}$ was achieved with T_a decreasing from 52.6^{o}C to 47.6^{o}C for the first 10 cycles and T_a of 57^{o}C for the final 25 cycles, producing a good yield of the expected 714 bp fragment with minimal non-specific amplification (**Figure 4.5, asterisk**). *mxiGC*$_{142\text{-}371}$ was then purified with the QIAquick PCR Purification Kit (Qiagen) to give ~150 ng/µl DNA.

Table 4.1 Primers used to amplify the *mxiGC*$_{142\text{-}371}$ insert	
Forward	**Reverse**
5'- GGC AGG GAT CCG TCC CTC TCT TTA AAG AAA GAT G -3’	5'- GCC GGA CTC GAG CTT ATT TTT ATC TAA AAA AAA CCA GTG TTT ATC -3’

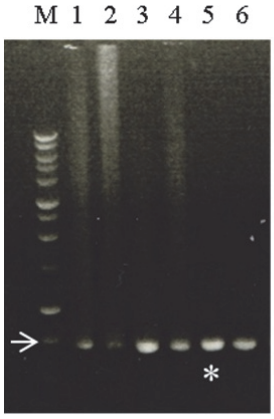


Figure 4.5 *1% agarose gel of mxiGC*$_{142\text{-}371}$ *insert amplified by touch-down PCR.*

Protocol optimised by varying the T_a gradient used in the first 10 cycles: T_a decreasing from 1 50.7^{o}C to 45.7^{o}C, 2 51.1^{o}C to 46.1^{o}C, 3 51.6^{o}C to 46.6^{o}C, 4 52.1^{o}C to 47.1^{o}C, 5 52.6^{o}C to 47.6^{o}C and 6 53.3^{o}C to 48.3^{o}C. The asterisk denotes the T_a range that resulted in optimal amplification of *mxiGC*$_{142\text{-}371}$. M 1 kb DNA ladder (Promega) with the 750 bp marker indicated by a white arrow.

The purified *mxiGC*$_{142-371}$ insert and empty pET22b vector were sequentially digested with *BamHI* then *XhoI*, as described previously for restriction digests during pBAD-*mxiGN*$_{2-125}$ and pGEX-*mxiGN*$_{2-125}$ cloning. Linearised vector was subsequently treated with alkaline phosphatase. Both digested insert and vector were subsequently purified using the QIAquick PCR Purification Kit (Qiagen), giving ~100 ng/µl and 15 µg/ul DNA respectively. The pET22b vector was ligated with *mxiGC*$_{142-371}$ and transformed into NovaBlue competent cells (Merck), also as outlined in the previous chapter. The presence of the correct insert in purified pET22b-*mxiGC*$_{142-371}$ was confirmed by DNA sequencing with the T7F primer (Source BioScience).

To reduce the formation of disulphide-bonded aggregates upon export to the periplasm, the putative free, surface-exposed C296 of MxiG-C$_{142-371}$ (**Figure 4.2B**) was mutated to serine. Site-directed mutagenesis (**Appendix 8.2.3**) was performed using the primers outlined in **Table 4.2**, a T_a of 55^{o}C and a t_e of 6 min. The presence of the mutation was subsequently confirmed by DNA sequencing, forming pET22b-*mxiGC*$_{142-371}$*(C296S)*.

Table 4.2 Primers used to create the C296S mutation in pET22b-*mxiGC*$_{142-371}$

Forward	**Reverse**
5'- G GTA AAT GTC CAG TAT AAA GAG ATT AGT GAA AAT AAT AAA GTT AC-3’	5'- GT AAC TTT ATT ATT TTC ACT AAT CTC TTT ATA CTG GAC ATT TAC C -3’

Expression and Purification from pET22b-mxiGC*$_{142-371}$*(C296S)

Expression and periplasmic localisation of MxiG-C$_{142-371}$(C296S) was tested by transformation using the heat-shock protocol (**Appendix 8.2.1**) of pET22b-*mxiGC*$_{142-371}$*(C296S)* into a variety of *E. coli* cell-lines: B834 (DE3) pLysS, C41 (DE3) pLysS, C43 (DE3) pLysS and Tuner (DE3) (**Table 4.3**). As a negative control for periplasmic export, B834 (DE3) pLysS cells were also transformed with pET14b-*mxiGN*$_{1-126}$. For each cell-line, 2 x 10 ml LB was inoculated with 50

µl of an overnight culture and cells were grown at 37°C in a shaking incubator until OD_{600} of ~0.6 was reached. Cells were induced with 0.1 mM IPTG and grown overnight at 21°C, so as to promote slow protein expression and prevent saturation of the Sec translocase in the inner membrane. 100 µg/ml ampicillin was present throughout cell growth. Analysis of whole-cell lysate confirmed overexpression in all cell-lines, although C41 (DE3) pLysS and C43 (DE3) pLysS expressed MxiG-$C_{142-371}$(C296S) at considerably lower levels (gel not shown).

Table 4.3 *E.coli* cell-lines used for expression of MxiG-$C_{142-371}$(C296S)

Cell-line	**Rationale for use**
B834 (DE3) pLysS*	Standard laboratory expression strain. Methionine auxotrophs allowing efficient selenomethione incorporation.
C41 (DE3) pLysS* / C43 (DE3) pLysS*	Mutations that prevent cell death from the expression of toxic recombinant proteins.
Tuner (DE3)	Uniform entry of IPTG into all cells in the population, allowing induction of expression in a concentration-dependent manner.

*Low levels of lysozyme are expressed throughout cell growth in the presence of the pLysS plasmid, reducing leaky expression prior to induction of recombinant protein expression.

Cells were subsequently spheroplasted in order to remove the cell-wall and release periplasmic proteins into solution. 1 ml of each culture was pelleted and resuspended in 90 µl 1 x PBS, 20% sucrose, 8 mM EDTA. The resuspension was subsequently supplemented with 20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 2% sucrose, 20 mg/ml lysozyme (chicken egg white; Sigma) to a total volume of 100 µl and incubated at 37°C for 1 h. Spheroplasts were then gently harvested by centrifugation at 2400 x g and the supernatant analysed by reducing SDS-PAGE (**Figure 4.6**). Bands corresponding to MxiG-$C_{142-371}$ were clearly visible in the periplasmic fractions from B834 (DE3) pLysS and Tuner (DE3) cells (**Figure 4.6, arrow**), potentially indicating successful targeting of unfolded protein for transport by the pelB leader sequence. However, any band corresponding to MxiG-N_{1-126} was obscured by lysozyme (**Figure 4.6, asterisk**), so the presence of cytosolic proteins in the supernatant due to spheroplast damage could not be discounted,

particularly as the intensities of MxiG-$C_{142-371}$ bands were proportional to expression levels. As the periplasmic fraction of B834 (DE3) pLysS pET22b-*mxiGC*$_{142-371}$*(C296S)* was relatively free from contaminating proteins, this cell-line was chosen for MxiG-$C_{142-371}$ expression.

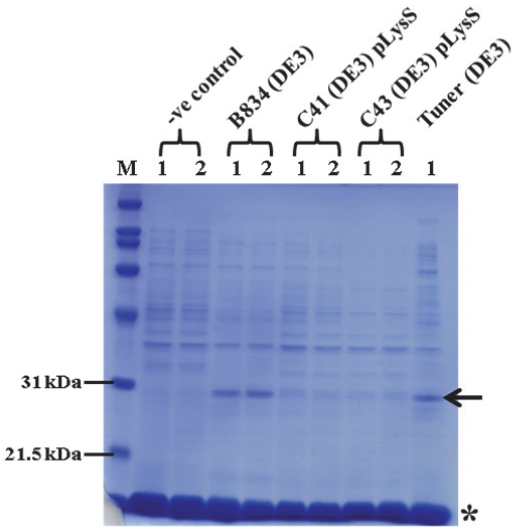


Figure 4.6 *13.5% SDS-PA reducing gel of periplasmic fractions from cells expressing pET22b-mxiGC*$_{142-371}$*(C296S)*

Samples taken from the supernatant obtained after the pelleting of spheroplasts. The –ve control was B834 (DE3) pLysS expressing MxiG-N_{1-126}, whilst the other *E. coli* cell types shown were expressing MxiG-$C_{142-371}$. Two repeat cell cultures for each cell-line were analysed. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad)). The arrow shows the bands corresponding to MxiG-$C_{142-371}$, whilst the asterisk shows the bands corresponding to lysozyme.

6 x 1 l LB cultures of B834 (DE3) pLysS pET22b-*mxiGC*$_{142-371}$*(C296S)* were grown as in the trial expressions described above, resulting in good overexpression of MxiG-$C_{142-371}$(C296S) (**Figure 4.7A, lanes 3&4**). This protein was subsequently purified as for MxiG-$C_{142-371}$ expressed by Origami 2 (DE3) pET14b-*mxiGC*$_{142-371}$ cells. No noticeable improvement in the solubility of MxiG-$C_{142-371}$(C296S) was observed (**Figure 4.7A, lane 6**), indicating the majority of the protein was incorporated into inclusion bodies within the cell before transport across the inner membrane could occur. However, a significant amount of soluble MxiG-$C_{142-371}$(C296S) eluted from the Ni-cartridge in the first fraction (**Figure 4.7A, lane 10**), potentially suggesting less sample aggregation. When this fraction was subjected to SEC, the main peak in the elution profile corresponded to MxiG-$C_{142-371}$(C296S) (**Figure 4.7B&C**). Although this elution peak was much less overlapped than observed with purification from pET14b-*mxiGC*$_{142-371}$, the high baseline still suggested a reasonably high level of background contaminants (**Figure 4.7C**). Estimation of the amount of MxiG-$C_{142-371}$(C296S) in a concentrated sample indicated that 50 µg protein could be purified from 1 l B834 (DE3) pLysS pET22b-*mxiGC*$_{142-371}$*(C296S)* and

therefore the yield had not been significantly improved. In addition, repeated purifications of MxiG-$C_{142-371}$(C296S) resulted in variable elution profiles and total yields, implying heterogeneity and instability in the sample. Comparison of protein purified from pET22b-*mxiGC*$_{142-371}$*(C296S)* and pET14b-*mxiGC*$_{142-371}$ by reducing SDS-PAGE showed that their molecular weights were similar (**Figure 4.7D**), suggesting that the molecular weight of MxiG-$C_{142-371}$(C296S) was more likely to be 28.9 kDa than 31.2 kDa and signal peptide cleavage had occurred. Therefore, although it appeared that the soluble fraction of MxiG-$C_{142-371}$(C296S) was

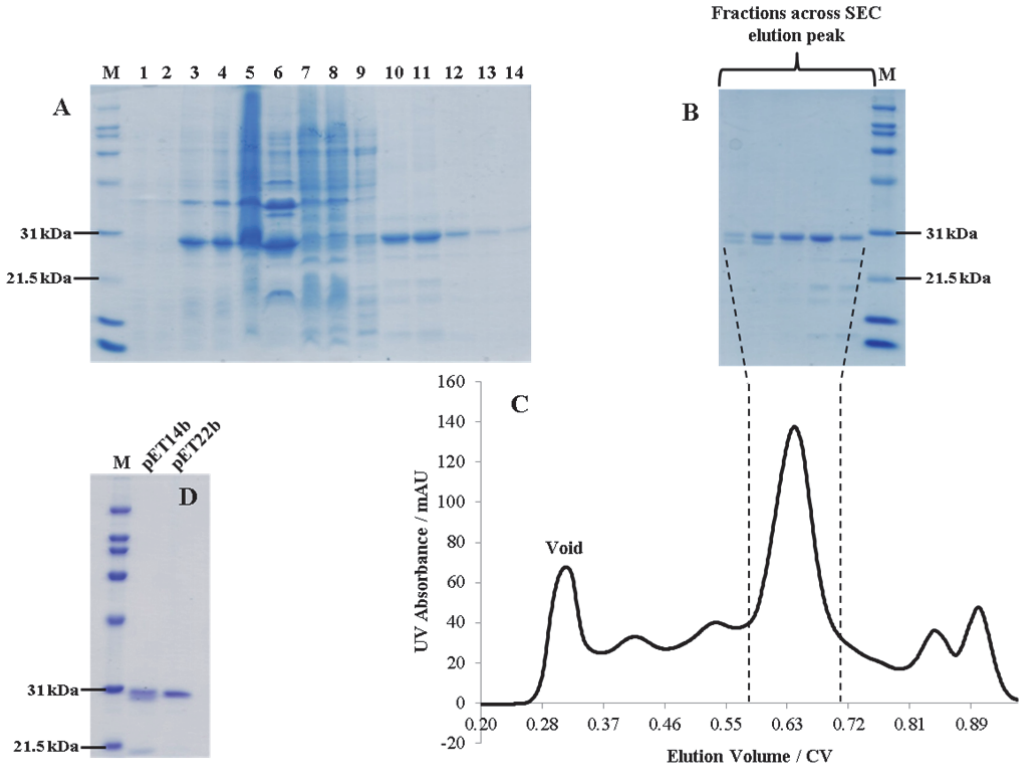


Figure 4.7 *Purification from pET22b-mxiGC*$_{142-371}$*(C296S)*.

M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

A 15% SDS-PA gel showing the expression and purification of MxiG-$C_{142-371}$. 1&2: pre-induction samples from B834 (DE3) pLysS pET22b-*mxiGC*$_{142-371}$*(C296S)*, 3&4: post-induction samples from B834 (DE3) pLysS pET22b-*mxiGC*$_{142-371}$*(C296S)*, 5: homogenate, 6: insoluble pellet after 26,000 x g centrifugation, 7: supernatant after 26,000 x g centrifugation, 8: flow-through from loading Ni-cartridge, 9: flow-through from 50 mM wash of Ni-cartridge and 10-14: fractions 1-5 from elution of Ni-cartridge with 300 mM imidazole.

B 15% SDS-PA gel showing fractions containing MxiG-$C_{142-371}$ after SEC. Samples are shown relative to the position of the fraction in the MxiG-$C_{142-371}$ elution profile.

C Elution profile for MxiG-$C_{142-371}$ from a Superdex 200 10/300 GL column. Elution with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl.

D 12% SDS-PA reducing gel comparing MxiG-$C_{142-371}$ purified from either pET14b-*mxiGC*$_{142-371}$ or pET22b-*mxiGC*$_{142-371}$*(C296S)*.

successfully localised to the periplasm, it did not seem to have been a successful strategy for purification of the stable, homogeneous, high yield sample required for structural characterisation, indicating another approach was needed.

Purification of MxiG-$C_{142-371}$ from the Insoluble Fraction

As the majority of MxiG-$C_{142-371}$ was found within the insoluble cell fraction, it was next decided to attempt a refold protocol for the purification of the protein. In order to maximise incorporation of MxiG-$C_{142-371}$ into inclusion bodies, protein expression in 2 x 1 l LB cultures of B834 (DE3) pLysS pET14b-*mxiGC*$_{142-371}$ was induced with 0.4 mM IPTG and left to proceed for 3 h at 37^{o}C. The cells were then harvested and resuspended in lysis buffer (50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 10 mM EDTA supplemented with standard amounts of lysozyme and deoxyribonuclease (DNase) I (bovine pancrease; Sigma)). Cells were homogenised and the resulting lysate mixed with 0.05% Tween for 5 min at 4^{o}C. The insoluble fraction was then separated by centrifugation at 28,000 x g for 30 min at 4^{o}C. Inclusion bodies found within the resulting pellet were then washed and solubilised as outlined in **Table 4.4**, resulting in a relatively clean solution of denatured MxiG-$C_{142-371}$ (**Figure 4.8**).

Table 4.4 Protocol for inclusion body preparation

1. Pellet resuspended in 40 ml lysis buffer with a hand-held homogeniser. Mixed with 0.05% Tween for 1 h at 4^{o}C.
2. Centrifugation at 28,000 x g for 20 min at 4^{o}C.
3. Pellet resuspended in 40 ml solubilisation buffer (100 mM Tris-HCl (pH 8.5), 8 M urea, 1 mM EDTA, 25 mM DTT) with a hand-held homogeniser. Mixed for 1 h at 4^{o}C.
4. Solution acidified quickly to pH 3-4 with HCl.
5. Centrifugation at 33,000 x g for 20 min at 4^{o}C.
6. Supernatant applied to 10000 MWCO SnakeSkin Dialysis Tubing (Thermo scientific) and dialysed against 3 x 1 l 6 M urea (pH 3-4) at 4^{o}C overnight.

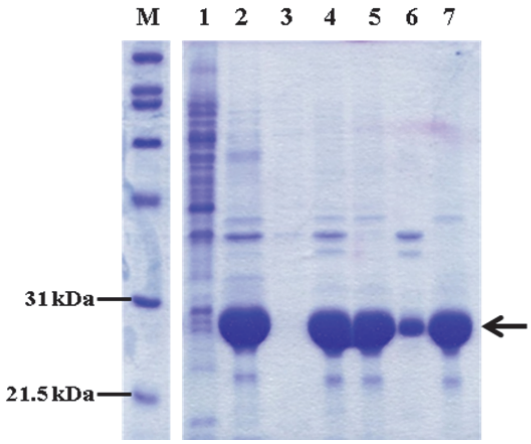


Figure 4.8 *Preparation of inclusion bodies from BS34 (DE3) pLysS pET14b-mxiGC$_{142-371}$.*

M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad), 1: supernatant after first 28,000 x g spin (of cell lysate), 2: pellet after first 28,000 x g spin (of cell lysate), 3: supernatant after second 28,000 x g spin (of insoluble fraction), 4: supernatant after second 28,000 x g spin (of insoluble fraction), 5: supernatant after 33,000 x g spin (of solublised inclusion bodies), 6: pellet after 33,000 x g spin (of solubilised inclusion bodies) and 7: post-dialysis against 6 M urea (pH 3-4). The arrow denotes the band corresponding to MxiG-C$_{142-371}$.

To find appropriate oxidising conditions for refolding MxiG-C$_{142-371}$, different refold buffers were trialled. Although each buffer contained 25 mM ethanolamine, 1 mM EDTA, 1 mM cysteine and 2 mM cystine, pH, additional buffering reagent and reagent used to minimise protein aggregation were varied (**Table 4.5**). The dialysed protein was divided into four 10 ml aliquots, each of which was added drop-wise using a peristaltic pump to 500 ml thoroughly degassed refold buffer. Refolding was left to proceed without stirring overnight at 4^{o}C.

Table 4.5 Refold buffers tested for MxiG-C$_{142-371}$

Buffer	Buffering Reagent	Reagent used to minimise aggregation	pH
A	50 mM Tris-HCl	10 % glycerol	9
B	50 mM Tris-HCl	10 % glycerol	11
C	500 mM arginine		9
D	500 mM arginine		11

12 ml was then extracted from each MxiG-C$_{142-371}$ refold and concentrated to 1.5 ml using an Amicon Ultra centrifugal device with a 10 kDa cut-off (Millipore). A 1 ml sample was subsequently injected on to a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. The elution profiles obtained from all

refold buffers had a single symmetric main peak (**Figure 4.9**), which indicated a high yield of a single species, in contrast to MxiG-$C_{142-371}$ purified from the soluble fraction. Assuming equivalent concentrating of the samples and their loading on to the column, the peak heights of the main elution peaks suggested the yield obtained with refold buffer C (500 mM arginine (pH 9.0), 25 mM ethanolamine, 1 mM EDTA, 1 mM cysteine, 2 mM cystine) was the highest (**Figure 4.9A, blue**), allowing this buffer to be selected for optimal MxiG-$C_{142-371}$ refolding. The presence of clean MxiG-$C_{142-371}$ in fractions eluted across the main peak from refold buffer C was subsequently confirmed by SDS-PAGE analysis (**Figure 4.9, inset**).

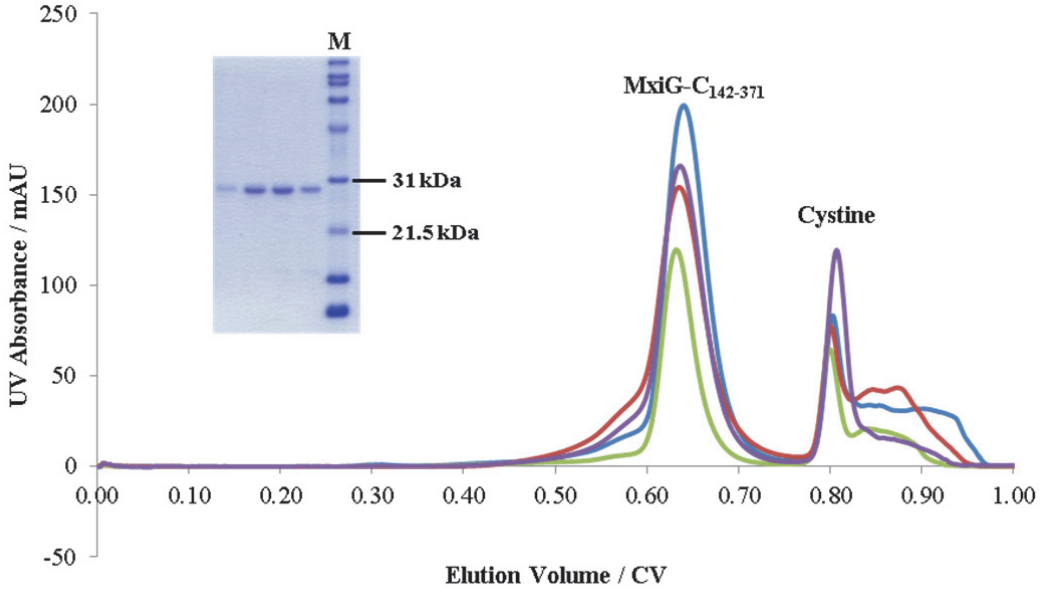


Figure 4.9 *Refold trials for MxiG-$C_{142-371}$.*

A Zoomed excerpt of the elution profile for refolded MxiG-$C_{142-371}$ after SEC using a Superdex 200 10/300 GL column. Refolded samples were concentrated eight times before injection on to the column. Elution with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. Green: refold buffer A (Tris/glycerol pH 9), purple: refold buffer B (Tris/glycerol pH 11), blue: refold buffer C (arginine pH 9) and red: refold buffer D (arginine pH 11). The identities of the peaks are shown. The inset shows a 15% SDS-PA gel of MxiG-$C_{142-371}$ fractions taken across the main elution peak obtained with refold buffer C. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

In subsequent purifications, the sample of denatured MxiG-$C_{142-371}$ obtained after inclusion body preparation (**Table 4.4**) was added to 2 l refold buffer. The refolded protein was then concentrated to ~50 ml using a Vivaflow 200 with a 10 kDa cutoff (Sartorius) attached to a Masterflex L/S pump. A significant amount of precipitation occurred during this step, indicating

aggregation of any MxiG-$C_{142-371}$ that was still denatured. However, estimation of the protein concentration before and after refolding showed only ~15 % of MxiG-$C_{142-371}$ had precipitated, indicating a high efficiency of protein refolding.

Refolded MxiG-$C_{142-371}$ was dialysed against 4 x 1 l final SEC buffer overnight at 4^{o}C, resulting in further precipitation and a ~20% reduction in yield with respect to pre-dialysis. This sample was subsequently concentrated to 5 ml with a centrifugal device and injected on to a HiLoad 16/60 Superdex 200 pg column (GE Healthcare). Although the initial SEC buffer used was the same as in the refold trials (**Figure 4.9**), the optimal buffer for MxiG-$C_{142-371}$ was found using a fluorescence-based thermal shift assay (ThermoFluor)(**Appendix 8.2.6**). In essence, increased binding of the SYPRO Orange dye (Sigma) to hydrophobic groups upon protein unfolding allowed a midpoint temperature for the transition to be determined (T_m), providing an assay to determine optimal buffer conditions for protein stability. In the initial calibration, MxiG-$C_{142-371}$ was found to undergo an unfolding transition with a T_m around 30^{o}C, with an optimal SYPRO Orange dye dilution of 1:300 and protein concentration of 0.18 mg/ml (**Figure 4.10A**). Although the presence of an unfolding transition indicated refolded MxiG-$C_{142-371}$ had tertiary structure, high initial dye binding and a low T_m suggested the protein was either partially unfolded, had exposed hydrophobic patches or that the dye was destabilising the protein fold. When the ThermoFluor assay was carried out under different buffer conditions (**Figure 4.10B**), those with a pH <6.0 were seen to dramatically destabilise MxiG-$C_{142-371}$. In addition, conditions with either lower salt concentrations or 30% glycerol seemed to lead to a small increase in T_m values. As glycerol could impede nucleation events during protein crystallisation (Sedgwick *et al.*, 2007), the SEC buffer chosen for the final purification of MxiG-$C_{142-371}$ based on the ThermoFluor results was 20 mM HEPES (pH 8.0), 50 mM NaCl, reproducibly leading to the elution profile shown in **Figure 4.10C**. Estimation of the concentration of the final MxiG-$C_{142-371}$ sample indicated a typical yield of 5 mg protein per 1 l B834 (DE3) pLysS pET14b-*mxiGC*$_{142-371}$, a 100-fold increase on the best yield obtained when MxiG-$C_{142-371}$ was purified from the soluble fraction.

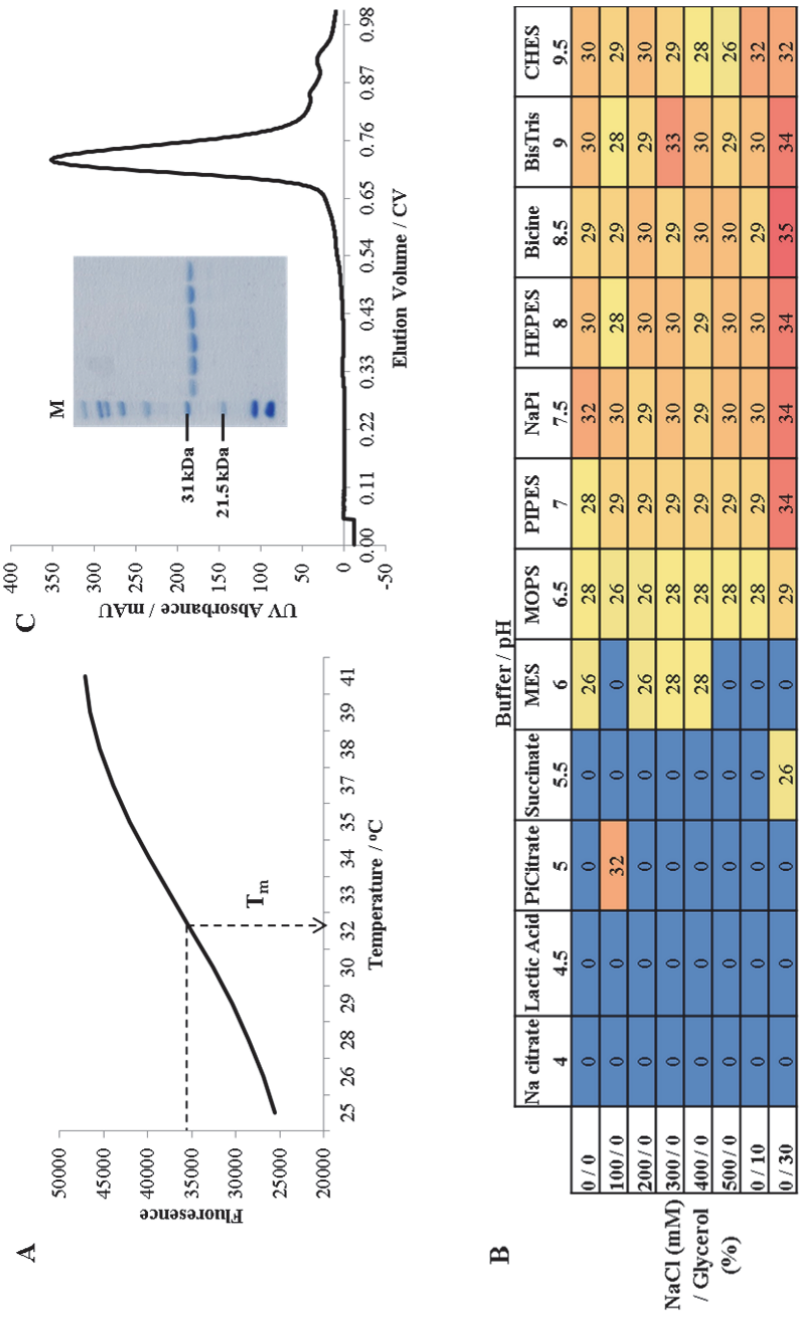


Figure 4.10 *Optimisation of the SEC buffer for MxiG-$C_{142-371}$ purification*.

A The unfolding transition of MxiG-$C_{142-371}$ obtained via ThermoFluor. The fluorescence of a reaction mixture (30% 1:100 SYPRO Orange dye (1:300 final) and 70% 0.26 mg/ml (0.18 mg/ml final) MxiG-$C_{142-371}$ in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl) equilibrated at 25°C for 5 min followed by 36 cycles with +1.3°C temperature increments was measured. An approximate T_m value for the transition is indicated.

B T_m values for the unfolding of MxiG-$C_{142-371}$ obtained from reaction mixtures containing 50% of the buffer screen shown, 30% 1:100 SYPRO Orange dye (1:300 final) and 20% 1 mg/ml (0.18 mg/ml final) MxiG-$C_{142-371}$ in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. Results are shown in the 96-well format used for the screen, with the buffer (100 mM) and pH above each column and other additives and their concentration next to each row. T_m values were taken as the minimum of the negative derivative of curves like that shown in **A** and are coloured according to magnitude from blue (low) to red (high).

C Elution profile for MxiG-$C_{142-371}$ from a HiLoad 16/60 Superdex 200 pg column. Elution with 20 mM HEPES (pH 8.0), 50 mM NaCl. The 15% SDS-PA gel shown in the inset shows samples taken from across the main elution peak. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

Limited Proteolysis of MxiG-$C_{142-371}$

MxiG-$C_{142-371}$ was not stable in solution, with ~20% of the total amount precipitating in one week at 4°C. Therefore limited proteolysis with trypsin was attempted, to see whether a more stable core structure could be obtained. 20 µl reactions containing 20 µg MxiG-$C_{142-371}$ in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl were supplemented with varying amounts of trypsin and incubated for 2 h at room temperature. SDS-PAGE analysis of these samples showed that limited proteolysis of MxiG-$C_{142-371}$ had occured (**Figure 4.11A**), forming a smaller single species and thus providing further evidence the refolded protein had tertiary structure. As optimal digestion was obtained with a w/w protein:trypsin ratio of 250:1 (**Figure 4.11A, asterisk**), mass spectrometry analysis of this sample was carried out as described in Chapter 2.

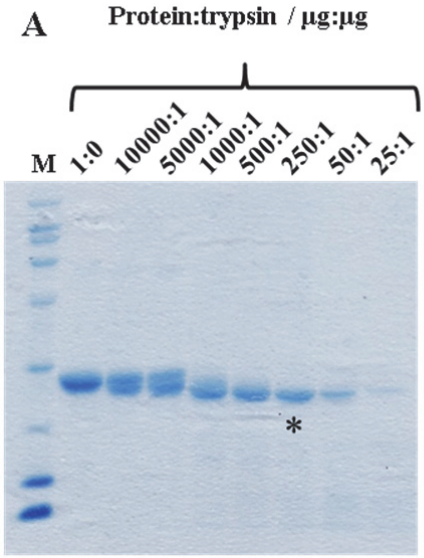
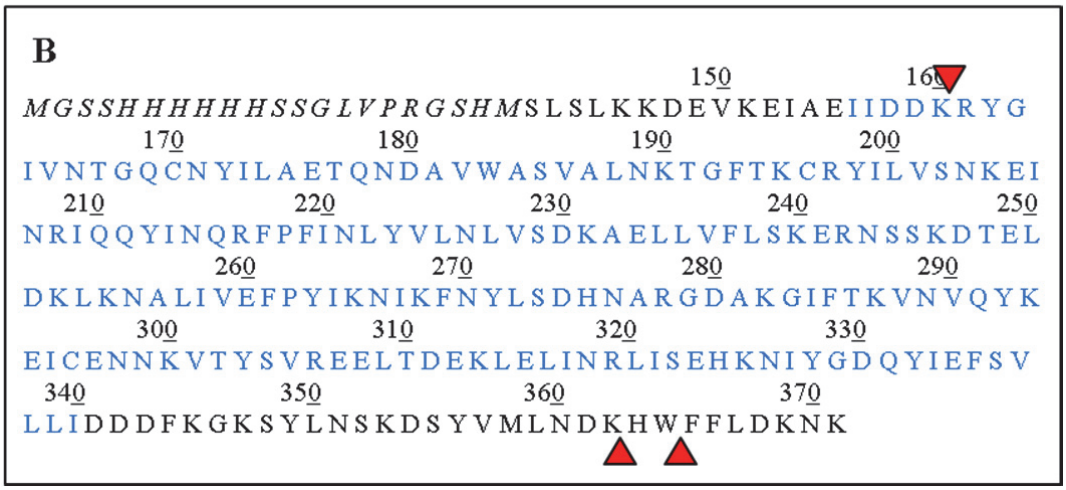


Figure 4.11 *Limited proteolysis of MxiG-$C_{142-371}$.*

A 15% SDS-PA gel showing limited proteolysis of MxiG-$C_{142-371}$ with varying amounts of trypsin. 20 µl reactions comprising 20 µg MxiG-$C_{142-371}$ and the indicated w/w amount of trypsin in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl were incubated for 2 h at room temperature before supplementing with 1 x loading dye. The protein:trypsin ratio giving optimal digestion of MxiG-$C_{142-371}$ is shown with an asterisk. M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad)

B The sequence of MxiG-$C_{142-371}$ numbered according to the native *S. flexneri* MxiG sequence. The trypsin cleavage sites used during limited proteolysis are shown by red triangles. There are two alternative cleavage sites at the C-terminus. The sequence homologous to $PrgH_{170-362}$ is shown in blue and the N-terminal His-tag is shown in italics.



This showed that a 23.6 kDa and a 23.9 kDa species had been formed, with the latter being more prevalent. Upon comparison to the sequence of MxiG-$C_{142-371}$, these species could be identified as being residues 161-362 and residues 161-364 of MxiG-C (MxiG-$C_{161-362/4}$), indicating removal of both the N-terminal His-tag and a small portion of the C-terminus (**Figure 4.11B)**.

In order to carry out a large-scale purification of MxiG-$C_{161-362/4}$, the concentration of MxiG-$C_{142-371}$ was estimated after dialysis into 20 mM HEPES (pH 8.0), 50 mM NaCl and the appropriate amount of trypsin for a w/w protein:trypsin ratio of 250:1 was added before incubation for 2 h at room temperature. The digested sample was then concentrated to 5 ml, a process which occurred more quickly than for MxiG-$C_{142-371}$, indicating less precipitation. SDS-PAGE analysis of this sample showed the digestion was complete, producing MxiG-$C_{161-362/4}$ (**Figure 4.12**). As the molecular weight of trypsin (23.3 kDa) is similar to those of MxiG-$C_{142-362/4}$, the proteins would not be separated by SEC. Therefore the sample was flowed through a 1

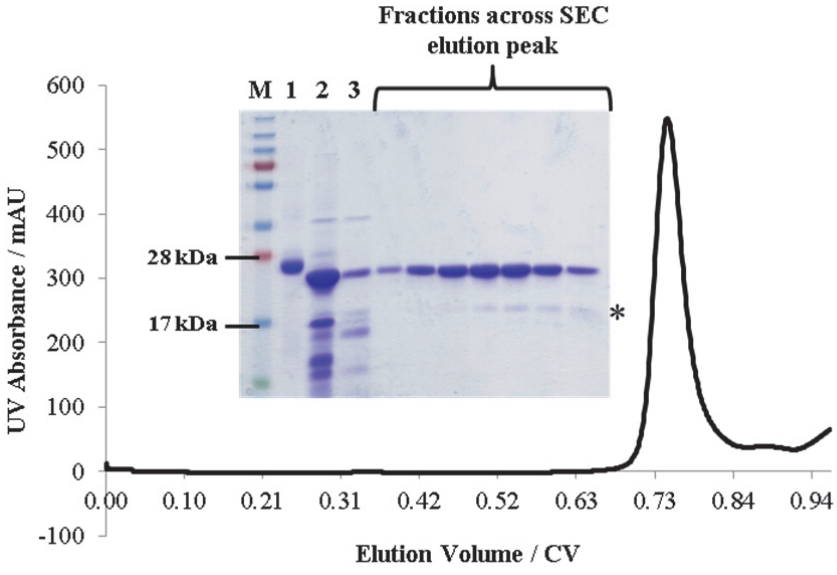


Figure 4.12 *Purification of MxiG-$C_{161-362/4}$.*

Elution profile for MxiG-$C_{161-362/4}$ from a HiLoad 26/60 Superdex 200 pg column. Elution with 20 mM HEPES (pH 8.0), 50 mM NaCl, 1 mM EDTA. The 15% SDS-PA gel shown in the inset shows samples taken from across the main elution peak and before SEC. M: PageRuler Plus Prestained Protein Ladder (Fermentas), 1: post-dialysis of MxiG-$C_{142-371}$, 2: post-digestion with trypsin and 3: post-removal of trypsin by the benzamidine column. The asterisk denotes a minor contaminating fragment of MxiG-$C_{142-371}$.

ml HiTrap Benzamidine FF column (GE Healthcare) at 1 ml/min using a peristaltic pump to remove trypsin from the sample. The column was washed with 5 ml 1 M NaCl, 15 mM HEPES (pH 8.0) to remove non-specifically bound MxiG-$C_{161-362/4}$, although some protein remained bound to the column (**Figure 4.12**). The total flow-through was then injected on to a HiLoad 26/60 Superdex 200 pg column (GE Healthcare) pre-equilibrated with SEC buffer supplemented with 1 mM EDTA. As for MxiG-$C_{142-371}$, the elution profile showed a single symmetric main peak, which corresponded to MxiG-$C_{161-362/4}$ when the fractions were analysed by SDS-PAGE (**Figure 4.12**). Although most of the smaller fragments produced during limited proteolysis were separated from MxiG-$C_{161-362/4}$ during SEC, later fractions within the main elution peak contained a small amount of a lower molecular weight contaminant (**Figure 4.12, asterisk**). When this gel band was subsequently subjected to trypsin digestion and mass spectrometry identification (Central Proteomics Facility, University of Oxford) it was shown to be MxiG-C (mascot score = 489), indicating a further minor proteolytic fragment.

The typical yield of MxiG-$C_{161-362/4}$ was 4 mg protein per 1 l B834 (DE3) pLysS pET14b-*mxiGC*$_{142-371}$ indicating relatively little was lost during limited proteolysis and subsequent purification. In addition, dialysis of 2 mg/ml MxiG-$C_{161-362/4}$ into buffers containing 20 mM HEPES (pH 7-8), 0-150 mM NaCl, 1 mM EDTA showed it was stable in solution for >1 week. In combination with the increased speed of concentrating, this implied an increased solubility of MxiG-$C_{161-362/4}$ with respect to MxiG-$C_{142-371}$, potentially due to the loss of the hydrophobic (HW)FFL motif at the C-terminus during limited proteolysis (**Figure 4.11B**). Interestingly, loss of just four residues at the C-terminus of PrgH is sufficient to destabilise the interaction between the IMR and OMR (Schraidt *et al.*, 2010), with subsequent cross-linking between the C-terminus of PrgH and the secretin InvG confirming their *in vivo* proximity (Sanowar *et al.*, 2010). The presence of an analogous HWYF motif at the extreme C-terminus of PrgH hints that the nature of the interaction with the secretin is conserved between MxiG-C homologues, allowing an interesting series of experiments for the MxiG and MxiD periplasmic domains to be hypothesised. An interaction with the MxiD periplasmic domain (**Figure 4.1**) could both

stabilise MxiG-$C_{142-371}$ and protect the C-terminus from trypsin cleavage (**Figure 4.11B**). In addition, MxiG-$C_{161-362/4}$ may have an altered affinity for MxiD relative to MxiG-$C_{142-371}$. Unfortunately, an apparent inaccessibility of the MxiD sequence to cloning, recombinant expression and purification (**Appendix 8.4.1**) precluded such experiments.

Comparison of the sequence of MxiG-$C_{161-362/4}$ to $PrgH_{170-362}$, showed that MxiG-$C_{161-362/4}$ has a significant portion at the C-terminus that is structurally uncharacterised (**Figure 4.11B**). Therefore MxiG-$C_{161-362/4}$ provided an ideal and interesting sample for further structural characterisation.

Characterisation of MxiG-C

Although ThermoFluor and limited proteolysis indicated refolded MxiG-$C_{142-371}$ had a core tertiary structure, circular dichroism (CD) spectroscopy was carried out to ascertain whether the protein had the expected proportions of secondary structure elements with respect to the $PrgH_{170-362}$ structure (Spreter et al., 2009). MxiG-$C_{142-371}$ was exchanged into 25 mM Na_2HPO_4 (pH 8.0) by running down a Superdex 200 10/300 GL column, resulting in a 0.1 mg/ml sample. CD spectroscopy was carried out at 25^{o}C on a JASCO J-715 Spectropolarimeter, with four repeats of wavelength scans between 195 and 300 nm. The CD spectrum for MxiG-$C_{142-371}$ showed characteristic troughs at 208 and 220 nm (**Figure 4.13A**), indicative of a dominant α-helical conformation. The KD2 server (Andrade *et al.*, 1993) subsequently predicted MxiG-$C_{142-371}$ to have 57% α-helix, 8% β-sheet and 34% random coil based on the far-UV portion of the spectrum (200-240 nm). Although this differed from the 47% α-helix and 23% β-sheet observed in the structure of $PrgH_{170-362}$ (Spreter *et al.*, 2009), the far-UV spectrum of MxiG-$C_{142-371}$ (**Figure 4.13B, black**) was very similar to the equivalent spectrum predicted by the K2D server for $PrgH_{170-362}$ from these proportions (**Figure 4.13B, red**). This indicated some limitations in the accuracy of predicting secondary structure from CD spectra. Therefore refolded MxiG-$C_{142-371}$ was likely to have the expected secondary structure, especially given that 66 residues of

MxiG-$C_{142-371}$ are not represented by the $PrgH_{170-362}$ structure and likely contributed to the differences in the CD spectra in **Figure 4.13B**.

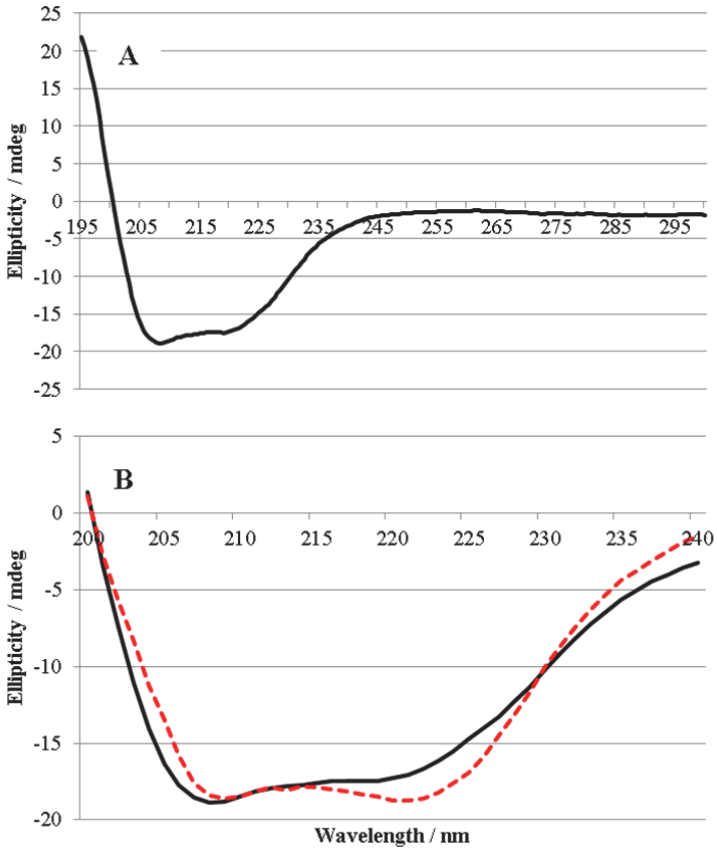


Figure 4.13 *CD spectroscopy of MxiG-$C_{142-371}$.*

A CD spectrum for 0.1 mg/ml MxiG-$C_{142-371}$ acquired at 25°C with four repeats of wavelength scans between 195 and 300 nm.

B Comparison of the far-UV (200-240 nm) region of the CD spectrum for MxiG-$C_{142-371}$ (black) and the predicted far-UV CD spectrum for $PrgH_{170-362}$ (red). The $PrgH_{170-362}$ spectrum was predicted using the K2D server (Andrade *et al.*, 1993), based on secondary structure proportions of 47% α-helix and 23% β-sheet in the crystal structure (Spreter *et al.*, 2009).

Subsequently, ^{15}N-labelled MxiG-$C_{142-371}$ was expressed in BL21 (DE3) pLysS cells as described for MxiG-N_{1-126} in Chapter 2, allowing purification of ^{15}N-MxiG-$C_{161-362/4}$ in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl. A 240 µM sample supplemented with 5% D_2O was used to collect a ^{1}H,^{15}N-HSQC spectrum for MxiG-$C_{161-362/4}$ (**Figure 4.14**). The spectrum showed good peak dispersion in both dimensions, providing further evidence that MxiG-$C_{161-362/4}$ had refolded into an ordered tertiary structure. However, severe peak overlap in the central region (**Figure 4.14, dashed box**) indicated that the spectrum would be difficult to assign and thus NMR spectroscopy would not be an ideal method for structural characterisation of MxiG-$C_{161-362/4}$.

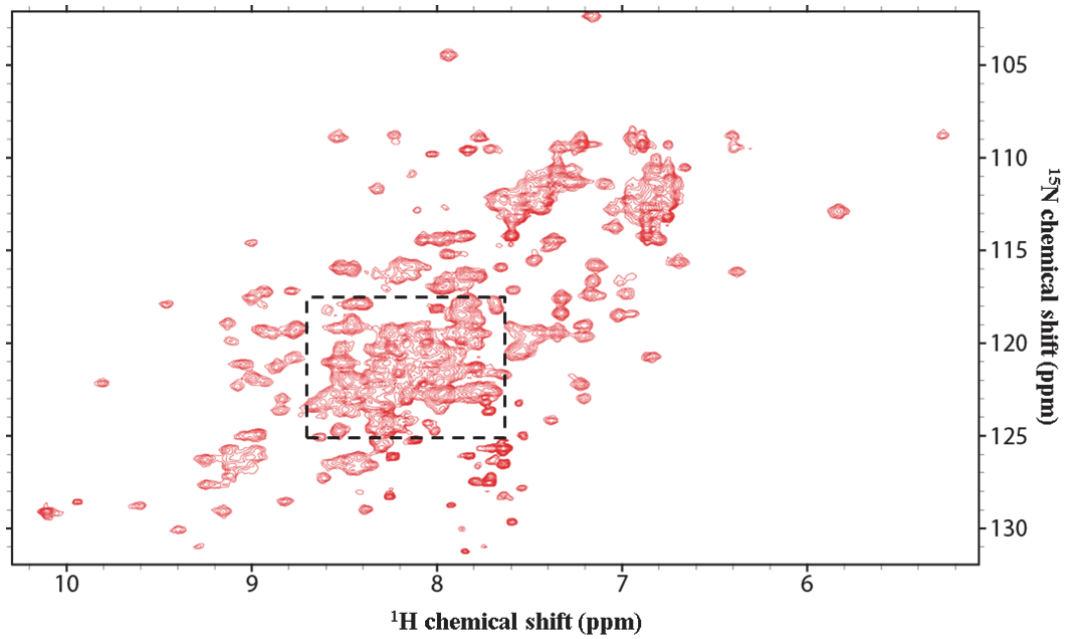


Figure 4.14 *^{1}H,^{15}N-HSQC spectrum of 240 µM ^{15}N-MxiG-C$_{161-362/4}$.*

Spectrum acquired in 19 mM Tris-HCl (pH 7.5), 48 mM NaCl, 5 % D_2O at 25°C. The dashed box indicates the region of most severe peak overlap.

The number of peaks and the signal-to-noise ratio seen in the ^{1}H,^{15}N-HSQC spectrum of MxiG-C$_{161-362/4}$ indicated the protein was more likely to be tumbling in solution as a ~23 kDa monomer rather than as a larger oligomer. To investigate the oligomeric state of MxiG-C further, Dr Steven Johnson carried out MALS on 65 µM samples of both MxiG-C$_{142-371}$ and MxiG-C$_{161-362/4}$. The in-line multi-angle scattering across the only elution peak showed a molecular weight consistent with a monomer for both constructs (**Figure 4.15**), indicating MxiG-C does not self-

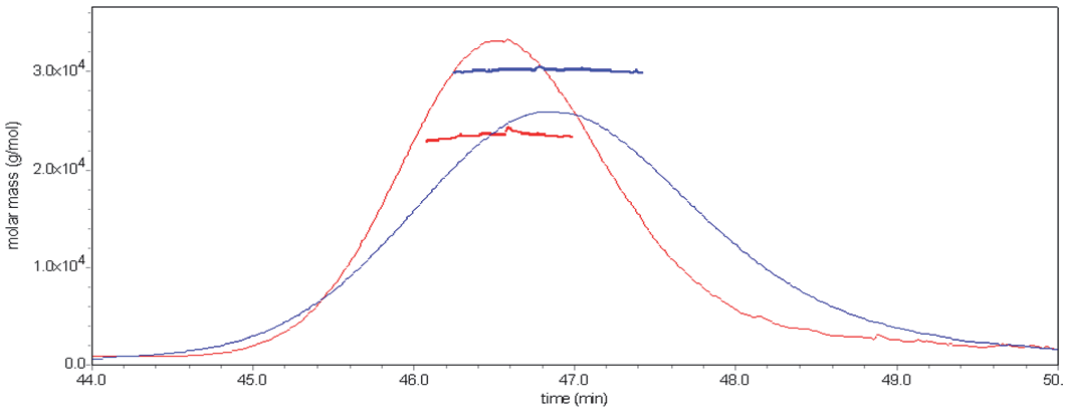


Figure 4.15 *Analysis of the oligomeric state of MxiG-C constructs by SEC and in-line multi-angle light scattering.*

The left axis represents the molecular mass at any given point in the elution profile. Elution profiles are shown for the only elution peak for 65 µM MxiG-C$_{142-371}$ (blue) and 65 µM MxiG-C$_{161-362/4}$ (red) and demonstrate these domains are monomeric at these concentrations.

associate via high affinity interactions. However, an interaction between MxiG-C domains with a K_d > 240 μM could still occur to stabilise MxiG within the assembled IMR, as was discussed in the previous chapter in the context of MxiG-N. In addition, MxiG-$C_{161-362/4}$ unexpectedly eluted from the column slightly earlier than MxiG-$C_{142-371}$ (**Figure 4.15**), indicating limited proteolysis had altered the protein shape or interaction with the column matrix.

Crystallisation Trials With MxiG-$C_{161-362/4}$

To summarise, MxiG-$C_{142-371}$ purified from the insoluble fraction of B834 (DE3) pLysS pET14b-*mxiGC*$_{142-371}$ provided a high yield, homogeneous sample that appeared to be correctly refolded. Limited proteolysis subsequently removed portions of the presumably flexible termini, further enhancing the sample stability and solubility. The resulting MxiG-$C_{161-362/4}$ species provided an ideal sample for structural characterisation, therefore crystallisation trials were carried out.

In general, crystallisation occurs when a protein slowly precipitates from solution, facilitating the formation of an ordered crystal lattice. Crystal nucleation and formation are hugely dependent on a large number of variables, including temperature, pH and the concentration and nature of the protein, salts and organic precipitants present in solution. Therefore high-throughput screening of purified MxiG-$C_{161-362/4}$ against a wide range of crystallisation conditions was required, with 96 conditions being tested per crystallisation plate. A Hydra 96 (Alpha Biotech Ltd) was used to transfer 75 μl of each condition from a 96-well screen to a MRC Crystallisation Plate (Molecular Dimensions), forming reservoirs of mother liquor. HT-96 broad screens from Molecular Dimensions (**Table 4.6**) were used in an attempt to identify conditions that encourage MxiG-$C_{161-362/4}$ crystallisation.

Crystallisation trials were conducted using the vapour-diffusion sitting-drop method, whereby a small drop of protein solution mixed with a precipitant-containing mother liquor was dispensed in a well positioned next to the larger reservoir of mother liquor, allowing gradual dehydration of the drop. 400 nl drops containing both 1:1 and 3:1 protein solution:mother liquor were

dispensed in the wells of the MRC Crystallisation Plate using an OryxNano Crystallisation Robot (Douglas Instruments). Plates were then sealed with a StarSeal Advanced Polyolefin Film (Type D; Star Lab) and incubated at 21^{o}C.

Table 4.6 HT-96 broad screens from Molecular Dimensions used in crystallisation trials

Screen	**Properties**
The Stura Footprint Screen + MacroSol	Screening of the protein solubility curve by systematic variation of pH and salt/organic precipitants.
JCSG-*plus*	Sparse matrix screen based on previous successful crystallisation conditions – wide range of precipitants, salts, buffers and pH.
Structure Screen I & II	Sparse matrix screen based on previous successful crystallisation conditions – wide range of precipitants, salts, buffers and pH.
PACT *premier*	Systematic screening of cations, anions and pH with PEG 3350 as the organic precipitant.
Morpheus	Low molecular weight ligands previously found to promote the formation of crystal contacts combined with a range of precipitants, salts, buffers and pH.
ProPlex	Sparse matrix screen based on previous successful crystallisation screens for protein complexes.

All the crystallisation trials with MxiG-$C_{161-362/4}$ are summarised in **Table 4.7**. To provide an estimate of an appropriate MxiG-$C_{161-362/4}$ concentration for initial crystallisation trials, a 'Stura-test' was carried out; an appropriate concentration is indicated by the presence of precipitation 60 to 180 s after mixing 1 µl protein solution with 1 µl condition A7 from the Stura Footprint Screen (20% PEG 4000, 0.2 M imidazole malate (pH 7.0)). The Stura-test indicated an initial MxiG-$C_{161-362/4}$ concentration around 2 mg/ml should be used for crystallisation trails, so the broad screens (**Table 4.6**) were set-up with 2.2 mg/ml protein. After 3 months no protein crystals had grown and most drops showed relatively minor precipitation, so screening was repeated with a slightly higher concentration of MxiG-$C_{161-362/4}$ (2.8 mg/ml). Unfortunately this still did not lead to any crystallisation hits.

Table 4.7 Crystallisation trials with MxiG-$C_{161-362/4}$

Sample	Purification Buffer	Broad Screens (Table 4.6)
2.2 mg/ml MxiG-$C_{161-362/4}$	20 mM HEPES (pH 8.0), 50 mM NaCl	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex
2.8 mg/ml MxiG-$C_{161-362/4}$	20 mM HEPES (pH 8.0), 50 mM NaCl, 1 mM EDTA	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex
2.7 mg/ml MxiG-$C_{161-362/4}$ and 1 M TMAO	20 mM HEPES (pH 8.0), 50 mM NaCl, 1 mM EDTA	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex
3.6 mg/ml methylated MxiG-$C_{161-362/4}$	20 mM HEPES (pH 8.0), 50 mM NaCl, 1 mM EDTA	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, ProPlex

As purified MxiG-$C_{161-362/4}$ did not crystallise alone, further sample modifications were tried. The osmolyte trimethylamine-N-oxide (TMAO) has been shown to entropically stabilise unstructured proteins and encourage their crystallisation (Hill *et al.*, 2002). The low T_m of MxiG-$C_{142-371}$ shown by ThermoFluor (**Figure 4.10A**) could imply intrinsic instability in the protein fold. Therefore 1 M TMAO in combination with 2.7 mg/ml MxiG-$C_{161-362/4}$ was subjected to crystallisation trials in an attempt to stabilise the protein backbone.

Methylation of lysine side-chains has also been shown to encourage the crystallisation of proteins that have been otherwise inaccessible to crystallography (Walter *et al.*, 2006). As MxiG-$C_{161-362/4}$ has 20 lysine residues, this modification had the potential to significantly affect protein properties. After MxiG-$C_{161-362/4}$ was passed through the HiTrap Benzamidine FF column, the protein was divided into 1 ml aliquots at 1 mg/ml. Each aliquot was subjected to two rounds of mixing with 40 µl 1 M formaldehyde and 20 µl 1 M dimethylamine-borane

complex (ABC) for 2 h in the dark at 4°C. 10 µl 1 M ABC was then added to each aliquot, before continued mixing overnight. The reaction was finally quenched with 20 mM Tris-HCl (pH 8.0) and centrifuged at 16,000 x g to remove precipitate. SEC of methylated MxiG-$C_{161-362/4}$ then proceeded as for the unmethylated protein. Mass spectrometry analysis confirmed the mass expected for complete methylation of MxiG-$C_{161-362/4}$ lysine side-chains. The Stura test subsequently indicated methylation had increased the solubility of the protein, so broad screening was carried out with 3.6 mg/ml methylated MxiG-$C_{161-362/4}$. However, neither methylation nor addition of TMAO led to crystallisation of MxiG-$C_{161-362/4}$.

In summary, all the MxiG-$C_{161-362/4}$ crystallisation trials outlined in this work were unsuccessful in producing protein crystals, suggesting further modification of the crystallisation variables may be required. With the current samples, additional temperatures, protein solution:mother liquor ratios or MxiG-$C_{161-362/4}$ concentrations could be tested. Alternatively, improving the homogeneity of the MxiG-$C_{161-362/4}$ samples could facilitate crystallisation, for example by removing the minor contaminating MxiG-C fragment that co-elutes with MxiG-$C_{161-362/4}$ during SEC (**Figure 4.12**). Similarly, the two species present in MxiG-$C_{161-362/4}$ could be eliminated by directly expressing the shorter construct. However it is also possible the current MxiG-$C_{161-362/4}$ construct may not have the intrinsic propensity to crystallise; the presence of significant additional sequence relative to that represented by the homologous $PrgH_{170-362}$ structure (**Figure 4.11B**) and the large amount of peak overlap in the $^{1}H,^{15}N$-HSQC spectrum (**Figure 4.14**) could indicate that large regions at the termini of MxiG-$C_{161-362/4}$ are unstructured and flexible, likely impeding crystal contact formation. In this case a MxiG-$C_{154-340}$ construct that encompasses the region homologous to $PrgH_{170-362}$ (**Figure 4.2B**) would perhaps be more amenable to crystallisation. However, as the sequence identity between MxiG-$C_{154-340}$ and $PrgH_{170-362}$ is only 24% (**Figure 4.2A**), there are also likely to be minor structural differences not represented by the homology model that could make MxiG-C a difficult crystallisation target in comparison to $PrgH_{170-362}$.

4.2 Purification and Characterisation of MxiJ Periplasmic Domains

Around 90% of MxiJ is predicted to be localised in the periplasm, with just the short portion at the extreme C-terminus being localised to the inner membrane and cytoplasm (**Figure 4.1**). High resolution structures of the periplasmic domain of EPEC EscJ provide the only insight into the structure of MxiJ homologues, showing two mixed α/β domains connected by a flexible linker (Crepin *et al.*, 2005, Yip *et al.*, 2005). Furthermore, the packing of EscJ as a superhelix within the crystal allowed a molecular model for a 24-subunit ring to be proposed, although a translation was required to flatten the ring (Yip *et al.*, 2005). Based on a sequence alignment of MxiJ with EscJ (Crepin *et al.*, 2005) which shows 23% sequence identity, residues 20 to 189 of MxiJ ($MxiJ_{20\text{-}189}$) were homologous to the region of EscJ represented by the crystal structure (pdb id 1yj7)(Yip *et al.*, 2005). This allowed a SCWRL homology model for $MxiJ_{20\text{-}189}$ to be constructed and the location of the flexible region shown by NMR spectroscopy (Crepin *et al.*, 2005) and the N- and C-terminal sub-domains to be predicted (**Figure 4.16**).

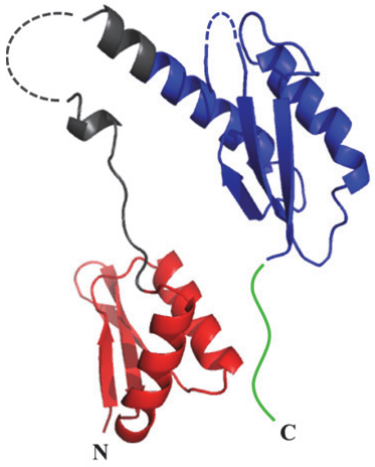


Figure 4.16 *Structural model for $MxiJ_{20\text{-}189}$.*

SCWRL homology model (Altschul *et al.*, 1990) for $MxiJ_{20\text{-}189}$ constructed using the crystal structure of EscJ (pdb id 1yj7)(Yip *et al.*, 2005) and the sequence alignment between MxiJ homologues (Crepin *et al.*, 2005). Regions homologous to the N-terminal subdomain (red), C-terminal subdomain (blue) and the flexible linker identified by NMR spectroscopy (grey) (Crepin *et al.*, 2005) are shown. Loop regions equivalent to those missing in the EscJ crystal structure are shown by dashed lines. The C-terminal extension present in the $MxiJ_{19\text{-}206}$ construct is represented by a green line.

In previous work by Harriet Allison (Part II Project), various regions of the MxiJ periplasmic domain were cloned, expressed and purified, in an attempt to identify a construct suitable for further characterisation. A protocol for the purification of residues 19 to 206 of MxiJ with a MGSSHHHHHHSSGLVPRGSHM N-terminal tag ($MxiJ_{19\text{-}206}$) was optimised, which produced a homogeneous, stable sample in amounts suitable for structural studies. This construct

encompassed more of the MxiJ periplasmic domain than the region homologous to the EscJ structure, with an additional 17 residues at the C-terminus that are structurally uncharacterised (**Figure 4.16**) and inclusive of the PrgK region that cross-links to PrgH (Sanowar *et al.*, 2010). However, subsequent crystallisation trials with $MxiJ_{19\text{-}206}$ were unsuccessful. As NMR spectroscopy was used to obtain the first structure of EscJ (Crepin *et al.*, 2005), characterisation of $MxiJ_{19\text{-}206}$ by this technique was attempted, as outlined in this section.

pET14b-*mxiJ*$_{19\text{-}206}$ (donated by S. Johnson) was transformed into BL21 (DE3) pLysS cells, allowing the expression of ^{15}N-labelled $MxiJ_{19\text{-}206}$ using the protocol described for previous constructs. ^{15}N-$MxiJ_{19\text{-}206}$ was extracted from the soluble fraction as outlined for MxiG-$C_{142\text{-}371}$, except for the elution of 5 x 5 ml fractions from a 5 ml Ni-NTA Superflow Cartridge (Qiagen).

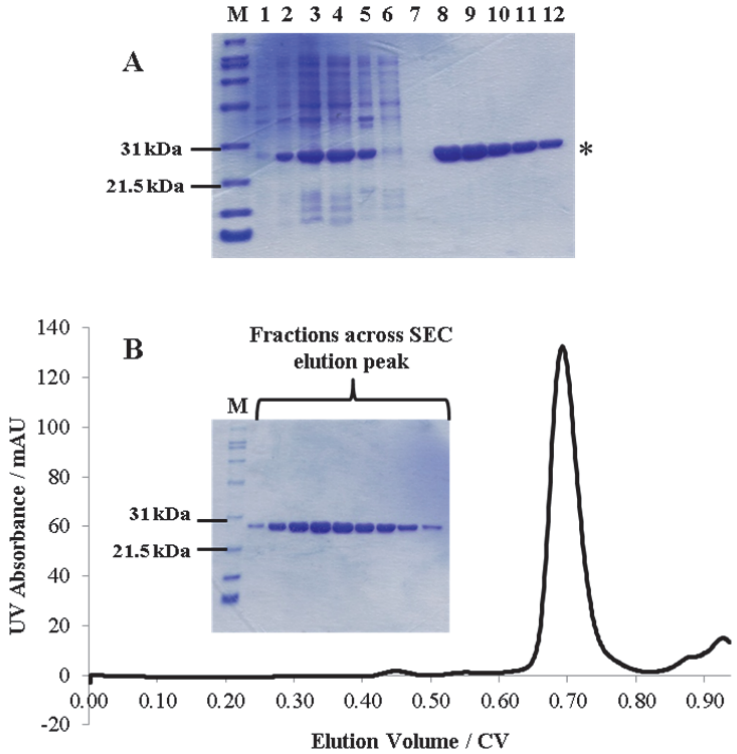


Figure 4.17 *The purification of ^{15}N-$MxiJ_{19\text{-}206}$.*

M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad)

A 15% SDS-PA gel showing the expression and purification of MxiG-$J_{19\text{-}206}$. 1: pre-induction sample from BL21 (DE3) pLysS pET14b-*mxiJ*$_{19\text{-}206}$, 2: post-induction sample from BL21 (DE3) pLysS pET14b-*mxiJ*$_{19\text{-}206}$, 3: homogenate, 4: supernatant after 26,000 x g centrifugation, 5: insoluble pellet after 26,000 x g centrifugation, 6: flow-through from loading Ni-cartridge, 7: flow-through from 50 mM wash of Ni-cartridge and 8-12: fractions 1-5 from elution of Ni-cartridge with 300 mM imidazole. The asterisk denotes the band corresponding to $MxiJ_{19\text{-}206}$.

B Elution profile for MxiG-$J_{19\text{-}206}$ from a HiLoad 26/60 Superdex 200 pg column. Elution with 25 mM Na_2HPO_4 (pH 6.8). The 15% SDS-PA gel in the inset shows samples taken across the main elution peak.

Although ^{15}N-MxiJ$_{19\text{-}206}$ runs slightly larger than expected for a 23.6 kDa protein during SDS-PAGE analysis, the protein was clearly overexpressed and present in the soluble fraction and subsequently in fractions 1 and 2 eluted from the Ni-cartridge (**Figure 4.17A**). These fractions were then injected on to a HiLoad 26/60 Superdex 200 pg column (GE Healthcare) and flowed at 3 ml/min with 25 mM Na_2HPO_4 (pH 6.8). The elution profile showed a single symmetric peak, corresponding to a homogeneous sample of ^{15}N-MxiJ$_{19\text{-}206}$ (**Figure 4.17B**). Concentration of ^{15}N-MxiJ$_{19\text{-}206}$ using an Amicon Ultra centrifugal device with a 10 kDa cut-off (Millipore) indicated a yield of ~3 mg protein per 1 l BL21 (DE3) pLysS pET14b-*mxiJ$_{19\text{-}206}$* culture grown.

A 150 µM sample of ^{15}N-MxiJ$_{19\text{-}206}$ supplemented with 5% D_2O was used to obtain a ^{1}H,^{15}N-HSQC spectrum for the protein (**Figure 4.18**). The peak dispersion in both dimensions was similar to that observed in the ^{1}H,^{15}N-HSQC spectrum for EscJ (Crepin *et al.*, 2005)(**Figure 4.18, inset**), suggesting MxiJ$_{19\text{-}206}$ had a core tertiary structure. However, the individual peaks were unresolved in the central region of the MxiJ$_{19\text{-}206}$ spectrum (**Figure 4.18, dashed box**), indicating assignment and structure determination by NMR spectroscopy would be challenging.

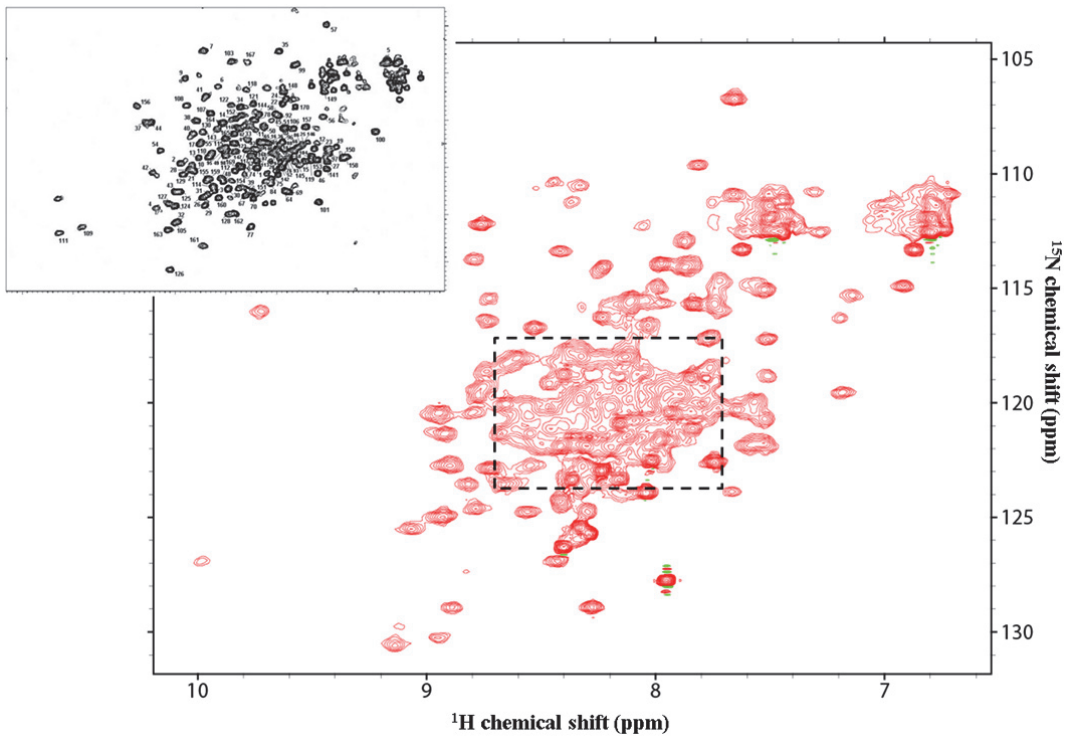
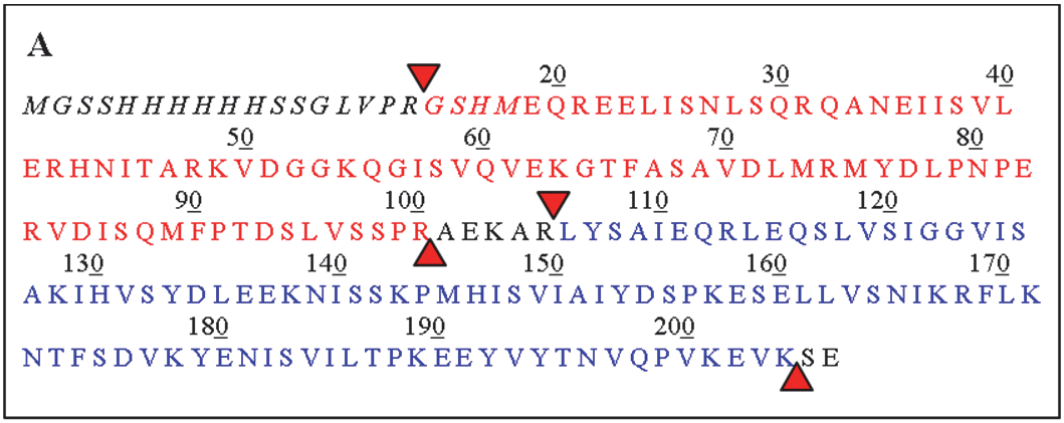


Figure 4.18 *^{1}H,^{15}N-HSQC spectrum of 150 µM ^{15}N-MxiJ$_{19\text{-}206}$.*

Spectrum acquired in 24 mM Na_2HPO_4 (pH 6.8), 5% D_2O at 25°C. The dashed box indicates the region of most severe peak overlap. The inset shows the ^{1}H,^{15}N-HSQC of EscJ extracted from Crepin et al., 2005.

Although there was severe peak overlap in the $^{1}H,^{15}N$-HSQC spectrum of the complete periplasmic domain of MxiJ, the two sub-domains in isolation could potentially produce spectra with improved resolution. Harriet Allison had previously optimised a protocol for limited proteolysis of $MxiJ_{19-206}$ with trypsin and subsequent purification of the fragments by SEC (Part II Project). Mass spectrometry analysis showed two fragments had been produced: a 9.6 kDa fragment encompassing residues 19-100 of MxiJ with a GSHM N-terminal tag ($MxiJ_{19-100}$) and a 11.3 kDa fragment encompassing residues 106-204 of MxiJ ($MxiJ_{106-204}$; **Figure 4.19A**). When these fragments were mapped on to the SCWRL model for $MxiJ_{20-189}$, $MxiJ_{19-100}$ corresponded to the N-terminal subdomain and the flexible linker, whilst $MxiJ_{106-204}$ corresponded to the C-terminal subdomain (**Figure 4.19B**). This provided a means by which to separate and characterise the individual subdomains of $MxiJ_{19-206}$.



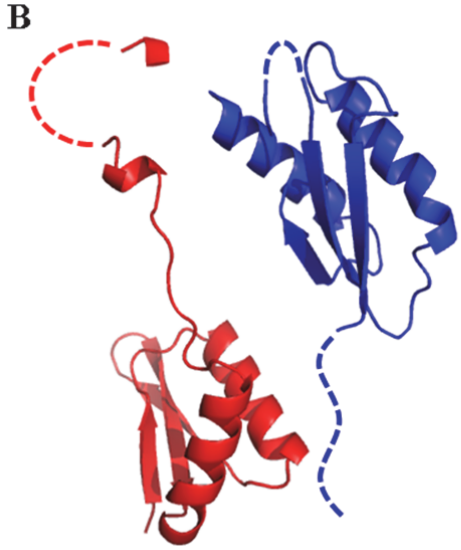


Figure 4.19 *Limited proteolysis of $MxiJ_{19-206}$.*

$MxiJ_{19-100}$ and $MxiJ_{106-204}$ fragments produced after limited proteolysis of $MxiJ_{19-206}$ with trypsin, as characterised by Harriet Allison (Part II Project).

A The sequence of $MxiJ_{19-206}$ numbered according to the native *S. flexneri* MxiJ sequence. Trypsin cleavage sites are shown by red triangles. Sequence corresponding to $MxiJ_{19-100}$ (red) and $MxiJ_{106-204}$ (blue), along with that lost during limited proteolysis (black) is indicated. The N-terminal His-tag is shown in italics.

B $MxiJ_{19-100}$ (red) and $MxiJ_{106-204}$ (blue) fragments mapped on to the SCWRL model of $MxiJ_{20-189}$ from **Figure 4.15**. Dashed lines indicate regions which are not structurally characterised for EscJ.

Trypsin was added to 5 mg of ^{15}N-MxiJ$_{19-206}$ at a w/w protein:trypsin ratio of 250:1, and the reaction left to proceed for 2 h at room temperature. The digest was subsequently injected on to a HiLoad 16/60 Superdex 75 pg column (GE Healthcare) and run through the matrix at 1 ml/min with 50 mM Tris-HCl (pH 7.5), 50 mM NaCl. The resulting elution profile showed two overlapping main peaks, which could be seen to correspond to MxiJ$_{19-100}$ and MxiJ$_{106-204}$ when the relevant fractions were analysed by SDS-PAGE (**Figure 4.20A**). When the indicated fractions were pooled and concentrated using an Amicon Ultra centrifugal device with a 3 kDa cut-off (Millipore), a MxiJ$_{19-100}$ sample was obtained with only a minor amount of

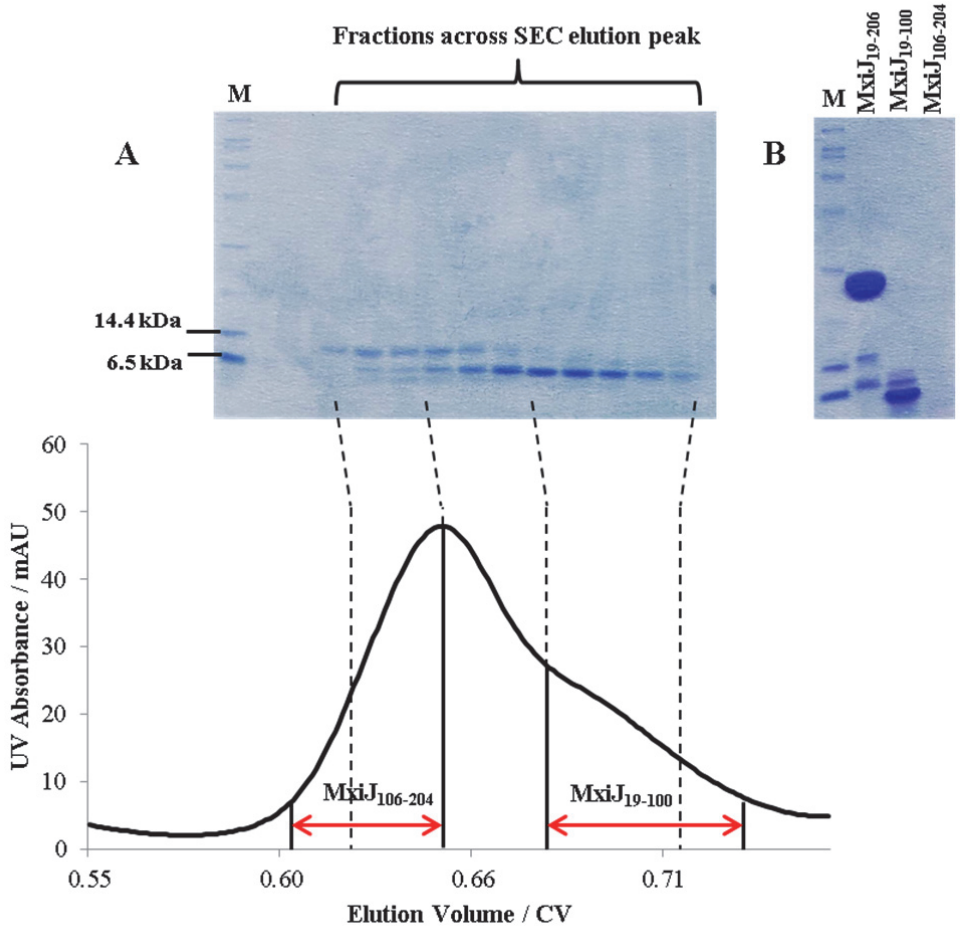


Figure 4.20 *Purification of MxiJ$_{19-100}$ and MxiJ$_{106-204}$.*

M: SDS-PAGE Molecular Weight Standards, Broad Range (Bio-Rad).

A Elution profile of the products from limited proteolysis of MxiJ$_{19-206}$ with trypsin from a HiLoad 16/60 Superdex 75 pg column. Elution with 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. The 15% SDS-PA gel shows samples taken from across the main elution peaks, with the dashed lines indicating the relative positions of their corresponding fractions in the elution profile. The solid lines indicate the fractions pooled to obtain MxiJ$_{206-204}$ and MxiJ$_{19-100}$ samples as indicated.

B 15% SDS-PA gel showing concentrated samples of MxiJ$_{19-206}$, MxiJ$_{19-100}$ and MxiJ$_{106-204}$ as indicated.

contaminating $MxiJ_{106-204}$ (**Figure 4.20B**). However, no $MxiJ_{106-204}$ was present after combining and concentrating the relevant fractions from the observed elution peak (**Figure 4.20B**), indicating the fragment had precipitated during the concentration step and was more unstable in solution than $MxiJ_{19-100}$.

A $^{1}H,^{15}N$-HSQC spectrum of 200 μM ^{15}N-$MxiJ_{19-100}$ supplemented with 5% D_2O showed good peak dispersion in both dimensions and clearly resolved peaks (**Figure 4.21A**). A small number of peaks arising from the backbone amides of small peptide fragments were present, presumably due to their formation as a by-product during limited proteolysis of $MxiJ_{19-206}$ with trypsin (**Figure 4.21A, dashed box**). However, 66 of the most intense peaks could still be unambiguously picked as candidates for the backbone amides of $MxiJ_{19-100}$. This left only 15 non-proline backbone amides unaccounted for, which could correspond to those sampling many different chemical environments within the unstructured flexible linker region (**Figure 4.16**)(Crepin *et al.*, 2005). Therefore $MxiJ_{19-100}$ appeared to be a good target for assignment and structure determination by NMR spectroscopy.

An overlay of the $^{1}H,^{15}N$-HSQC spectrum of ^{15}N-$MxiJ_{19-100}$ with that of ^{15}N-$MxiJ_{19-206}$ showed that the majority of peaks predicted to correspond to the backbone amides of the N-terminal subdomain were coincident with peaks corresponding to the full-length periplasmic domain (**Figure 4.21B**). This indicated that the chemical environment of the backbone amides of $MxiJ_{19-100}$ did not change between the sub-domain being present in isolation or within of the entire periplasmic domain. Therefore, determining the tertiary structure of $MxiJ_{19-100}$ would provide relevant information for characterising the structure of the entire periplasmic domain. In addition, this consistency of peak positions implied there was no significant interaction between the N- and C-terminal subdomains of MxiJ, in agreement with their elution as separate species during SEC (**Figure 4.19A**) and the lack of interdomain NOEs between the equivalent subdomains in EscJ (Crepin *et al.*, 2005). Therefore, the interaction between the N- and C-terminal domains of MxiJ *in vivo* requires further investigation.

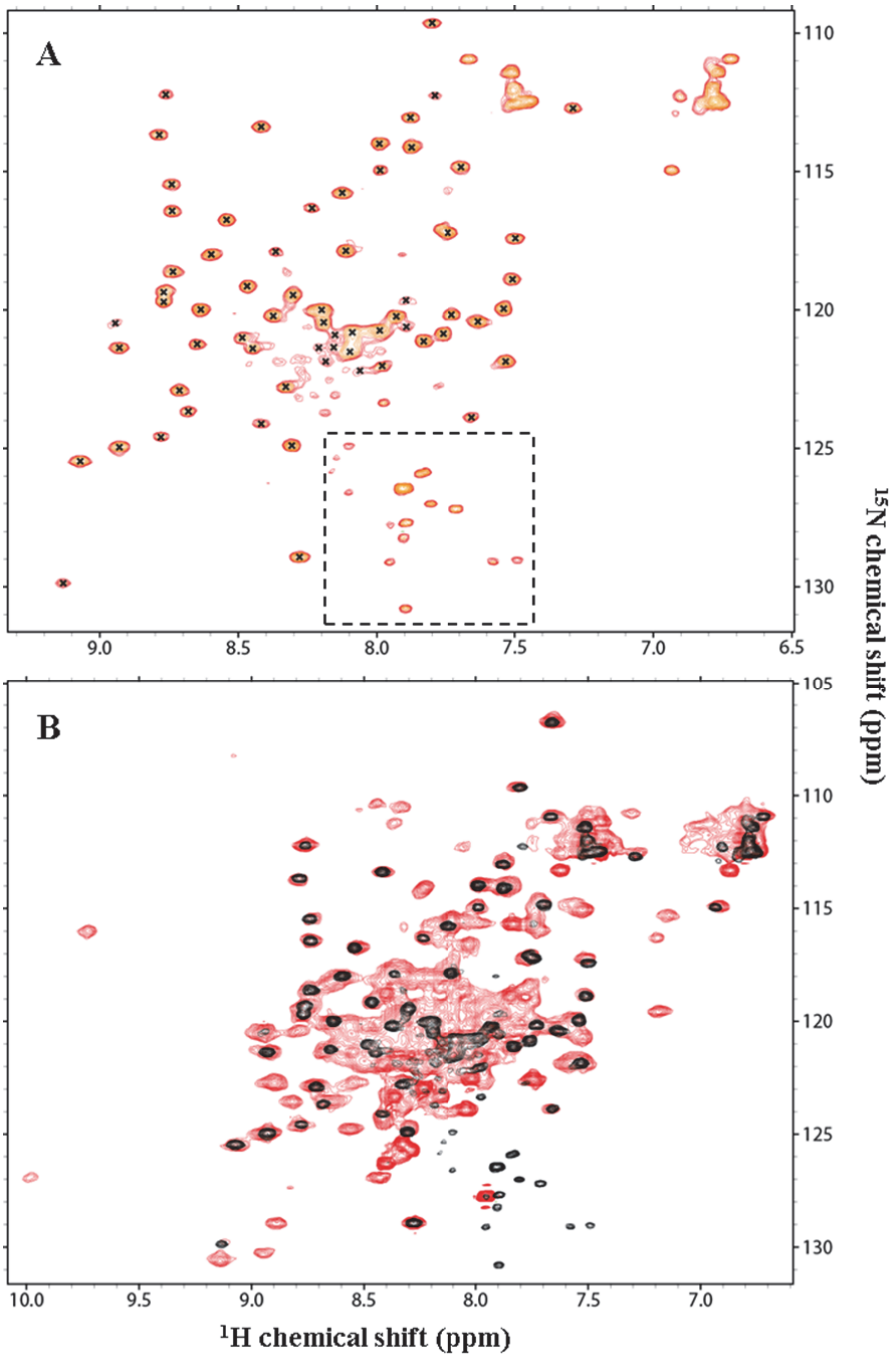


Figure 4.21 *^{1}H,^{15}N-HSQC spectrum of 200 µM ^{15}N-MxiJ$_{19-100}$.*

Spectrum acquired in 48 mM Tris-HCl (pH 7.5), 48 mM NaCl, 5% D_2O at 25°C.

A Peaks from the spectrum coloured according to peak height from lowest (red) to highest (yellow) values. Black crosses indicate the 66 peaks picked as those most likely to represent the backbone amides of MxiJ$_{19-100}$. The dashed box indicates the region where peaks corresponding to the backbone amides of small proteolytic fragments are found.

B Overlay of the ^{1}H,^{15}N-HSQC spectra of 150 µM ^{15}N-MxiJ$_{19-206}$ (red; **Figure 4.18**) and 200 µM ^{15}N-MxiJ$_{19-100}$ (black). A spectrum of 60 µM ^{15}N-MxiJ$_{19-206}$ collected in 48 mM Tris-HCl (pH 7.5), 48 mM NaCl, 5% D_2O subsequently showed the same peak positions as the spectrum shown here.

Furthermore, the $^{1}H,^{15}N$-HSQC spectrum overlay also implied that the 94 non-proline backbone amide peaks of $MxiJ_{106-204}$ were largely present within the central overlapped region of the $MxiJ_{19-206}$ spectrum (**Figure 4.21B**). This suggested the C-terminal subdomain was reasonably unstructured, in agreement with large loop regions observed in both the crystal and NMR structures of EscJ (Yip *et al.*, 2005, Crepin *et al.*, 2005). In combination with the observed instability of the isolated domain in solution (**Figure 4.20B**), this indicated that $MxiJ_{106-204}$ perhaps required an interaction with the other periplasmic components of the T3SS to stabilise the structure and therefore would not be a good target for structural characterisation by NMR spectroscopy.

4.3 Probing the Putative Interaction Between the Periplasmic Domains of MxiG and MxiJ

Although close proximity between PrgH and PrgK periplasmic domains is implied by *in vivo* cross-links between their membrane-proximal regions (Sanowar *et al.*, 2010) and by docking of atomic models into the 10 Å EM reconstruction of the *S. typhimurium* NC (Schraidt & Marlovits, 2011), a direct interaction between the periplasmic domains of MxiG and MxiJ homologues has never been observed *in vitro*. Furthermore, the scope of *in vivo* cross-linking for characterising their interaction interface at the atomic level is severely limited by the inaccessibility to solution of the majority of the MxiJ periplasmic domain within the assembled T3SS (Schraidt *et al.*, 2010), with accessible portions being within structurally uncharacterised regions (Sanowar *et al.*, 2010). Therefore NMR spectroscopy had the potential to characterise both the affinity and the interface of the putative interaction between the MxiG and MxiJ periplasmic domains.

All $MxiJ_{19-206}$ and MxiG-$C_{161-362/4}$ samples were dialysed into a buffer with a higher salt concentration (20 mM Tris-HCl (pH 7.5), 150 mM NaCl) using a Slide-A-Lyzer Dialysis

Cassette (Thermo Scientific), in an attempt to increase the protein stability at high concentrations. A $^{1}H,^{15}N$-HSQC spectrum obtained at 25°C for 50 μM ^{15}N-MxiJ$_{19-206}$ titrated with a 5-fold molar excess of MxiG-C$_{161-362/4}$ showed no changes in peak positions relative to the spectrum of 510 μM ^{15}N-MxiJ$_{19-206}$ alone (**Figure 4.22A**), indicating no interaction between the periplasmic domains was occurring in solution. As this was close to the maximum concentration of MxiG-C$_{161-362/4}$ that could be achieved given the solubility limit, the reverse titration was carried out in an attempt to increase the molar difference further. However a $^{1}H,^{15}N$-HSQC spectrum obtained at 20°C for 50 μM ^{15}N-MxiG-C$_{161-362/4}$ titrated with 1 mM MxiJ$_{19-206}$ was of poor quality (**Figure 4.22B**), due to the limited stability of the proteins at the concentrations required to achieve this 20-fold molar excess. However, the ^{15}N-MxiG-C$_{161-362/4}$ peaks that were present appeared to show no change in position upon addition of MxiJ$_{19-206}$. Taken together these spectra indicate that the periplasmic domains of MxiG and MxiJ are not interacting with an affinity that is amenable to study by NMR spectroscopy, so their characterisation by this technique was pursued no further.

EM visualisation of the dissolution of *S. typhimurium* NCs into intact IMRs and OMRs (Schraidt *et al.*, 2010, Schraidt & Marlovits, 2011) and of ring structures purified from *E. coli* cells recombinantly overexpressing PrgH and PrgK (Kimbrough & Miller, 2000), indicate that the concentric rings of PrgH and PrgK stably associate. Although equivalent observations have never been obtained for the *S. flexneri* IMR, its similarity to the *S. typhimurium* IMR in diameter and morphology from EM reconstructions (Hodgkinson *et al.*, 2009) suggests MxiG and MxiJ are also likely to be closely associated. This would appear to be in contrast to the NMR titrations presented here and in the previous chapter, which unambiguously showed that neither the cytoplasmic nor periplasmic domains of MxiG and MxiJ interact at 5-fold molar excess of each other. However, as was discussed for the self-association of MxiG-N in the previous chapter, low affinity interactions between MxiG and MxiJ could still be important for stabilising the IMR, where both proteins are effectively localised at high concentrations.

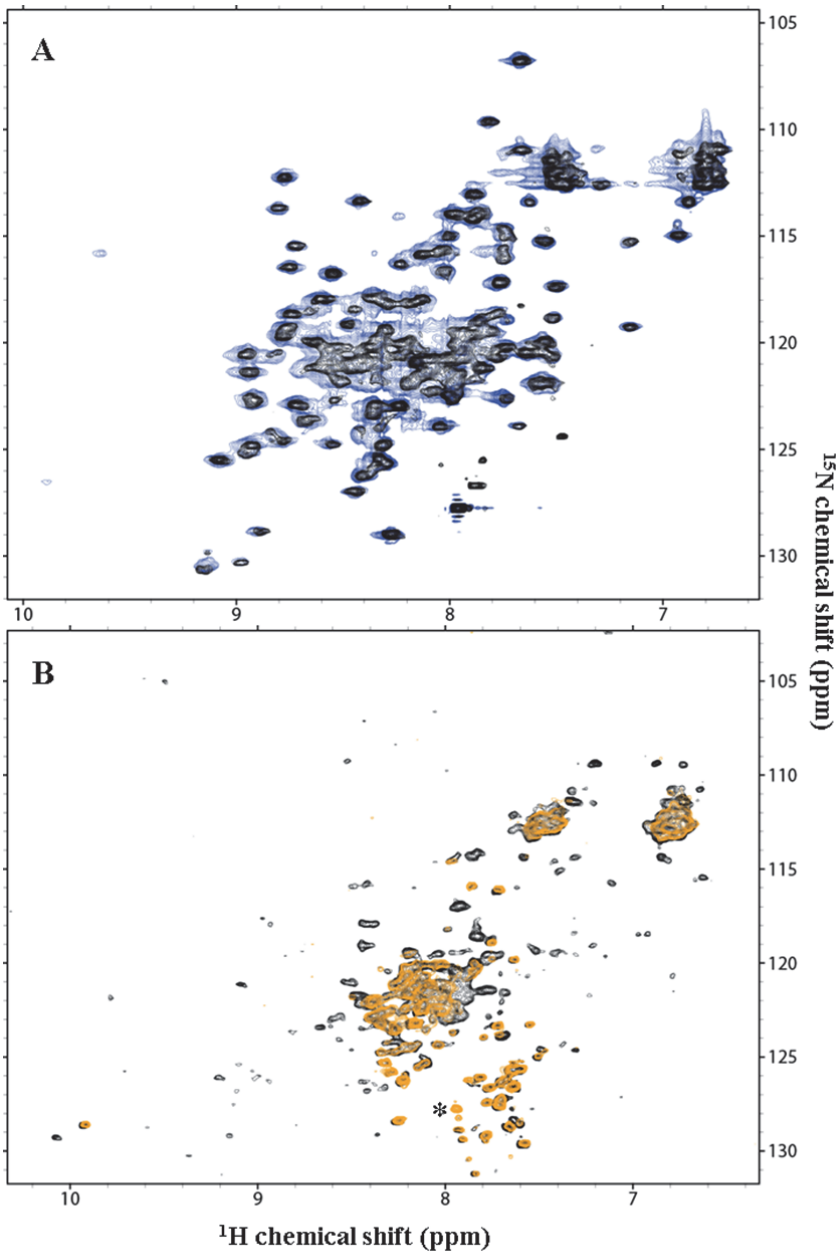


Figure 4.22 *NMR titrations with MxiJ$_{19-206}$ and MxiG-C$_{161-362/4}$.*

^{1}H,^{15}N-HSQC spectra collected with samples in 19 mM Tris-HCl (pH 7.5), 142 mM NaCl, 5% D_2O.

A Overlay of spectra from 510 µM ^{15}N-MxiJ$_{19-206}$ (blue) and from 50 µM ^{15}N-MxiJ$_{19-206}$ titrated with 245 µM MxiG-C$_{161-362/4}$ (black), both collected at 25^{o}C.

B Overlay of spectra from 50 µM ^{15}N- MxiG-C$_{161-362/4}$ (black) and of 50 µM ^{15}N- MxiG-C$_{161-362/4}$ titrated with 1 mM MxiJ$_{19-206}$ (orange), both collected at 20^{o}C. The asterisk shows a peak that has arisen from the natural abundance ^{1}H,^{15}N-HSQC spectrum of 1 mM MxiJ$_{19-206}$ (not shown).

During the mathematical modelling of the assembly of three-membered heteromeric ring structures, the presence of one weak interaction was found to be optimal for the most efficient assembly time and yield, as it enabled a balance between having the stable intermediates needed for assembly and having unstable intermediates that breakdown to replenish the required monomeric building blocks (Deeds *et al.*, 2012). In the context of the assembly of the basal body, a weak interaction between MxiG and MxiJ could therefore be beneficial, with the relatively uncharacterised interaction between MxiG and MxiD periplasmic domains providing the more stable interaction required for assembly. However this model would assume the simultaneous formation of all the ring structures of the basal body in a relatively stochastic manner. Alternatively, assembly of the basal body could occur in a step-wise manner, with assembly of the one ring structure being dependent on completed formation of the previous. Co-purification experiments using various knockout strains of *Y. enterocolitica* indicated that prior insertion of the OMR was required for the association of the IMR with the NC, with the YscD ring facilitating the assembly of the YscJ ring (Diepold *et al.*, 2010). This model could imply that the formation of a MxiJ ring requires a pre-formed MxiG ring to act as a structural scaffold and likewise that MxiG ring formation requires a pre-formed MxiD ring. This would suggest that MxiD is the only protein of the *S. flexneri* basal body that must undergo self-association at a high enough affinity to mediate independent ring formation, reconciling the absence of high affinity self-association of MxiG-N and MxiG-C domains found in this work and of MxiJ periplasmic domains observed previously (Harriet Allison, Part II Project). However, this model disagrees with the observed formation of PrgH and PrgK ring structures in isolation (Kimbrough & Miller, 2000), indicating the mechanism of basal body assembly may vary between bacterial species. As very little species-specific experimental data is available for the T3SS assembly mechanism in *S. flexneri*, it is difficult to speculate on the correct model for MxiJ and MxiG ring assembly.

Regardless, this work has shown that any interactions between the cytoplasmic and periplasmic domains of MxiJ and MxiG must be low affinity and therefore not readily amenable to

structural and biophysical characterisation *in vitro*. However an interaction between the transmembrane regions of MxiG and MxiJ remains uncharacterised and could still prove to be significant for IMR assembly and formation. Indeed, the self-association of the inner membrane protein OutC within the *Erwinia chrysanthemi* type II secretion system was found to be mediated by the single transmembrane segment and not the cytoplasmic or periplasmic globular domains (Login & Shevchik, 2006). Additionally, the cytoplasmic and periplasmic domains of MxiJ and MxiG have only been studied in isolation, therefore it is still possible that addition of low affinity interactions between the individual domains results in a significant overall interaction avidity between the full-length proteins. Preliminary characterisation of full-length MxiG and MxiJ hinted at their co-purification, however experiments were hampered by poor sample yields and stability (Matthew Cottee, Part II Project). In an attempt to improve the stability of MxiJ and MxiG *in vitro* and provide the best recapitulation of their *in vivo* association in the IMR, full-length MxiG and MxiJ were purified within the context of intact basal bodies, as presented in the next chapter.

4.4 Perspectives

A protocol was established for purifying large amounts of a homogenous sample of MxiG-$C_{142-371}$ from the insoluble fraction of *E. coli*. The MxiG-$C_{161-362/4}$ construct formed during limited proteolysis with trypsin appeared to be stable in solution, but subsequent crystallisation trials and NMR spectroscopy indicated that further construct optimisation would be required for structural characterisation. Similarly the N-terminal sub-domain appeared to be the only part of the MxiJ periplasmic domain that was amenable to structure determination. Investigation of the putative interactions of MxiG-C with both itself and the periplasmic domain of MxiJ indicated their affinity was too low for *in vitro* characterisation, relying on the high local concentration of MxiG and MxiJ within the IMR to be significant.

5. Purification and Characterisation of the T3SS Basal Body

The T3SS basal body of *S. flexneri* comprises the OMR secretin MxiD associated with the concentrically arranged IMR proteins MxiG and MxiJ. The 10 Å EM reconstruction of the *S. typhimurium* NC with C3 symmetry applied provides the highest resolution representation of basal body morphology, indicating 24-fold and 15-fold symmetry for the IMR and OMR respectively. However, the multiple interaction interfaces present within this hetero-oligomeric protein complex have only been characterised by docking high resolution structures of isolated periplasmic domains into this density. (Schraidt & Marlovits, 2011) Furthermore, no high resolution structures are currently available for any membrane-spanning region of the T3SS, so structural information is limited to the resolution of the EM reconstruction. For *S. flexneri*, available EM reconstructions of the NC are at a significantly lower resolution (Sani *et al.*, 2007, Hodgkinson *et al.*, 2009) and suggest a contradictory pseudo-24-fold and 12-fold symmetry for the IMR and OMR respectively after 12-fold symmetry is applied (Hodgkinson *et al.*, 2009). Therefore, there is a general need for structural characterisation of the T3SS basal body at atomic resolution, but particularly for the *S. flexneri* system. It was postulated that the basal body could be amenable to characterisation by X-ray crystallography, as the high symmetry of the ring structures may promote the formation of the regular lattice found within a crystal. In an attempt to provide a sample suitable for crystallisation trials, a protocol for the purification of the *S. flexneri* basal body was established, as outlined in this chapter.

5.1 Purification of the T3SS Basal Body

When intact NCs were purified from *S. flexneri* by density gradient centrifugation, 8% of the particles visualised by transmission EM (TEM) were found to be basal bodies lacking the needle appendage. Furthermore, a homogenous sample of basal bodies could be obtained from a *mxiH*$^{-}$ strain and was found to comprise MxiG, MxiJ and MxiD, but not the inner rod protein

MxiI. (Sani *et al.*, 2007) Similarly, basal bodies were also observed as a sub-population of NC preparations from *S. typhimurium* (Marlovits *et al.*, 2004). Therefore, these results indicate that the basal body is a stable sub-complex of the *S. flexneri* NC, which can be made exclusively by the bacterium in the absence of the needle protein.

In general, the homogeneity and concentration of a sample used for crystallisation trials must be much higher than one used for EM, therefore a stringent, efficient protocol was required for the purification of basal bodies produced by *S. flexneri* at native levels. It was found that inducible expression of full-length MxiG with a MGHHHHHHGSR N-terminal tag (His-MxiG) from the pBAD-*mxiG* construct in the *mxiG*$^-$ strain of *S. flexneri* (**Table 3.2**) led to the formation of a functional T3SS, indicating His-MxiG can be inserted in the inner membrane and substituted for natively expressed MxiG in the IMR (Zenk *et al.*, 2007). Furthermore, purification of NCs under near-physiological conditions via Ni-affinity chromatography and SEC produced a highly enriched sample when visualised by TEM (Zenk *et al.*, 2007, Hodgkinson *et al.*, 2009). A sub-population of structures lacking the needle were also observed, indicating basal bodies incorporating His-MxiG were stable (Hodgkinson *et al.*, 2009). Therefore the approaches of Sani et al., 2007 and Zenk et al., 2007 were combined in a protocol for the purification of His-tagged basal bodies from a MxiH deficient strain of *S. flexneri*, as outlined in this section.

A *mxiG*$^-$*/mxiH*$^-$ strain of *S. flexneri* (donated by A. Blocker) was transformed with pBAD-*mxiG* (Zenk *et al.*, 2007) by the 'gentle transformation' protocol (**Appendix 8.2.1**) and propagated overnight at 37^{o}C on TCSB agar plates (14 mg/ml agar, 30 mg/ml TCSB). The resulting *mxiG*$^-$*/mxiH*$^-$ pBAD-*mxiG* colonies were used to inoculate 4 x 10 ml TCSB and the resulting cultures grown overnight at 37^{o}C. These overnight cultures were subsequently used to inoculate 6 x 1 l growth medium (**Table 5.1**). Cells were grown in a shaking incubator at 37^{o}C until an OD_{600} of ~1.0 was reached, at which point they were harvested and washed once with PBS. 50 μg/ml kanamycin and 100 μg/ml ampicillin were present throughout all cell growth.

Table 5.1 Growth medium used for *mxiG⁻/mxiH⁻* pBAD-*mxiG*

Component	Concentration	Justification
TCSB without Dextrose (Fluka)	27.5 g/l	Dextrose present within standard TCSB medium represses His-MxiG expression from ara operon.
Glycerol	0.4%	Carbon source compensating for lack of dextrose.
Arabinose	0.02%	Continued induced low-level expression of His-MxiG. Concentration selected as: • produces expression of His-MxiG in *mxiG⁻* comparable to MxiG expression in wild-type *S. flexneri* (Zenk *et al.*, 2007); • highest concentration that could be present before toxicity during *mxiG⁻/mxiH⁻* pBAD-*mxiG* growth was apparent (Matthew Cottee, Part II Project).
Kanamycin	50 μg/ml	Maintenance of *mxiG⁻/mxiH⁻*
Ampicillin	100 μg/ml	Maintenance of pBAD-*mxiG*

Basal bodies were extracted from *mxiG⁻/mxiH⁻* pBAD-*mxiG* immediately after harvesting and purified largely following the protocol of Zenk et al., 2007, as outlined in detail in **Table 5.2**. Briefly, cells were spheroplasted to remove the outer cell wall, before solublisation of the inner and outer membranes with the non-ionic detergent Triton X-100 (Sigma), which was present at a 130-fold excess of its critical micelle concentration (CMC) of 0.015%. This allowed incorporation of basal bodies into Triton X-100 micelles. A low-speed centrifugation step allowed the removal of cell debris, whilst a high-speed centrifugation step pelleted the basal body-containing micelles. His-tagged basal bodies were then selectively bound overnight to Ni-NTA Agarose (Qiagen) and contaminating proteins were subsequently removed by washing the

Table 5.2 Protocol for the purification of basal bodies from *mxiG⁻/mxiH⁻* pBAD-*mxiG*

Day 1

1. Resuspension of *mxiG⁻/mxiH⁻* pBAD-*mxiG* cells in 30 ml sucrose solution (0.5 M sucrose, 100 mM Tris-HCl (pH 8.0)) and division into 2 x 15 ml aliquots.
2. 1 mM EDTA and 1 mg/ml lysozyme added.
3. Spheroplasting of cells for 45 min at 37°C in shaking incubator.
4. EDTA-free protease inhibitor and 2% Triton X-100 added.
5. Solubilisation of membranes for 30 min at 4°C with continual mixing.
6. 0.05 mg/ml DNase I and 10 mM $MgSO_4$ added.
7. Centrifugation of lysate at 18,000 x g for 20 min at 4°C to pellet cell debris.
8. Centrifugation of supernatant at 120,000 x g for 120 min at 10°C to pellet the membranes.
9. Resuspension of each pellet in 5 ml membrane resuspension buffer (10 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM imidazole (pH 8.0), 0.5% Triton X-100, EDTA-free protease inhibitor) by drawing through a needle attached to a syringe.
10. Membrane resuspension mixed with 1 ml Ni-NTA Agarose (pre-equilibrated in membrane resuspension buffer) and total volume increased to 10 ml with membrane resuspension buffer.
11. Batch binding of His-tagged basal bodies to Ni-NTA Agarose overnight at 4°C with continual mixing.

Day 2

1. Membrane resuspension and Ni-NTA Agarose applied to PD-10 gravity-flow column (GE Healthcare).
2. Washing of Ni-NTA Agarose with 30 ml wash buffer (10 mM Tris (pH 8.0), 150 mM NaCl, 50 mM imidazole, EDTA-free protease inhibitor) with 0.1% lauroylsarcosin added to remove contaminating outer membrane proteins.
3. Washing of Ni-NTA Agarose with 30 ml wash buffer with 0.1% Triton X-100 added.
4. Elution of His-tagged basal bodies with elution buffer (10 mM Tris (pH 8.0), 500 mM imidazole, 0.1% Triton X-100), collecting 5 x 1 ml fractions. 1 mM EDTA added to each fraction.
5. Fractions combined and concentrated to 1 ml using an Amicon Ultra centrifugal device with a 100 kDa cut-off (Millipore).
6. Concentrated sample immediately injected on to a Superdex 200 10/300 GL column (GE Healthcare) and flowed at 1 ml/min with SEC buffer (25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% Triton X-100).

slurry with buffers containing lauroylsarcosin and Triton X-100. The sample eluted from the Ni-NTA Agarose was then subjected to SEC to separate intact basal bodies from accummulated empty Triton X-100 micelles and micelles containing His-MxiG not associated with basal bodies.

Samples taken throughout the purification of His-tagged basal bodies were separated on a 15% SDS-PA gel and analysed by western blotting (**Appendix 8.2.7**) against His-MxiG and MxiJ, using a 1:10 dilution of mouse α-MxiG and α-MxiJ primary antibodies (donated by A. Blocker), a 1:1500 dilution of the peroxidase-conjugated goat anti-mouse secondary antibody (Thermo scientific) and a 15 s exposure to the Amersham Hyperfilm ECL (GE Healthcare). The α-MxiG blot showed that full-length His-MxiG was predominantly present in the pellet after the 120,000 x g spin and therefore had been effectively solubilised from the inner membrane by Triton X-100 (**Figure 5.1A, lane 5**). Although a large amount of His-MxiG had not bound to the Ni-NTA agarose after batch binding (**Figure 5.1A, lane 6**), a significant proportion was still specifically eluted with 500 mM imidazole (**Figure 5.1A, lane 9**). A high background was generally detected by α-MxiG, although two degradation products were particularly prominent (**Figure 5.1A, asterisks**). As α-MxiG was epitope mapped to MxiG-N (**Figure 5.1A, lane 13**), these His-MxiG degradation products are likely comprised of at least MxiG-N and the transmembrane region (Matthew Cottee, Part II Project), providing an explanation for their affinity purification from the membrane fraction of *mxiG*⁻*/mxiH*⁻ pBAD-*mxiG*. The α-MxiJ blot had a much cleaner background and showed full-length MxiJ to have a similar profile to His-MxiG in the purification samples. The specific elution of native MxiJ from the Ni-NTA agarose suggested it had been purified as a complex from the inner membrane with His-MxiG (**Figure 5.1A, lane 9**).

His-tagged basal bodies could be expected to have a molecular weight of 2.6 MDa, based on the molecular weights of His-MxiG (44.3 kDa), lipidated MxiJ (27 kDa) and MxiD (60.7 kDa) and their proposed ring symmetries of 24-fold, 24-fold and 15-fold respectively (Schraidt & Marlovits, 2011). Therefore they would be expected to be eluted in the void of a Superdex 200

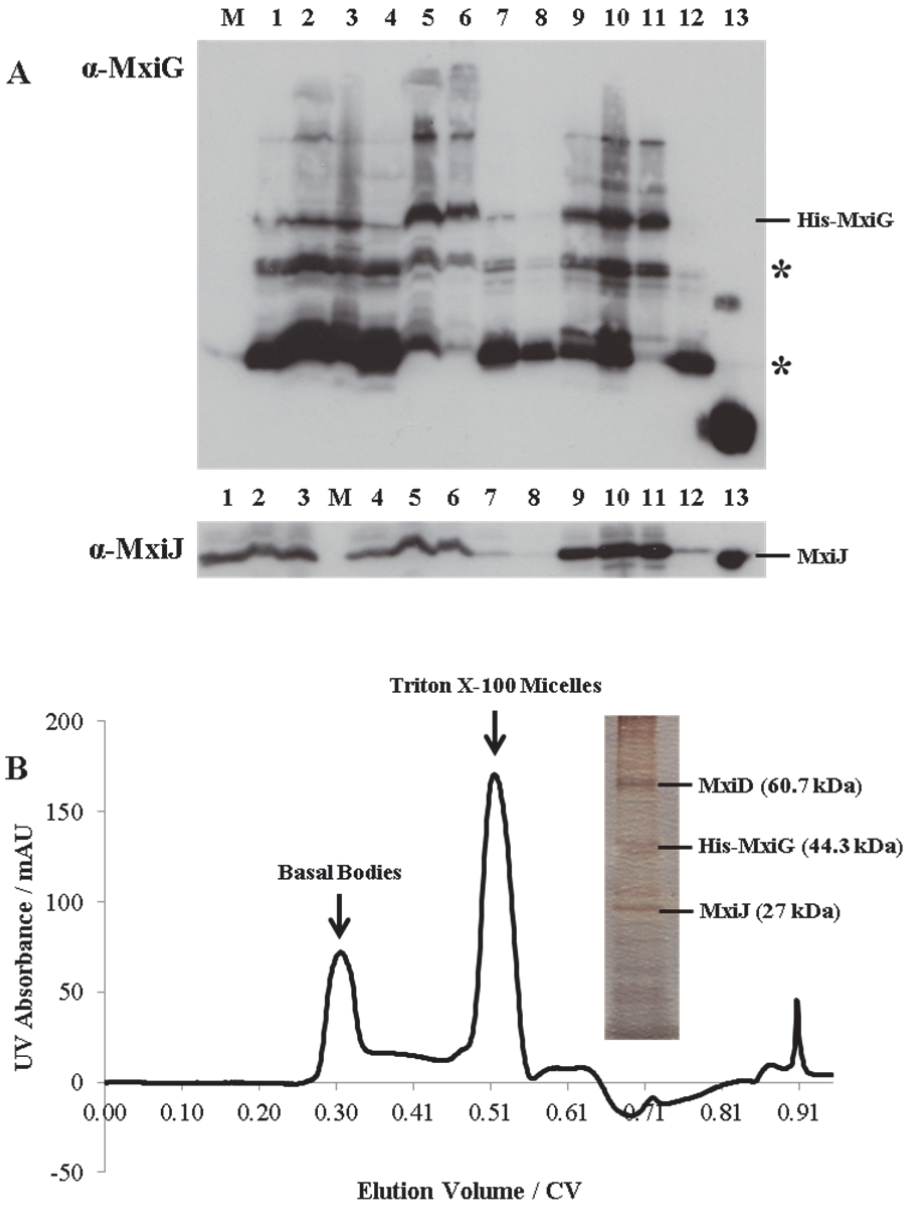


Figure 5.1 *Purification of His-tagged basal bodies mxiG⁻/mxiH⁻ pBAD-mxiG.*

A Western blots of samples taken throughout the purification using mouse α-MxiG and α-MxiJ primary antibodies. M: PageRuler Plus Prestained Protein Ladder (Fermentas), 1: whole cell lysate, 2: supernatant after 18,000 x g spin, 3: pellet after 18,000 x g spin, 4: supernatant after 120,000 x g spin, 5: pellet after 120,000 x g spin, 6: flow-through after batch-binding to Ni-NTA agarose, 7: flow-through from Ni-NTA agarose wash with 0.1% lauroylsarcosin, 8: flow-through from Ni-NTA agarose wash with 0.1% Triton X-100, 9: elution from Ni-NTA agarose, 10: concentrated fractions from Ni-NTA agarose elution, 11: void fraction from SEC, 12: Triton X-100 micelle fraction from SEC and 13: positive control for epitope mapping. The positive control for α-MxiG was purified MxiG-$N_{6\text{-}112}$ and for α-MxiJ was purified $MxiJ_{19\text{-}206}$. The positions of MxiJ and His-MxiG are indicated, whilst prominent degradation products of His-MxiG are shown by asterisks.

B Elution profile of His-tagged basal bodies from a Superdex 200 10/300 GL column. Elution with 25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% Triton X-100. The void fraction from SEC was run on 15% SDS-PA gel that was silver-stained using SilverQuest Staining kit (Invitrogen), as shown in the inset. The bands identified to be the basal body components by mass spectrometry are indicated with their relative molecular weights.

10/300 GL column, as previously observed for whole NCs (Zenk *et al.*, 2007). Indeed, the elution profile showed a significant void peak (**Figure 5.1B**), which was subsequently found to contain both His-MxiG and MxiJ when subjected to western blotting (**Figure 5.1A, lane 11**). Furthermore, analysis of the void fraction on a silver-stained 15% SDS-PA gel showed three bands, which were shown to be His-MxiG (mascot score = 436), MxiJ (mascot score = 312) and MxiD (mascot score = 296) after trypsin digestion and mass spectrometry identification (Central Proteomics Facility, University of Oxford), indicating all of the expected components of the basal body were co-purified (**Figure 5.1B, inset**).

The elution profile of the His-tagged basal body also showed a prominent peak in the position expected for the elution of Triton X-100 micelles, indicating their effective separation from basal body-containing micelles (**Figure 5.1B**). However, unlike the elution peak obtained for empty Triton X-100 micelles after an unsuccessful purification attempt (**Figure 5.2**), the elution peak had a positive absorbance, suggesting the Triton X-100 micelles contained protein. Indeed, western blotting of a fraction eluted within this peak with α-MxiG indicated that the smallest degradation product of His-MxiG was present (**Figure 5.1, lane 12**). Therefore, SEC had also allowed the effective separation of His-tagged basal bodies from this N-terminal degradation product, which had been unable to associate with the basal body in the absence of a complete periplasmic domain.

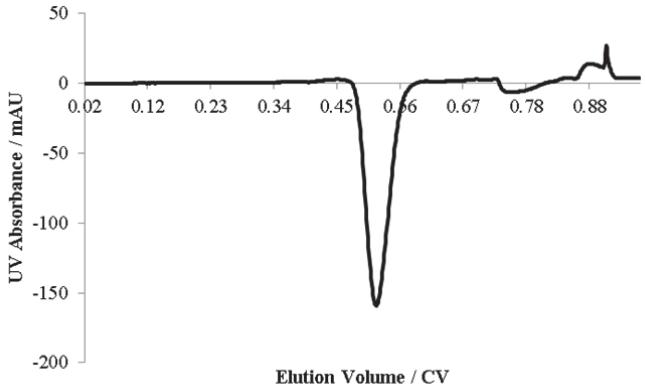


Figure 5.2 *Elution profile for empty Triton X-100 micelles*.
Obtained from a Superdex 200 10/300 GL column with elution in 25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% Triton X-100.

To ascertain whether intact His-tagged basal bodies had been purified, TEM was carried out on the void fraction containing His-MxiG, MxiJ and MxiD by Michael Shaw (Sir William Dunn School of Pathology Bioimaging Facility, University of Oxford). A formvar-carbon coated

glow-discharged copper grid (300 mesh) was coated with 1 µl undiluted void fraction, washed once with distilled water and negatively stained with 2% uranyl acetate (pH 4.0), with filter paper being used to remove excess liquid at each step. Images were then acquired using a Philips Tecnai T12 TEM, with micrographs being recorded at 30,000 x magnification on a 2K ProScan CCD camera.

The images showed dispersed circular structures with a diameter of ~250 Å, equivalent to that of the IMR from *S. flexneri* EM reconstructions (Sani *et al.*, 2007, Hodgkinson *et al.*, 2009)(**Figure 5.3A**). Further inspection revealed these structures to have a smaller ring sitting on top, which has a size consistent with an OMR of 137 Å diameter (Hodgkinson *et al.*, 2009)(**Figure 5.3A, inset**). These images were very similar to those obtained of basal bodies purified from the *mxiH*⁻ strain of *S. flexneri* (Sani *et al.*, 2007)(**Figure 5.3B**), indicating the visualised structures were top views of intact His-tagged basal bodies. Unlike the mixture of top views and side views obtained for the untagged basal bodies (Sani *et al.*, 2007), there were no side-views seen for His-tagged basal bodies. This is in agreement with top view only orientations obtained both for the basal body sub-population present after negative staining of His-MxiG containing NCs with 2% uranyl acetate (pH 4.0)(Hodgkinson *et al.*, 2009) and for His-PrgH containing NCs purified from *S. typhimurium* under acidic conditions (Schraidt & Marlovits, 2011). The observed uniform top views of the basal bodies were hence likely due to an attraction between the positively-charged His-tags, which were protonated during negative staining at pH 4.0, and the negatively-charged glow-discharged grids, which tethered the N-terminus of His-MxiG to the grid. Therefore, the TEM images indicated a protocol for the purification of intact His-tagged basal bodies had been established, providing a sample for further characterisation.

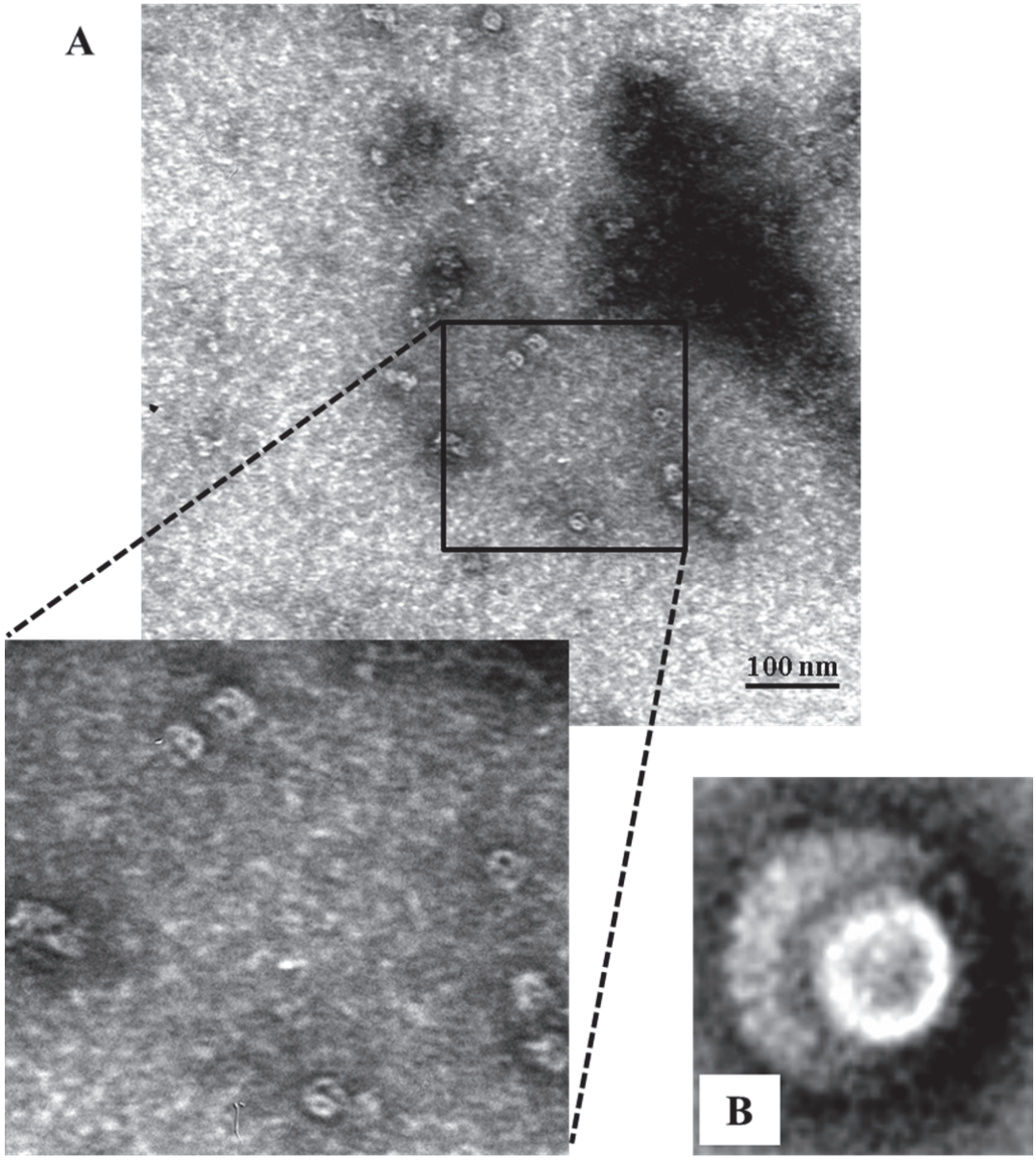


Figure 5.3 *TEM images of the S. flexneri basal body*.

A TEM image showing His-tagged basal bodies acquired at 30,000 x magnification. The void fraction after SEC was applied to a formvar-carbon coated copper grid and negatively stained with 2% uranyl acetate (pH 4.0), before imaging with a Philips Tecnai T12 TEM. The zoomed-in view given in the inset shows a cluster of basal bodies sitting on their base.

B TEM image of untagged basal bodies acquired at 66,850 x magnification taken from Sani et al., 2007. Basal bodies purified from a *mxiH*⁻ strain of *S. flexneri* were used to prepare grids via the same protocol used in this study. This view of a basal body sitting on its base is equivalent to the orientation observed for His-tagged basal bodies in **A**.

Whilst EM images can be acquired rapidly after sample preparation, protein crystallisation can take from hours to months, hence requiring the sample to be much more stable in solution. This implies that purified His-MxiG, MxiJ and MxiD must remain stably associated within the basal body over the timescales required for crystallisation, in order for the complex to be amenable to structural characterisation by X-ray crystallography. The stability of the purified His-tagged basal bodies was tested by Dr Steven Johson using MALS. Fractions taken across the void peak from SEC were immediately concentrated 10-fold and injected on to a Superose 6 10/300 GL column (GE Healthcare). The in-line multi-angle scattering across the resulting void peak showed a continuous spectrum of molecular masses, indicating a highly heterogeneous mixture of macromolecular complexes that likely correspond to various composite species of the basal bodies (**Figure 5.4**). Therefore His-tagged basal bodies present within Triton X-100 micelles do not appear to be stable over a period of a few hours and do not provide an ideal sample for protein crystallisation.

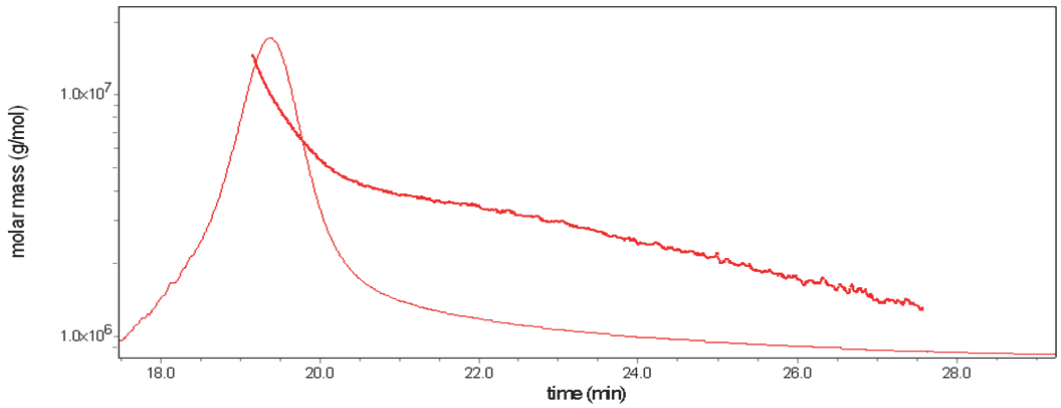


Figure 5.4 *Analysis of the mass of His-tagged basal bodies by SEC and in-line multi-angle light scattering.*

The left axis represents the molecular mass at any given point in the elution profile. The elution profile is shown for the void elution peak from a concentrated sample of purified His-tagged basal bodies run down a Superose 6 10/300 GL in 25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.05% Triton X-100.

Although an accurate measurement of basal body concentration was unfeasible given the complexity of the system, the structures were highly dispersed in EM images (**Figure 5.3A**). This concentration of basal bodies was deemed insufficient for crystallisation trials, especially given that rapid nucleation events would be required in such an unstable sample. This yield is in

stark contrast to the EM image obtained from the unconcentrated void fraction of purified His-tagged NCs, which shows a high density of structures on the EM grid (Zenk *et al.*, 2007). Similarly, the purification of native basal bodies in Triton X-100 from *mxiH*⁻ led to a much lower number of particles per image compared to when intact NCs were purified from wild-type *S. flexneri* by an identical protocol (Sani *et al.*, 2007). Together these results imply either that the basal body is intrinsically unstable in the absence of the inner rod and needle or that the optimal protocol for basal body purification differs from that for NC purification. To ascertain whether the latter is the case, a series of modifications were made to the detergents used in the His-tagged basal body purification protocol shown in **Table 5.2**, as outlined in the next section.

5.2 Optimisation of the Detergents used for T3SS Basal Body Purification

The nature of the detergent used to solubilise and purify membrane proteins is a key determinant in the yield and stability of the final sample, along with its suitability for protein crystallisation (Arnold & Linke, 2001). Previous optimisation of the protocol for purification of His-tagged NCs from *S. flexneri* tested the detergents Triton X-100, n-octyl β-D-glucopyranoside (OG), 3-lauroamido-N,N'-dimethyl-3-n-propylamine (LAPAO) and n-dodecyl-N,N-dimethylamine-N-oxide (DDAO), ascertaining that Triton X-100 gave the highest sample yield, stability and purity (Zenk *et al.*, 2007). Although Triton X-100 enabled the successful purification of His-tagged basal bodies (**Figure 5.3A**), the resulting sample yield and stability was not optimal for crystallisation trials, suggesting further detergents should be tested.

n-dodecyl-N,N-dimethylamine-N-oxide (LDAO) has been the detergent of choice for purification and EM visualisation of NCs from *S. typhimurium* (Marlovits *et al.*, 2004, Schraidt *et al.*, 2010, Schraidt & Marlovits, 2011). Therefore His-tagged basal bodies were purified from *mxiG*⁻*/mxiH*⁻ pBAD-*mxiG* as outlined in **Table 5.2**, except 2% LDAO (Anatrace) was used to solubilise membranes and 0.1% LDAO replaced Triton X-100 in all subsequent purification buffers. However when samples taken throughout the purification were separated by SDS-

PAGE and analysed by western blotting with α-MxiJ as previously described, solubilisation of membrane-associated MxiJ complexes with LDAO was seen to be less efficient than with Triton X-100, resulting in no significant elution of His-tagged basal bodies in the void fraction during SEC (**Figure 5.5**). Therefore, either the absence of the inner rod and needle or the composition of the *S. flexneri* membranes resulted in unsuccessful purification of His-tagged basal bodies with LDAO.

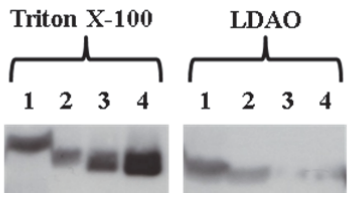


Figure 5.5 *Purification of His-tagged basal bodies with LDAO.*

Western blot of samples taken throughout the purification using mouse α-MxiJ primary antibody. 1: pellet after 120,000 x g spin, 2: flow-through after batch-binding to Ni-NTA agarose, 3: elution from Ni-NTA agarose and 4: void fraction from SEC.

The hydrophilic polyethylene oxide chain of Triton X-100 is variable in length, making it a non-optimal detergent for obtaining a homogenous sample for crystallisation trials. Detergents with a defined headgroup type and size are therefore preferred, with the most membrane protein crystal structures being obtained using n-dodecyl-β-D-maltopyranoside (DDM)(Privé, 2007). Therefore initial small-scale purifications of His-tagged basal bodies each from 1.5 l *mxiG*$^{-}$*/mxiH*$^{-}$ pBAD-*mxiG* were carried out, using different DDM (Anatrace) concentrations to solubilise the membranes and 0.03% DDM in all subsequent buffers. A western blot with α-MxiJ indicated that an equivalent amount of MxiJ was present in the void fractions from SEC (**Figure 5.6A**), indicating solubilisation had been equally effective with all concentrations of DDM and 2% Triton X-100. Therefore, 0.25% DDM was selected for further membrane solubilisation to minimise the amount of excess detergent present in the purified sample. This led to successful solubilisation of MxiJ from the membrane during a large-scale purification of His-tagged basal bodies from 6 l *mxiG*$^{-}$*/mxiH*$^{-}$ pBAD-*mxiG* (**Figure 5.6B, lane 2**). However, only a small amount of MxiJ was present in the resulting void fraction from SEC, visualised after a prolonged 75 s exposure to the Amersham Hyperfilm ECL during western blotting (**Figure 5.6B, lane 8**). Furthermore, SEC elution profiles for the affinity-purified basal bodies obtained an hour apart showed a decrease in the intensity of the basal body-containing void

peak and an increase in the intensity of peaks corresponding to smaller molecular weight complexes (**Figure 5.6C**). Together these results indicated that His-tagged basal bodies were highly unstable once present within DDM micelles and did not remain intact over the timescales required for crystallography.

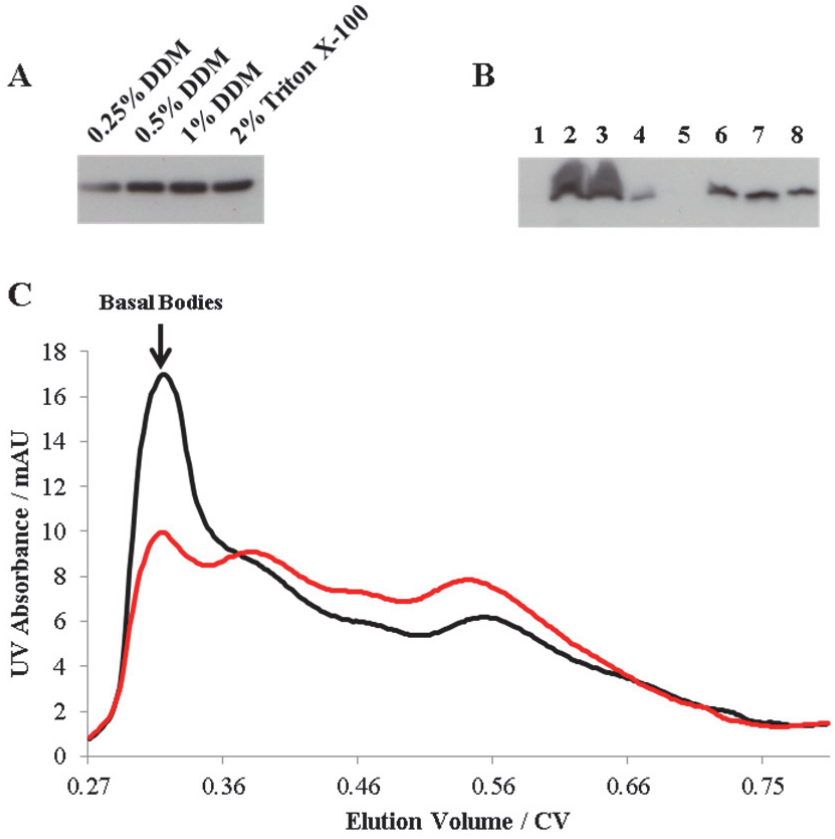


Figure 5.6 *Purification of His-tagged basal bodies with DDM.*

A Western blot of the void fractions from SEC using mouse α-MxiJ primary antibody. The detergent type and concentration used for membrane solubilisation during a small-scale purification are indicated.

B Western blot of samples taken throughout a large-scale purification using mouse α-MxiJ primary antibody. 1: supernatant after 120,000 x g spin, 2: pellet after 120,000 x g spin, 3: flow-through after batch-binding to Ni-NTA agarose, 4: flow-through from Ni-NTA agarose wash with 0.1% lauroylsarcosin, 5: flow-through from Ni-NTA agarose wash with 0.03% DDM, 6: elution from Ni-NTA agarose, 7: concentrated fractions from Ni-NTA agarose elution and 8: void fraction from SEC.

C Elution profile of His-tagged basal bodies from a Superdex 200 10/300 GL column. Concentrated fractions from the Ni-NTA agarose elution were divided into two equal aliquots and subjected either immediately (black) or after 1 h (red) to SEC. Elution with 25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.03% DDM. The void peak corresponding to elution of His-tagged basal bodies is indicated.

In an attempt to stabilise the His-tagged basal bodies within detergent micelles, a milder detergent than DDM was tested; Cymal-4 retains the same hydrophilic headgroup as DDM, but has a cyclohexyl substitution at position four of the alkyl chain, resulting in a CMC that is ~40-fold higher. However, membrane solubilisation of MxiJ with 1% Cymal-4 (Anatrace) was slightly less efficient than with 2% Triton X-100 (**Figure 5.7, lane 1**), indicating harsher detergents such as DDM and Triton X-100 were required for the most efficient extraction of His-tagged basal bodies from the IMR and OMR. Nonetheless, His-tagged basal bodies were subsequently purified from the Cymal-4 solubilised membranes by using 0.8% Cymal-4 in all buffers. In addition, the Triton X-100 solubilised basal bodies were affinity purified and detergent exchanged by including 0.8% Cymal-4 during the second wash of the Ni-NTA agarose, resulting in no loss of MxiJ in the flow-through (**Figure 5.7, lane 4**). Again, only a small amount of MxiJ was present in the void fractions after SEC of both samples (**Figure 5.7, lane 6**), indicating that His-tagged basal bodies were not more stable within Cymal-4 micelles.

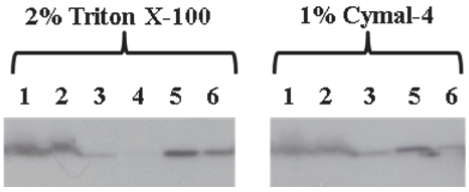


Figure 5.7 *Purification of His-tagged basal bodies with Cymal-4*.

Western blot of samples taken throughout the purification using mouse α-MxiJ primary antibody. 1: pellet after 120,000 x g spin, 2: flow-through after batch-binding to Ni-NTA agarose, 3: flow-through from Ni-NTA agarose wash with 0.1% lauroylsarcosin, 4: flow-through from Ni-NTA agarose wash with 0.8% Cymal-4, 5: elution from Ni-NTA agarose, 6: void fraction from SEC. The detergent type and concentration used for membrane solubilisation is shown.

Therefore, purification of His-tagged basal bodies using Triton X-100, LDAO, DDM and Cymal-4 did not produce a sample with either the yield or stability required for crystallisation trials. This is in contrast to His-tagged NCs, which could be purified in large numbers with OG, LAPAO, DDAO and Triton X-100 using an equivalent protocol (Zenk *et al.*, 2007). This suggests that intact NCs are more amenable to purification than basal bodies, in agreement with the lower yield of untagged basal bodies purified from *mxiH*⁻ relative to NCs purified from wild-type *S. flexneri* using the same protocol (Sani *et al.*, 2007). The implication is that NCs

form a more robust macromolecular complex, likely mediated by the inner rod and needle that are absent from basal bodies. This is hardly surprising given that difference EM images of the NC reconstruction minus the basal body reconstruction for both *S. flexneri* (Sani *et al.*, 2007) and *S. typhimurium* (Marlovits *et al.*, 2004) show that the inner rod and needle are integral to the basal body structure. Indeed, EM analysis of complexes isolated from a *invG*$^{-}$ strain of a *S. typhimurium* shows a needle-like structure extending from the IMR (Schraidt *et al.*, 2010), implying the inner rod and needle are anchored firmly to the IMR via the previously visualised socket structure (Marlovits *et al.*, 2004, Marlovits *et al.*, 2006). Furthermore, a sub-population of complexes isolated from wild-type *S. flexneri* lacked the IMR, which when classified showed a variable position of the OMR relative to the needle, indicating a weak interaction between them (Sani *et al.*, 2007). Therefore the stability of the isolated NC relative to the basal body is likely to be largely dependent on the interaction between the IMR, inner rod and needle, which locks the IMR and OMR into their positions within the basal body.

5.3 Purification of Complexes Lacking MxiG-N

To ascertain whether NCs can form in the absence of MxiG-N, residues F12-H104 were deleted from pBAD-*mxiG* by site-directed mutagenesis, as outlined in Chapter 3 for the construction of pAB13-ΔN. The resulting pBAD-*mxiGΔN* construct afforded the expression of MxiGΔN with a MGHHHHHHGSR N-terminal tag (His-MxiGΔN), allowing affinity purification of any complexes containing the construct in an analogous manner to the purification of His-tagged basal bodies. His-MxiGΔN was expressed from pBAD-*mxiGΔN* in the *mxiG*$^{-}$ strain of *S. flexneri* (**Table 3.2**) as described previously for the expression of His-MxiG in *mxiG*$^{-}$*/mxiH*$^{-}$, then the resulting His-tagged complexes were purified as outlined in **Table 5.2**.

Although His-MxiGΔN lacks the epitope for α-MxiG, western blotting of samples taken throughout the purification with α-MxiJ showed that MxiJ was affinity purified from the membrane fraction of *mxiG*$^{-}$ (**Figure 5.8A, lane 8**). Furthermore, the small void peak from SEC

(**Figure 5.8B**) was found to contain MxiJ (**Figure 5.8A, lane 10**), indicating purification of His-MxiGΔN and MxiJ within a macromolecular complex. These void fractions were subsequently concentrated ~7-fold using a Vivaspin 6 centrifugal device with a 1 MDa cut-off (Sartorius) to avoid accumulation of empty Triton X-100 micelles. Analysis of this concentrated sample on a silver-stained SDS-PA gel showed three main bands (**Figure 5.8B, inset**); the lower molecular weight bands were in the expected positions for MxiJ and His-MxiGΔN, whilst the highest molecular weight band would correspond to MxiD if the protein was running at a slightly higher molecular weight than expected, as observed within the context of the His-tagged basal bodies (**Figure 5.1B, inset**). Therefore it appeared that basal body components had co-purified with His-MxiGΔN.

To ascertain whether the purified His-MxiGΔN complexes were forming ordered structures, TEM was carried out on the unconcentrated and concentrated void fractions (stored at -80^{o}C in 10% trehalose), as described previously for the His-tagged basal bodies. The images for both samples showed clusters of circular structures with a ~250 Å diameter (**Figure 5.8C, asterisk**), consistent with the diameter of the IMR from *S. flexneri* EM reconstructions (Sani *et al.*, 2007, Hodgkinson *et al.*, 2009). However, given the poor background contrast in these images, it was difficult to unambiguously discern whether smaller ring structures corresponding to the OMR were present, as had been observed for the His-tagged basal bodies (**Figure 5.3A, inset**). Therefore, it could be concluded that the minimal T3SS sub-structure formed with His-MxiGΔN was the IMR, indicating that MxiG can still self-associate and interact with MxiJ in the absence of MxiG-N. This is in agreement with evidence both from Chapter 3, which suggested that MxiG-N did not participate in high affinity interactions with either itself or $MxiJ_{232\text{-}241}$, and from the flexibility of the MxiG-N ring implicated by *S. flexneri* and *S. typhimurium* EM reconstructions (Hodgkinson *et al.*, 2009, Marlovits *et al.*, 2004, Schraidt & Marlovits, 2011). However, the abundance of His-MxiGΔN subcomplexes was much lower than for the His-tagged basal bodies, as assessed by the relative sizes of the void peaks from SEC

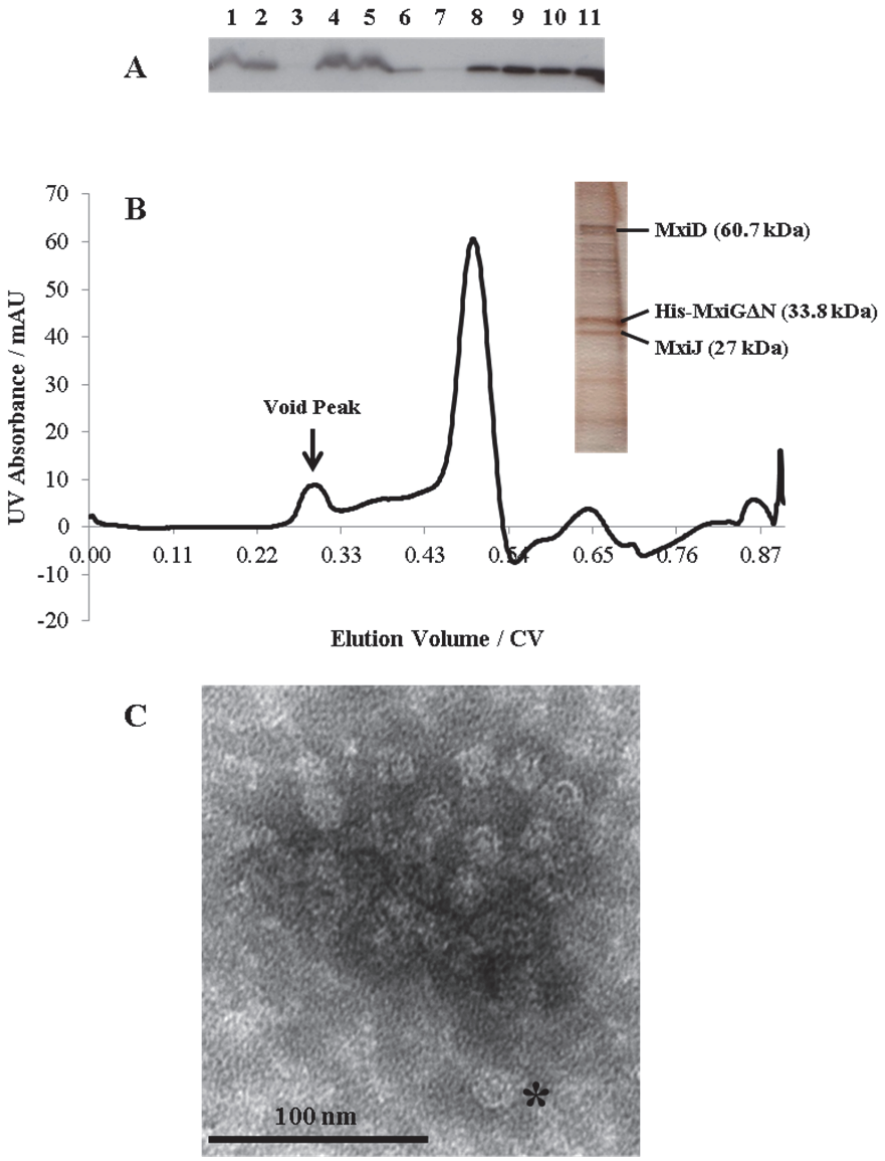


Figure 5.8 *Purification of complexes containing His-MxiGΔN.*

A Western blot of samples taken throughout the purification using mouse α-MxiJ primary antibody. 1: supernatant after 18,000 x g spin, 2: pellet after 18,000 x g spin, 3: supernatant after 120,000 x g spin, 4: pellet after 120,000 x g spin, 5: flow-through after batch-binding to Ni-NTA agarose, 6: flow-through from Ni-NTA agarose wash with 0.1% lauroylsarcosin, 7: flow-through from Ni-NTA agarose wash with 0.1% Triton X-100, 8: elution from Ni-NTA agarose, 9: concentrated fractions from Ni-NTA agarose elution, 10: void fractions from SEC and 11: concentrated void fractions from SEC.

B Elution profile of His-MxiGΔN complexes from a Superdex 200 10/300 GL column. Elution with 25 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% Triton X-100. The void fraction from SEC was run on 15% SDS-PA gel that was silver-stained using SilverQuest Staining kit (Invitrogen), as shown in the inset. The putative bands for the basal body components are indicated with their relative molecular weights.

C TEM image showing His-MxiGΔN complexes acquired at 30,000 x magnification. The void fractions after SEC were concentrated ~7-fold, applied to a formvar-carbon coated copper grid and negatively stained with 2% uranyl acetate (pH 4.0), before imaging with a Philips Tecnai T12 TEM. The asterisk denotes a typical circular structure obtained.

(**Figure 5.1B, Figure 5.8B**) and the dispersion of complexes on the EM grids (**Figure 5.3A, Figure 5.8C**). This would imply that His-MxiGΔN subcomplexes are less robust, suggesting MxiG-N could still be involved in low affinity interactions that make a significant stabilising contribution within the assembled IMR, as discussed in Chapter 3.

The apparent purification of basal body sub-complexes from the membrane fraction of *mxiG*$^{-}$ pBAD-*mxiGΔN* indicates that His-MxiGΔN is able to be correctly inserted into the inner membrane. This validates the conclusion that MxiG-N was essential for T3SS assembly and activity, as shown by the impaired binding to CR and overnight leakage of early effectors by *mxiG*$^{-/\Delta N}$ presented in Chapter 3 (**Figure 3.1**). Interestingly, it appeared that no needle appendages were protruding from the His-MxiGΔN complexes (**Figure 5.8C**), indicating that the needle either does not form or is significantly destabilised in the absence of MxiG-N. The observed phenotype of *mxiG*$^{-/\Delta N}$ would suggest that the former is more likely, implicating MxiG-N in a role in needle assembly. Indeed, the density corresponding to the N-terminal domain of PrgH was seen to undergo a clamping movement when EM images of intact NCs and basal bodies from *S. typhimurium* were compared (Marlovits *et al.*, 2004), suggesting conformational changes in the MxiG-N ring may directly contribute to the substrate specificity switching that must occur during T3SS assembly. This would correlate with the archetypal cell-signalling role for a FHA domain (Tsai, 2002), although it is unknown how this functionality would occur given that MxiG-N lacks the canonical P-Thr binding site, as discussed in Chapter 3. Alternatively, and perhaps more likely, the MxiG-N ring may simply act as a structural scaffold, tethering the cytoplasmic components which regulate needle assembly to the base of the T3SS. Although the attempts to characterise the *in vivo* interaction partners of MxiG-N presented in Chapter 3 were unsuccessful, previous studies have indicated an interaction with the middle portion of the C-ring component Spa33 (Morita-Ishihara *et al.*, 2006, Barison *et al.*, 2012), which is seen to be localised at the base of the T3SS (Morita-Ishihara *et al.*, 2006). Indeed a *spa33*$^{-}$ strain was shown by both *in vitro* and *in vivo* TEM analysis to only produce basal bodies lacking the needle appendage, indicating Spa33 was required for needle assembly

(Morita-Ishihara *et al.*, 2006). Therefore, based on results from this study, it is proposed that *S. flexneri* strains deficient in MxiG-N produce basal bodies that are unable to localise Spa33 at their base, emulating the phenotype observed for *spa33*$^{-}$. This hypothesis implies that the putative interaction between Spa33 and MxiG-N is essential for the functionality of Spa33, although both the molecular nature of the interaction and whether the interaction is direct or mediated by a mutual interaction partner are yet to be unambiguously determined. Regardless, the C-ring formed by Spa33 appears to be crucial for the correct assembly and functionality of the *S. flexneri* T3SS and was therefore an important target for further characterisation, as will be discussed in the final chapters.

5.4 Perspectives

His-tagged basal bodies were successfully isolated from *mxiG*$^{-}$*/mxiH*$^{-}$ and purified by affinity chromatography and SEC, enabling their visualisation by TEM. However, the yield and stability of samples obtained with a variety of detergents was insufficient for crystallisation trials and suggested the basal bodies were only weakly associated in the absence of the inner rod and needle. Purification and visualisation of His-MxiGΔN complexes from *mxiG*$^{-}$ indicated limited formation of the IMR, suggesting any low affinity interactions mediated by MxiG-N would likely have a role in stabilisation rather than assembly of the basal body. It was hypothesised that the absence of needle structures in His-MxiGΔN complexes was due to the loss of the putative interaction between MxiG-N and Spa33 preventing the C-ring mediating secretion by the T3SS.

6. Structural Characterisation of Spa33-C

As outlined in the introduction, Spa33 is an essential cytoplasmic regulator of the *Shigella* T3SS, being required from the onset of secretion for assembly of the needle (Morita-Ishihara *et al.*, 2006). Although Spa33 appears to be a central component of the T3SS interaction network (**Table 1.6**), the entire protein remained structurally uncharacterised. Homologous structures of Spa33-C were available from flagellar (*T. maritima* FliN)(Brown *et al.*, 2005) and plant-associated (*P. syringae* $HrcQ_B$-C)(Fadouloglou *et al.*, 2004) T3SSs, however there were no equivalent structures from either an animal pathogen or a T3SS with a single C-ring protein. In addition, a multiple sequence alignment of residues 208-293 of Spa33 ($Spa33_{208-293}$) with the C-termini of homologous proteins from animal, plant and flagellar T3SSs showed a pairwise sequence identity of 16-34%, indicating structural differences could exist between Spa33-C and previously determined structures (**Figure 6.1**). Therefore a high resolution structure of Spa33-C was determined by X-ray crystallography, as outlined in this chapter.

```
Spa33_shigella  --ESEHTEVSLALFNYDDINVKVDFILLEKNMTINELKMYVENELFKFPDDI-VKHVNIK 264
SpaO_salmonella ---------AETLPGLNQLPVKLEFVLYRKNVTLAELEAMGQQQLLSLPTNV-ELNVEIM 270
YscQ_yersinia   --HESDELNPEPLTDLNQLPVQVSFEVGRQILDWHTLTSLEPGSLIDLTTPV-DGEVRLL 276
FliN_thermotoga SETRKTEVPSDKLELLLDIPLKVTVELGRTRMTLKRVLEMIHGSIIELDKLT-GEPVDIL 117
FliN_brady      ------------LEAVFDVPVQVSAVLGRSKMDVGELLKLGPGTVLELDRRV-GEAIDIY 85
HrcQ_ralstonia  ---------AGQGVPLEQLEVPVHLELAVMGMPLAELAALQPQHVLTLPVKIRDVSVRLV 322
HrcQB_pseudomonas   GAMDPQDEPPALDSLALDLTLRCGELRLTLAELRRLDAGTILEVTGIS-PGHATLC 99
                                .: : :        :    :       :: .          :

Spa33_shigella          VNGSLVGHGELVSIEDGYGIEISSWMVKE-------- 293
SpaO_salmonella         ANGVLLGNGELVQMNDTLGVEIHEWLSES-------- 299 (34%)
YscQ_yersinia           ANGRLLGHGRLVEIQGRLGVRIERLTEVTIS------ 307 (20%)
FliN_thermotoga         VNGKLIARGEVVVIDENFGVRITEIVSPKERLELLNE 154 (22%)structure
FliN_bradyrhizobium     VNNKLVARGEVVLVEDKLGVTMTEIIKT-ERT----- 116 (24%)
HrcQ_ralstonia          CHGQTLGHGQLVAVGEQLGLQIASIGKHH-------- 351 (22%)
HrcQB_pseudomonas       HGEQVVAEGELVDVEGRLGLQITRLVTRS-------- 128 (16%)structure
                             :..*.:* :    *: :
```

Figure 6.1 *Sequence alignments of Spa33-C with the C-termini of homologous T3SS proteins*.

Sequence alignments of the C-termini of Spa33_shigella (*S. flexneri*), SpaO_salmonella (*S. typhimurium*), YscQ_yersinia (*Yersinia pseudotuberculosis*), FliN_thermotoga (*T. maritima*), FliN_brady (*Bradyrhizobium japonicum*), HrcQ_ralstonia (*Ralstonia solanacearum*) and HrcQB_pseudomonas (*P. syringae*) were performed using ClustalW2 (Chenna *et al.*, 2003), with identical, conserved and semi-conserved residues being indicated by asterisks, colons and full-stops respectively. Sequence numbers are shown relative to the full-length protein. The pairwise identity with the Spa33 sequence and sequences for which a crystal structure was available prior to this work are indicated in bold after each sequence. Proteins from animal-associated, flagellar and plant-associated T3SSs are shown in red, blue and green respectively.

6.1 Purification of $Spa33_{208-293}$

In previous work by Miriam Friede (Part II Project), a protocol for the expression and purification of $Spa33_{208-293}$ with a LEHHHHHH C-terminal tag ($Spa33_{208-293}$-His) had been established, providing a high yield, homogeneous sample. Furthermore, the His-tag could be selectively digested to produce $Spa33_{208-293}$, which was subsequently subjected to crystallisation trials. Protein crystal formation was observed in several broad screen conditions, indicating the construct could be amenable to characterisation by X-ray crystallography. However, datasets obtained were typically low resolution (≤3.5 Å), resulting in unsuccessful molecular replacement (MR) attempts. In an attempt to obtain a higher resolution dataset, $Spa33_{208-293}$ was re-purified and subjected to further crystallisation trials, as outlined in the following sections.

pET28b-*$spa33_{208-293}$* (donated by M. Friede) was transformed into B834 DE3 (pLysS) cells using the heat-shock protocol (**Appendix 8.2.1**) in the presence of 50 µg/ml kanamycin. 2 x 1 l LB with 50 µg/ml kanamycin were inoculated with a resuspension of colonies from the agar plates and the resulting cultures grown at 37°C in a shaking incubator. When an OD_{600} of ~0.6 was reached, expression of $Spa33_{208-293}$-His was induced with 1 mM IPTG and allowed to proceed overnight at 21°C. SDS-PAGE analysis confirmed overexpression of $Spa33_{208-293}$-His at the expected molecular weight of 11.1 kDa (**Figure 6.2**). Cells were then harvested and stored at -20°C.

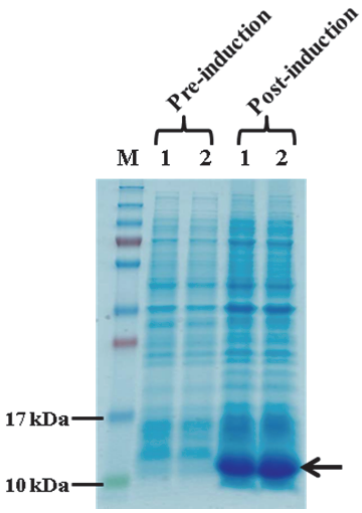


Figure 6.2 *15% SDS-PA gel showing over-expression of $Spa33_{208-293}$-His in B834 (DE3) pLysS.*

Samples taken from repeat cultures 1 and 2 pre- and post-induction with 1 mM IPTG are labelled. The arrow indicates the band corresponding to $Spa33_{208-293}$-His. M: PageRuler Plus Prestained Protein Ladder (Fermentas).

B834 DE3 (pLysS) pET28b-*spa33*$_{208\text{-}293}$ cells were resuspended in 50 mM Tris-HCl (pH 7.5), 300 mM NaCl and supplemented with lysozyme, DNase I and an EDTA-free protease inhibitor. Cells were subsequently lysed, fractionated and Spa33$_{208\text{-}293}$-His purified from the soluble fraction by Ni-affinity chromatography, as described for the purification of MxiG-N$_{1\text{-}126}$ in Chapter 2. The first three fractions eluted from the Ni-cartridge were applied to 10000 MWCO SnakeSkin Dialysis Tubing (Thermo scientific) and dialysed against 4 x 1 l 20 mM Tris-HCl (pH 7.5), 150 mM NaCl overnight at 4^{o}C. Although a small amount of precipitate had to be subsequently removed using a 0.22 µm syringe filter, only a 10% reduction in the pre-dialysis yield had occurred. The dialysed sample was then concentrated to ~10 ml using an Amicon Ultra centrifugal device with a 3 kDa cut-off (Millipore) and the concentration of Spa33$_{208\text{-}293}$-His was estimated, allowing carboxypeptidase A (type II bovine pancreas, Sigma) to be added to a final w/w protein:enzyme ratio of 25:1. The digest was once again passed through a 0.22 µm syringe filter before injection on to a HiLoad 26/60 Superdex 75 pg column (GE Healthcare) pre-equilibrated in 20 mM HEPES (pH 7.5), 150 mM NaCl. The resulting elution profile had one major elution peak, which corresponded to a single species of the correct molecular weight for Spa33$_{208\text{-}293}$ (10.0 kDa) from SDS-PAGE (**Figure 6.2A**). Fractions eluted within this peak were pooled and subsequently analysed by mass spectrometry, as outlined in Chapter 2. This confirmed that carboxypeptidase A had digested the C-terminal His-tag to produce Spa33$_{208\text{-}293}$, although E293 was additionally removed in ~60% species. Estimation of the protein concentration of this sample indicated that ~45 mg Spa33$_{208\text{-}293}$ had been purified per 1 l B834 DE3 (pLysS) pET28b-*spa33*$_{208\text{-}293}$. In addition, a species eluted in a minor peak after Spa33$_{208\text{-}293}$ during SEC had the expected molecular weight for carboxypeptidase A (35 kDa) from SDS-PAGE, indicating the contaminating enzyme had been successfully separated from Spa33$_{208\text{-}293}$ during SEC (**Figure 6.2A, asterisk**).

Protein crystals had been previously obtained with both the unmodified Spa33$_{208\text{-}293}$ construct and a construct with the lysine residues methylated. Therefore half of the purified Spa33$_{208\text{-}293}$

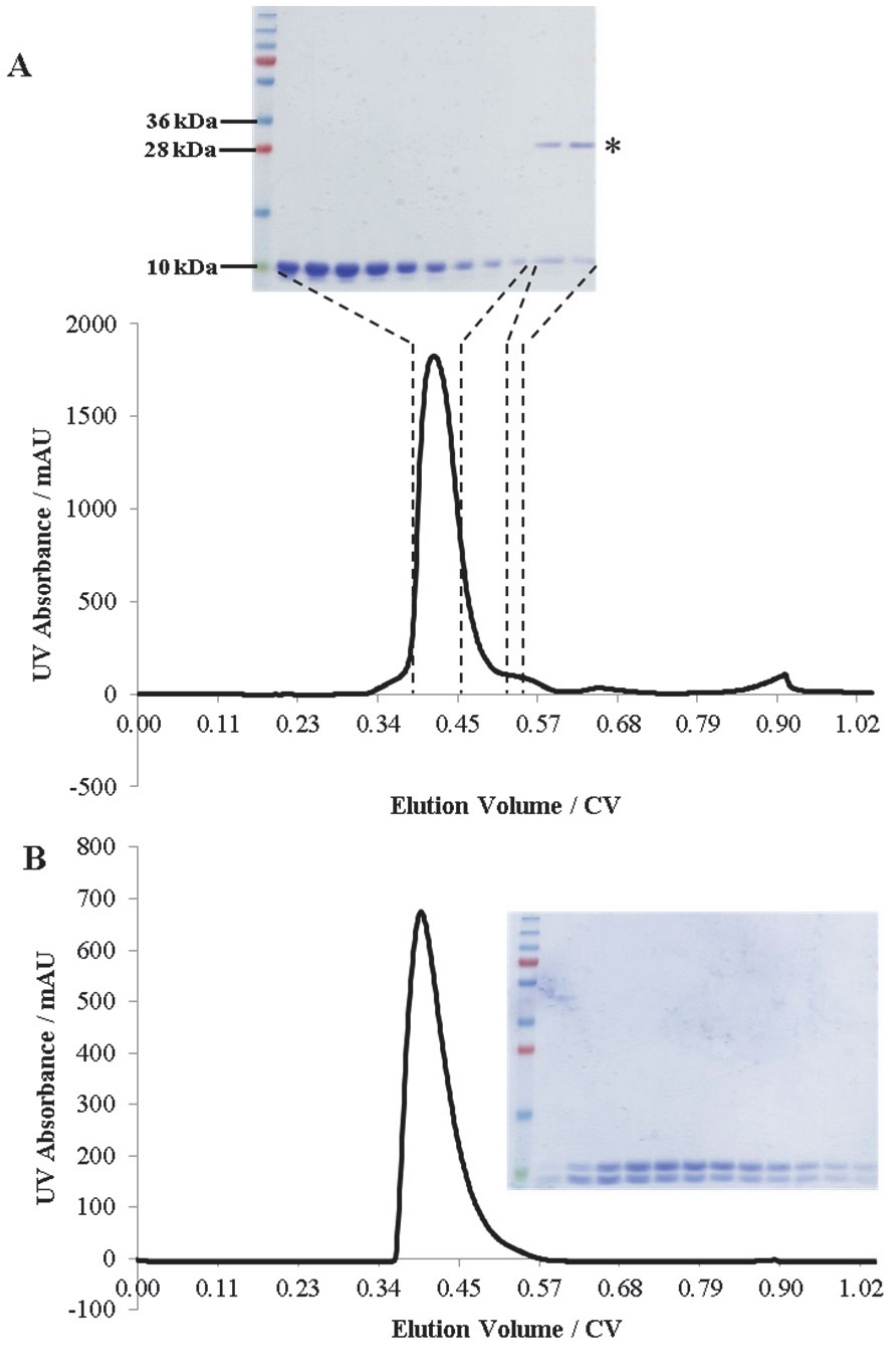


Figure 6.2 *Purification of Spa33$_{208-293}$.*

M: PageRuler Plus Prestained Protein Ladder (Fermentas).

A Elution profile of Spa33$_{208-293}$ after digestion with carboxypeptidase A from a HiLoad 16/60 Superdex 75 pg column. Elution with 20 mM HEPES (pH 7.5), 150 mM NaCl. The 15% SDS-PA gel shows samples taken from across the major and minor elution peaks, with the dashed lines indicating the relative positions of their corresponding fractions in the elution profile. The asterisk denotes the putative band for carboxypeptidase A.

B Elution profile for methylated Spa33$_{208-293}$ from a HiLoad 26/60 Superdex 75 pg column. Elution with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. The 15% SDS-PA gel shown in the inset shows samples taken from across the main elution peak.

was methylated according to the protocol outlined in Chapter 4 for MxiG-$C_{161-362/4}$ and in Walter et al., 2006. The quenched reactions were concentrated to 10 ml and the SEC purification step was repeated, instead using 20 mM Tris-HCl (pH 7.5), 150 mM NaCl as the buffer. The single peak in the elution profile was asymmetric, indicating a heterogeneous species (**Figure 6.2B**). Indeed, when the corresponding fractions were subjected to SDS-PAGE, two equimolar species were evident (**Figure 6.2B, inset**). However, when these fractions were pooled and analysed by mass spectrometry, a single species was identified (10.2 kDa) that had the mass increase expected for complete methylation of the amines of the seven lysine residues and the N-terminus of $Spa33_{208-293}$. Therefore the underlying heterogeneity of methylated $Spa33_{208-293}$ suggested by SEC and SDS-PAGE was intriguing and possibly indicated either that one species was not identifiable by mass spectrometry or that the methylated construct did not fully denature during SDS-PAGE, exhibiting conformational variation.

6.2 Crystallisation Trials with $Spa33_{208-293}$

Crystal trays with $Spa33_{208-293}$ samples were set up as described for MxiG-$C_{161-362/4}$ samples, using 400 nl drops containing both 1:1 and 3:1 protein solution:mother liquor and an incubation temperature of 21^{o}C. Broad screening of potential crystallisation conditions was conducted using HT-96 broad screens from Molecular Dimensions (**Table 4.6**) and a range of protein concentrations. As previous crystallisation trials with ≤10 mg/ml unmethylated and methylated $Spa33_{208-293}$ had resulted in slow crystal formation, protein concentrations up to 23 mg/ml were screened in an attempt to increase the speed of nucleation events and crystal growth. Additionally, various concentrations of unmethylated and methylated $Spa33_{208-293}$ were screened against the home-made screens, Optimisation Screen 1 (**Table 6.1**) and Optimisation Screen 2 (**Table 6.2**), which were designed to provide minor variations on the three crystallisation conditions that had previously yielded the best crystals of $Spa33_{208-293}$.

Table 6.1 Optimisation Screen 1
(based on 20% v/v isopropanol, 0.1 M Na Citrate (pH 5.6), 20% v/v PEG 4000 and 10% v/v isopropanol, 0.1 M Na Citrate (pH 5.5), 16% v/v PEG 4000 crystallisation conditions)

		% v/v PEG 4000											
		14	15	16	17	18	20	14	15	16	17	18	20
% v/v isopropanol	10	0.1 M Na Citrate (pH 5.3)						0.1 M Na Citrate (pH 5.5)					
	12.5												
	15												
	20												
	10	0.1 M Na Citrate (pH 5.6)						0.1 M Na Citrate (pH 5.7)					
	12.5												
	15												
	20												

Table 6.2 Optimisation Screen 2
(based on 0.2 M $MgCl_2$, 0.1 M Na Cacodylate (pH 6.5) 10% v/v PEG 3000 crystallisation condition)

		% v/v PEG 3000											
		8	9	10	11	12	13	8	9	10	11	12	13
$MgCl_2$ concentration / M	0.19	0.1 M Na Cacodylate (pH 6.4)						0.1 M Na Cacodylate (pH 6.5)					
	0.20												
	0.21												
	0.22												
	0.19	0.1 M Na Cacodylate (pH 6.6)						0.1 M Na Cacodylate (pH 6.7)					
	0.20												
	0.21												
	0.22												

Several $Spa33_{208-293}$ samples were also supplemented with subtilisin (protease, from *Bacillus licheniformis*; Sigma) at a w/w enzyme:protein ratio of 1:10000, in an attempt to remove any flexible termini or loops that could be impeding crystallisation. In total, 3840 crystallisation conditions were screened with various $Spa33_{208-293}$ samples, as summarised in **Table 6.3**.

Table 6.3 Crystallisation trials with $Spa33_{208-293}$

Sample	**Purification Buffer**	**96-Well Screens**
10 mg/ml $Spa33_{208-293}$	20 mM HEPES (pH 7.5), 150 mM NaCl	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex, Optimisation Screen 1, Optimisation Screen 2.
5 mg/ml $Spa33_{208-293}$	20 mM HEPES (pH 7.5), 150 mM NaCl	Optimisation Screen 1, Optimisation Screen 2.
15 mg/ml $Spa33_{208-293}$	20 mM HEPES (pH 7.5), 150 mM NaCl	Optimisation Screen 1, Optimisation Screen 2.
7 mg/ml $Spa33_{208-293}$ and 0.7 µg/ml subtilisin	20 mM HEPES (pH 7.5), 150 mM NaCl	Optimisation Screen 1, Optimisation Screen 2.
23 mg/ml methylated $Spa33_{208-293}$	20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex, Optimisation Screen 1, Optimisation Screen 2.
15 mg/ml methylated $Spa33_{208-293}$	20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex, Optimisation Screen 1, Optimisation Screen 2.
10 mg/ml methylated $Spa33_{208-293}$	20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Stura Footprint Screen & MacroSol, JCSG-*plus*, Structure Screen I and II, PACT *premier*, Morpheus, ProPlex, Optimisation Screen 1, Optimisation Screen 2.
8 mg/ml methylated $Spa33_{208-293}$	20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Optimisation Screen 1, Optimisation Screen 2.

After one month, a rod-shaped 200 x 50 x 50 µm crystal had grown amongst heavy precipitate within a 1:1 drop of 10 mg/ml $Spa33_{208-293}$ and condition B7 of the ProPlex screen (10% v/v isopropanol, 0.1 M Na HEPES (pH 7.0), 10% PEG 4000; **Figure 6.3A**). Promisingly, the presence of isopropanol and PEG 4000 in this condition mimicked a condition where protein crystals had previously been obtained (**Table 6.1**). In addition, several drops containing conditions from Optimisation Screen 1 had only grown needle-shaped crystals in the presence of subtilisin, indicating proteolysis had accelerated crystal formation. Needles of approximately 100 µm length within the 3:1 drop of 7 mg/ml $Spa33_{208-293}$ and 0.7 µg/ml subtilisin and condition E10 of Optimisation Screen 1 (10% v/v isopropanol, 0.1 M Na Citrate (pH 5.7), 17% PEG 4000) were the largest (**Figure 6.3B**). No crystal formation was observed for methylated $Spa33_{208-293}$, perhaps suggesting that the observed sample heterogeneity had impeded crystallisation (**Figure 6.2B**).

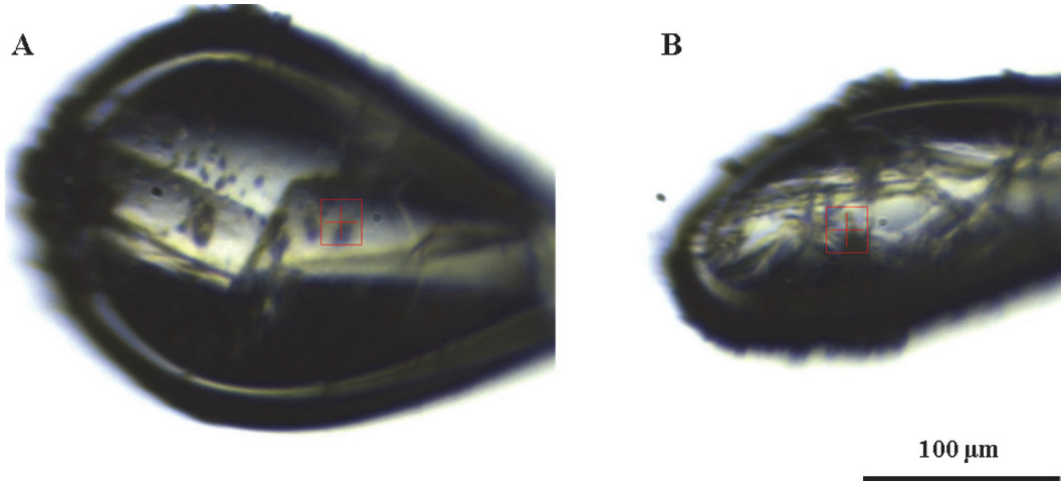


Figure 6.3 *Crystals of $Spa33_{208-293}$*. Visualisation after fishing with a Cryoloop.

A Rod-shaped crystal grown at 21°C within a 400 nl drop of 10 mg/ml $Spa33_{208-293}$ mixed 1:1 with condition B7 of the ProPlex HT-96 broad screen (10% v/v isopropanol, 0.1 M Na HEPES (pH 7.0), 10% PEG 4000). The crystal is seen to have broken into two pieces during fishing.

B Needle-shaped crystals grown at 21°C within a 400 nl drop of 7 mg/ml $Spa33_{208-293}$ and 0.7 µg/ml subtilisin mixed 3:1 with condition E10 of Optimisation Screen 1 (10% v/v isopropanol, 0.1 M Na Citrate (pH 5.7), 17% PEG 4000).

1:4 v/v ethylene glycol:mother liquor was selected as the cryo-protectant for the crystals shown in **Figure 6.3**, based on the similarity of their crystallisation conditions to conditions for which

optimal cryo-protectants had previously been determined (McFerrin & Snell, 2002). These crystals were subsequently fished with CrystalCap nylon Cryoloops (Hampton research) and briefly soaked in cryo-protectant, before flash-freezing and storage in liquid N_2.

6.3 Determination of the Crystal Structure of $Spa33_{208-293}$

Data Collection

Data collection was performed by Dr Pietro Roversi at beamline I04 of the Diamond Light Source (Oxfordshire, UK). Diffraction images were collected at 100 K to minimise radiation damage and by rotating the centred crystal around a single axis while being exposed to a 0.979 Å wavelength X-ray beam. The four initial images that were acquired for each of the crystals shown in **Figure 6.3** both showed reflections characteristic of a protein crystal (**Figure 6.4**). Only the rod-shaped crystal was used to acquire a full dataset, as the images (**Figure 6.4A**) showed a significantly higher resolution (~2.3 Å) than those for the needle-shaped crystals (~40 Å; **Figure 6.4B**).

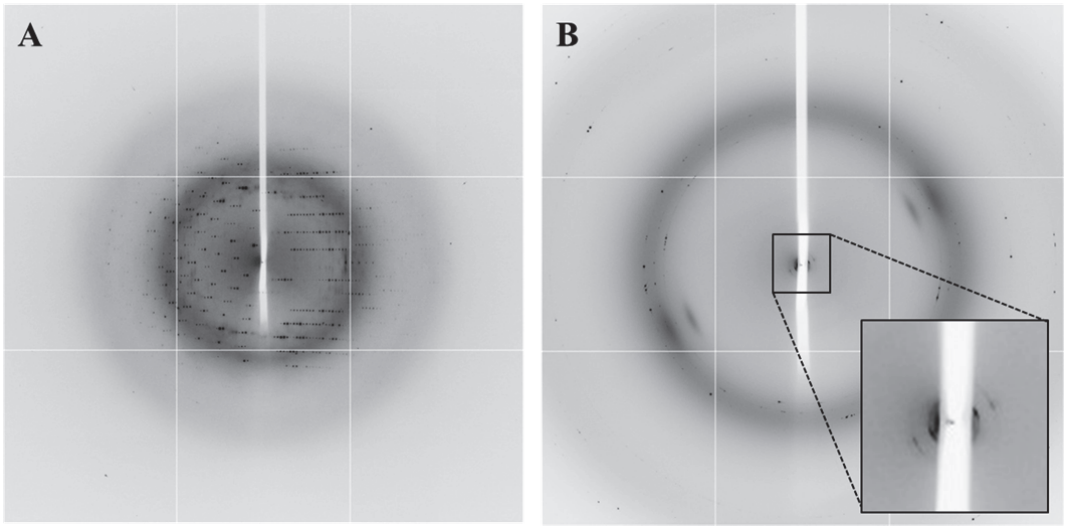


Figure 6.4 *Diffraction patterns obtained for $Spa33_{208-293}$ crystals*.

Images acquired at 100 K after exposure of the crystal to the X-ray beam at beamline I04 (Diamond Light Source). **A** Reflections obtained for the rod-shaped crystal (**Figure 6.3A**) were at a higher resolution (~2.3 Å) than **B** reflections obtained for the needle-shaped crystals (**Figure 6.3B**). The zoomed inset in **B** shows reflections at ~40 Å resolution.

The initial high resolution images for the rod-shaped crystal (**Figure 6.4A**) were indexed by iMosflm (Leslie, 2006) at the beamline, suggesting the crystal had a mosaicity of 0.57^{o} and a P222 lattice symmetry. This indicated an appropriate strategy for data collection and allowed centring of the crystal in the beam. Two datasets were obtained from a 90^{o} rotation of a distinct region of the crystal: one dataset with an oscillation angle ($\Delta\varphi$) of 0.25^{o} (360 images) and one with a $\Delta\varphi$ of 0.5^{o} (180 images).

Data Processing

Diffraction images are processed in order to assess the quality of the data and extract the experimental parameters required for structure determination. The output is a list of the Miller indices (h,k,l) of the unique reflections and the corresponding intensity (I_{hkl}) and standard deviation (σ_{hkl}) for each. In addition, unit cell dimensions and Laue group are determined, which characterise the crystal lattice. There are four main stages during the processing of diffraction images:

1. *Indexing* – reflections are identified on each image, allowing a prediction of unit cell dimensions and Laue group from repeating distances and symmetry within the diffraction pattern.
2. *Integration* – intensity and intensity standard deviation for each reflection are determined via a summation of pixel values within the spot region relative to the background. The distribution of partially recorded reflections over several images must be taken into account in determining accurate reflection positions and intensities.
3. *Scaling* – the intensities of symmetry-related and multiply-measured reflections are brought on to the same scale. The many symmetry-related reflections present within a dataset should have the same intensities, but in practice differ between images due to variations in beam intensity, crystal orientation, anisotropic absorption, radiation damage and random noise (Evans, 2006).

4. *Merging* – intensities from symmetry-related and multiply-measured reflections are averaged to provide a value I_{hkl} for each unique reflection.

The two $Spa33_{208\text{-}293}$ datasets were processed using Xia2 with the -3dii command implemented (Winter, 2010). Within this pipeline, XDS was used for indexing and integration via 3D profiling of reflections present over several images, before scaling and merging using XSCALE (Kabsch, 2010). The data quality and resolution were subsequently assessed using average I/σ(I), R_{merge} and R_{pim} (**Table 6.4**).

Table 6.4 Data quality statistics used during processing.

Data quality statistic	Definition
$\langle I/\sigma(\mathrm{I})\rangle = \frac{1}{N}\sum_{h}^{N}\frac{\lvert I_h\rvert}{\sigma(I_h)}$	Indication of signal-to-noise ratio for a resolution shell given by average I/σ(I) summed over N reflections.
$R_{merge} = \frac{\sum_h \sum_i^N \lvert I_{(h)i} - \langle I_h\rangle\rvert}{\sum_h \sum_i^N I_{(h)i}}$	Measurement of the internal consistency of the data after scaling and merging. Average of the deviations of the intensity of each redundant reflection h from the scaled intensity $\langle I_h\rangle$ for each unique reflection. Increases with increasing multiplicity.
$R_{pim} = \frac{\sum_h (\frac{1}{N-1})^{1/2} \sum_i^N \lvert I_{(h)i} - (I_h)\rvert}{\sum_h \sum_i^N I_{(h)i}}$	$1/(N\text{-}1)^{1/2}$ term takes multiplicity into account, so the expected increase in precision with increase in redundancy is reflected by this statistic.

The statistics from the processing of $Spa33_{208\text{-}293}$ datasets are shown in **Table 6.5**. Indexing showed this $Spa33_{208\text{-}293}$ crystal form to be primitive orthorhombic, with a minimal point group of 222. Subsequently systematic absences along all reciprocal axes indicated that three screw axes were present in the lattice and therefore that the likely space group was $P2_12_12_1$ (**Figure 6.5A-C**). However, an anomaly in the systematic absences along k suggested that absolute

confidence in the space group assignment should not be taken until the final structure solution was obtained (**Figure 6.5B**).

Data quality statistics (**Table 6.4**) averaged over all the data were favourable and did not show significant variation between images, allowing them all to be included during processing (not shown). Furthermore, the combined datasets provided almost a complete set of unique reflections. The average I/σ(I) showed that a measurable signal was present above the noise up to a resolution of 2.28 Å, which was taken as the highest resolution shell. R_{merge} was seen to be consistently <3% for low resolution reflections, but increased dramatically for reflections at a resolution higher than ~2.7 Å (**Figure 6.5D**). R_{merge} of 79% in the highest resolution shell did not correlate with the signal-to-noise ratio via the expected relationship ($R_{merge} \sim 0.8 / \langle I/\sigma(I) \rangle$), which would predict a value of ~30%. However, R_{merge} was likely inflated by the 7-fold redundancy in the data, therefore the precision of the data was better represented by R_{pim}, which took multiplicity into account. Indeed, an R_{pim} of 32% for the outer shell was in good agreement with the expected value, providing an internal consistency check of the data quality. Together, these acceptable data quality statistics indicated that the processed $Spa33_{208\text{-}293}$ data were suitable for further analyses.

Table 6.5 Statistics from processing of $Spa33_{208\text{-}293}$ datasets.

Space group	$P2_12_12_1$
Unit cell dimensions (Å)	a = 27.8, b = 51.3, c = 104.5
Unit cell angles	$\alpha = \beta = \gamma = 90^o$
Resolution (Å)*	51.3-2.28 (2.34-2.28)
Number of unique reflections	7291
Multiplicity*	6.7 (7.0)
Completeness (%)*	99.8 (100)
R_{merge}*	0.057 (0.79)
R_{pim}*	0.024 (0.32)
Average I/σ(I)*	19.0 (2.6)

*Values in brackets are for the highest resolution shell.

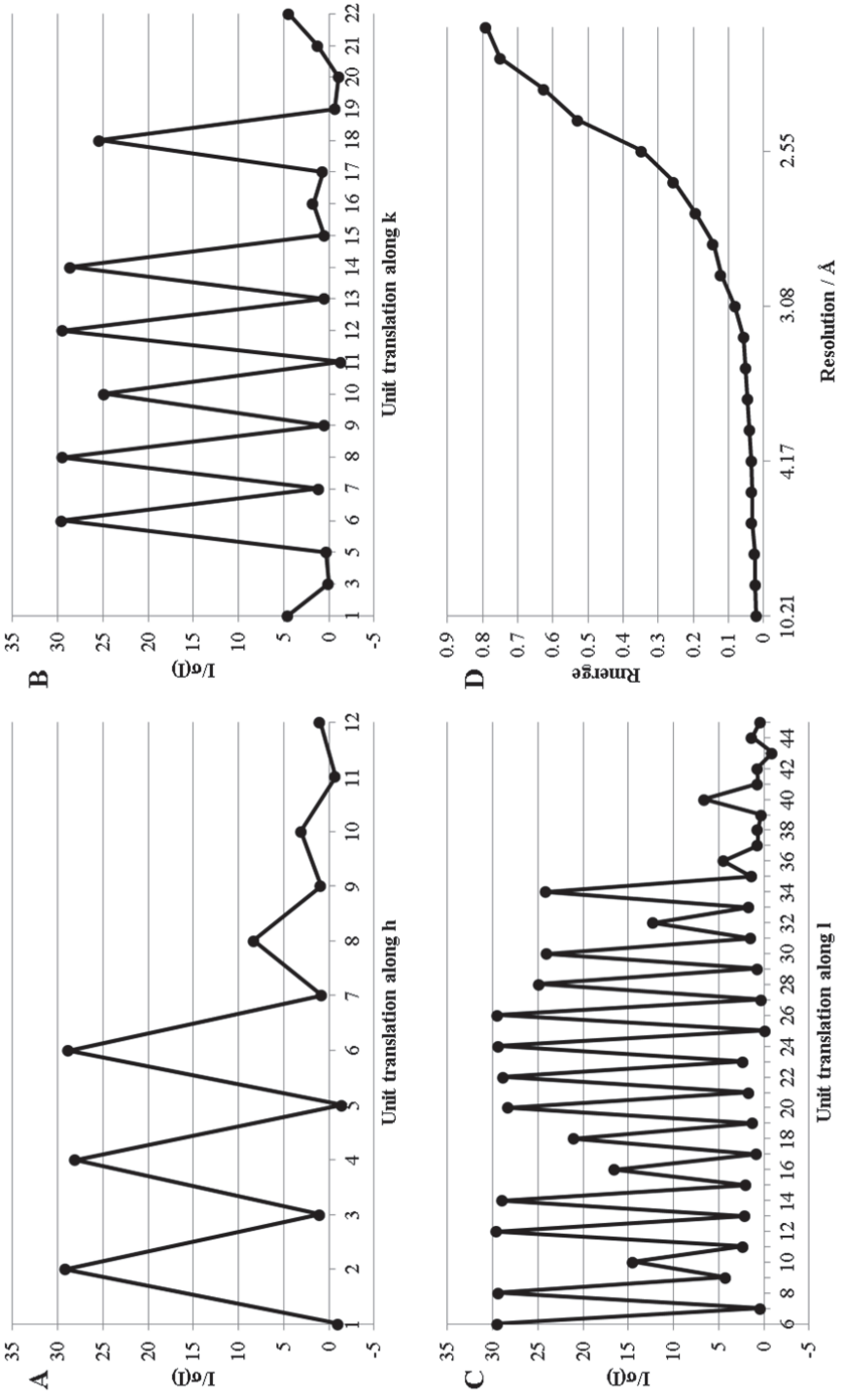


Figure 6.5 *Processing of Spa33$_{208-293}$ datasets.*

A-C Systematic absences along the **A** h, **B** k and **C** l reciprocal axes are diagnostic of three screw axes within the crystal lattice.

D Analysis of R_{merge} against resolution for the highest resolution shells.

Characterisation of the Asymmetric Unit

The asymmetric unit can contain multiple copies of the protein molecule related by translational or rotational non crystallographic symmetry (NCS), resulting in further redundant information that can assist structure determination. The number of $Spa33_{208-293}$ molecules within the asymmetric unit was initially assessed by estimation of the solvent content of the crystal. The Matthews coefficient (V_M), or ratio of unit cell volume to protein molecular weight, is related to the solvent content (V_S) of the asymmetric unit ($V_S = 1 - (1.23/V_M)$)(Matthews, 1968, Matthews, 1976). V_M and V_S for different possible copy numbers of $Spa33_{208-293}$ in the asymmetric unit was assessed using the Collaborative Computational Project Number 4 (CCP4) software suite (Winn *et al.*, 2011), as outlined in **Table 6.6**. Most crystals have a V_S of 27-78%, with the most common value being 43% (Matthews, 1968, Matthews, 1976), therefore a V_S of 34% for the $Spa33_{208-293}$ crystal was determined as most likely (**Table 6.6, yellow**). This indicated that a $Spa33_{208-293}$ dimer was most likely to be present within the asymmetric unit, as had been observed previously in homologous structures (Fadouloglou *et al.*, 2004, Brown *et al.*, 2005). However, it was noted that the presence of a monomer within the asymmetric unit was also possible from the V_S calculation (**Table 6.6**) and that any potential $Spa33_{208-293}$ dimer could be crystallographic if the $P2_12_12_1$ space group was subsequently found to be incorrect.

Table 6.6 Estimation of the number of copies of $Spa33_{208-293}$ in the asymmetric unit.

Number $Spa33_{208-293}$ molecules per asymmetric unit (molecular weight)	**V_M / Å/Da**	**V_S / %**	**Probability across all resolution ranges**
1 (10.0 kDa)	3.73	67	0.35
2 (20.0 kDa)	1.87	34	0.63
3 (30.0 kDa)	1.24	1	0.01

The nature of the NCS that related the likely two copies of $Spa33_{208-293}$ within the asymmetric unit was probed further with the Patterson function, as this gave a representation of all

interatomic vectors in the unit cell. If pure or nearly pure translational NCS was present, a native Patterson map calculated from experimental I_{hkl} would show a strong peak corresponding to the translational vector. The native Patterson map for $Spa33_{208-293}$ generated by fast fourier transform in CCP4 only showed a strong intensity peak (89σ) at the origin corresponding to the zero distance self vector peaks. All other peaks were <7σ intensity, indicating pure translational NCS did not relate two copies of $Spa33_{208-293}$ within the asymmetric unit.

The self-rotation function (SRF) monitors the rotation of the native Patterson map upon itself, giving peaks at rotation angles where the maps align. When the SRF is calculated over a volume near the origin, significant peaks aside from those corresponding to crystallographic symmetry are indicative of rotational NCS. The SRF for the $Spa33_{208-293}$ crystal was calculated from the native Patterson map using MOLREP (Vagin & Teplyakov, 1997) within CCP4 and data with a 4-15 Å resolution. The $\kappa = 180^{o}$ section showed the expected peaks corresponding to the three two-fold crystallographic axes (**Figure 6.6**), however no other peaks were found to be of an intensity significantly above the noise. This was surprising given that the crystal structures of the $HrcQ_B$-C (Fadouloglou *et al.*, 2004) and FliN (Brown *et al.*, 2005) dimers showed them to exhibit 2-fold rotational symmetry. Therefore, either only one copy of $Spa33_{208-293}$ was present in the asymmetric unit, the two copies of $Spa33_{208-293}$ exhibited a different type of NCS or peaks corresponding to 2-fold rotational NCS were obscured by the peaks corresponding to space group symmetry in the $\kappa = 180^{o}$ section (**Figure 6.6**).

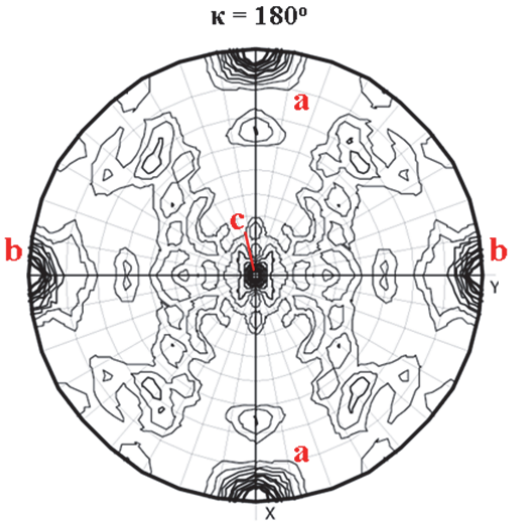


Figure 6.6 *$Spa33_{208-293}$ self rotation function.*

The $\kappa = 180^{o}$ section is shown as produced by MOLREP (Vagin & Teplyakov, 1997). Peaks labelled **a**, **b** and **c** correspond to the three two-fold crystallographic axes, whilst all other peaks are below the noise.

Phasing by Molecular Replacement

In order to reconstruct the electron density of $Spa33_{208-293}$, the 'phase problem' had to be overcome; I_{hkl} derived from the datasets only provided the amplitudes of the scattered waves, requiring an initial estimate of their phase angles to proceed further. As the sequences of FliN from *T. maritima* and $HrcQ_B$-C from *P. syringae* showed 22% and 16% identity with $Spa33_{208-293}$ (**Figure 6.1**), it was likely these proteins showed some structural homology to $Spa33_{208-293}$. Indeed, the crystal structures of FliN (Brown *et al.*, 2005) and $HrcQ_B$-C (Fadouloglou *et al.*, 2004), which share a 23% sequence identity, superimposed with a C^{α} RMSD value of 1.62 Å over 138 atoms (**Figure 6.7A**). Therefore phasing via MR using FliN (pdb id 1yab) and $HrcQ_B$-C (pdb id 1o9y) as search models was likely to be feasible but challenging due to the low sequence identity.

MR was carried out with Phaser (McCoy *et al.*, 2007) in CCP4, which uses maximum likelihood functions to rotate then translate the search model within the unit cell, until the best agreement between model- and data-derived structure factor amplitudes are reached. The success of MR with different search models was then assessed by the log-likelihood gain (LLG) and the rotation and translation function Z-scores (RFZ and TFZ respectively). The LLG indicated how much better the data could be predicted from the oriented model relative to a random-atom model, whilst the Z-scores showed how many standard deviations the LLG for the solution was above the mean LLG for randomly oriented models. The number of clashes <3 Å between symmetry related molecules was also assessed.

Despite the absence of a clear peak corresponding to 2-fold rotational NCS in the SRF (**Figure 6.6**), the presence of a $Spa33_{208-293}$ dimer in the asymmetric unit could be predicted both from the calculated V_S (**Table 6.6**) and the conserved dimer observed in crystal structures of homologous proteins (Brown *et al.*, 2005, Fadouloglou *et al.*, 2004). Therefore dimers of FliN and $HrcQ_B$-C were assessed as suitable search models for MR. The *T. maritima* FliN dimer (pdb id 1yab) was initially used as a search model as it exhibited higher sequence homology with

$Spa33_{208-293}$ than *P. syringae* $HrcQ_B$-C (**Figure 6.1**). However, the two solutions had six clashes <3 Å and a TFZ of 4.5 (**Table 6.7**), indicating it had a low signal-to-noise ratio and was likely to be incorrect. In an attempt to increase the structural homology between $Spa33_{208-293}$ and the search model and therefore the likelihood of finding a minimally biased solution by MR, the search models were improved as follows:

- *Ensembles* – FliN and $HrcQ_B$-C structures were superimposed by SSM within Superpose (Krissinel & Henrick, 2004). Phaser then used the ensemble to calculate an average structure, weighted in different regions according to the correlation between the structures.
- *Chainsaw models* – the sequence alignment with $Spa33_{208-293}$ (**Figure 6.1**) was used to create homology models based on the FliN and $HrcQ_B$-C backbone and prune non-conserved side-chains back to the C^{γ} atom within Chainsaw (Stein, 2008).
- *Trimming* – structurally dissimilar loop regions highlighted by the FliN and $HrcQ_B$-C superimposition (**Figure 6.7A**) were likely to also show variation in the $Spa33_{208-293}$ structure, so they were removed from several search models.
- *Monomers* – it was possible a single copy of $Spa33_{208-293}$ was present in the asymmetric unit from V_S calculation (**Table 6.6**) or a dimer could have a different arrangement from homologous proteins, so FliN and $HrcQ_B$-C monomers were used as search models.

Key examples of these modified search models used for phasing by MR are shown (**Figure 6.7**), along with the statistics associated with their solutions from Phaser (**Table 6.7**). Whilst using the $FliN_{CHAINSAW}/HrcQ_BC_{CHAINSAW}$ trimmed monomer ensemble (**Figure 6.7C**) as a search model yielded a high number of possible solutions, the $FliN_{CHAINSAW}/HrcQ_B$-$C_{CHAINSAW}$ dimer ensemble (**Figure 6.7A**) and the $FliN_{CHAINSAW}/HrcQ_B$-$C_{CHAINSAW}$ trimmed dimer ensemble (**Figure 6.7B**) gave a single solution with favourable LLG and Z-scores, so were chosen for further investigation.

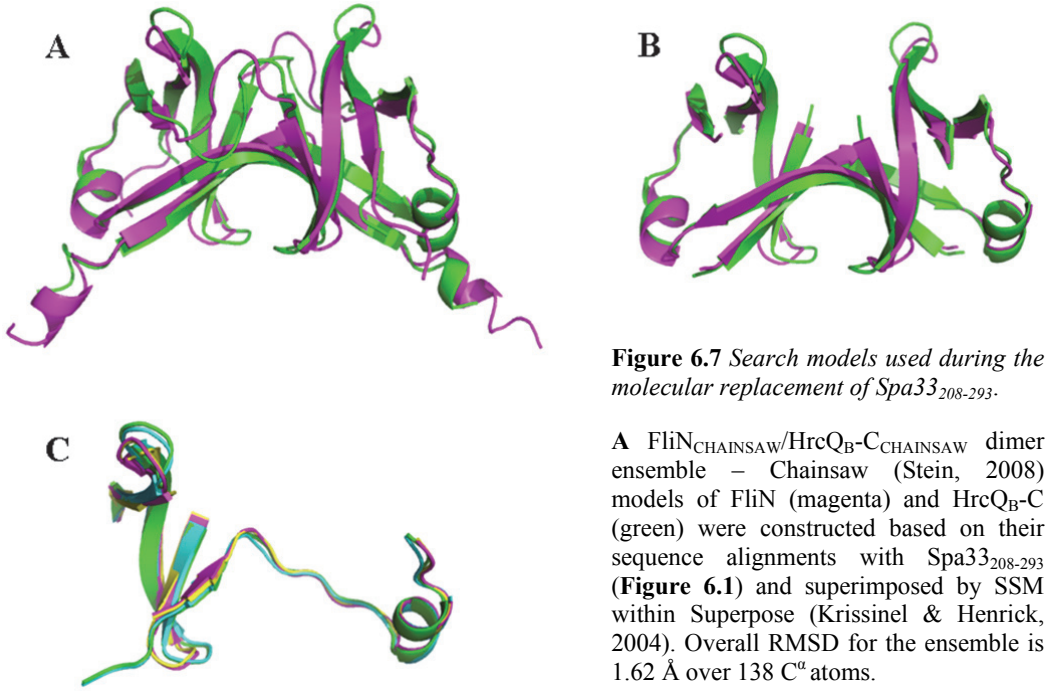


Figure 6.7 *Search models used during the molecular replacement of Spa33$_{208-293}$.*

A FliN$_{CHAINSAW}$/HrcQ$_B$-C$_{CHAINSAW}$ dimer ensemble – Chainsaw (Stein, 2008) models of FliN (magenta) and HrcQ$_B$-C (green) were constructed based on their sequence alignments with Spa33$_{208-293}$ (**Figure 6.1**) and superimposed by SSM within Superpose (Krissinel & Henrick, 2004). Overall RMSD for the ensemble is 1.62 Å over 138 C$^{\alpha}$ atoms.

B FliN$_{CHAINSAW}$/HrcQ$_B$-C$_{CHAINSAW}$ trimmed dimer ensemble – residues 19-46 and 55-82 of the FliN Chainsaw model (magenta) and residues 19-45 and 54-82 of the HrcQ$_B$-C Chainsaw model (green) were superimposed by SSM within Superpose. Overall RMSD for the ensemble is 1.37 Å over 110 C$^{\alpha}$ atoms.

C FliN$_{CHAINSAW}$/HrcQ$_B$C$_{CHAINSAW}$ trimmed monomer ensemble – residues 22-41 and 55-82 of chain B (yellow) of the FliN Chainsaw model and residues 19-41 and 55-82 of chains A (green) & D (cyan) of the HrcQ$_B$-C Chainsaw model were superimposed on residues 22-41 and 55-82 of chain A (magenta) of the FliN Chainsaw model by SSM within Superpose. Overall RMSD compared to FliN chain A over 48 C$^{\alpha}$ atoms is 0.33 Å, 1.28 Å and 1.12 Å for FliN chain B, HrcQ$_B$-C chain A and HrcQ$_B$-C chain D respectively.

Table 6.7 Molecular replacement solutions for Spa33$_{208-293}$ from Phaser.

Search model	Number of solutions	LLG	TFZ	RFZ	Number of clashes <3 Å
FliN dimer (pdb id 1yab)	2	24	4.5	5.5	6
FliN$_{CHAINSAW}$/HrcQ$_B$-C$_{CHAINSAW}$ dimer ensemble	1	37	7.1	4.5	0
FliN$_{CHAINSAW}$/HrcQ$_B$-C$_{CHAINSAW}$ trimmed dimer ensemble	1	31	7.0	5.4	0
FliN$_{CHAINSAW}$/HrcQ$_B$C$_{CHAINSAW}$ trimmed monomer ensemble	35	30	6.2	3.2	2

MR in Phaser with the $FliN_{CHAINSAW}/HrcQ_B\text{-}C_{CHAINSAW}$ trimmed dimer ensemble as the search model (**Table 6.7, yellow**) produced both an initial model for the structure of $Spa33_{208\text{-}293}$ and an initial corresponding electron density map, whose structure factors were obtained from experimentally-derived amplitudes and the model-derived phases. Comparison of the model and electron density map in Coot (Emsley *et al.*, 2010) revealed regions of electron density which could not be accounted for by the initial $Spa33_{208\text{-}293}$ model, but could be explained by expected additional $Spa33_{208\text{-}293}$ features not present in the search models. For example, there was likely density for I224 at the N-terminus of chain A, which was not present in the initial $Spa33_{208\text{-}293}$

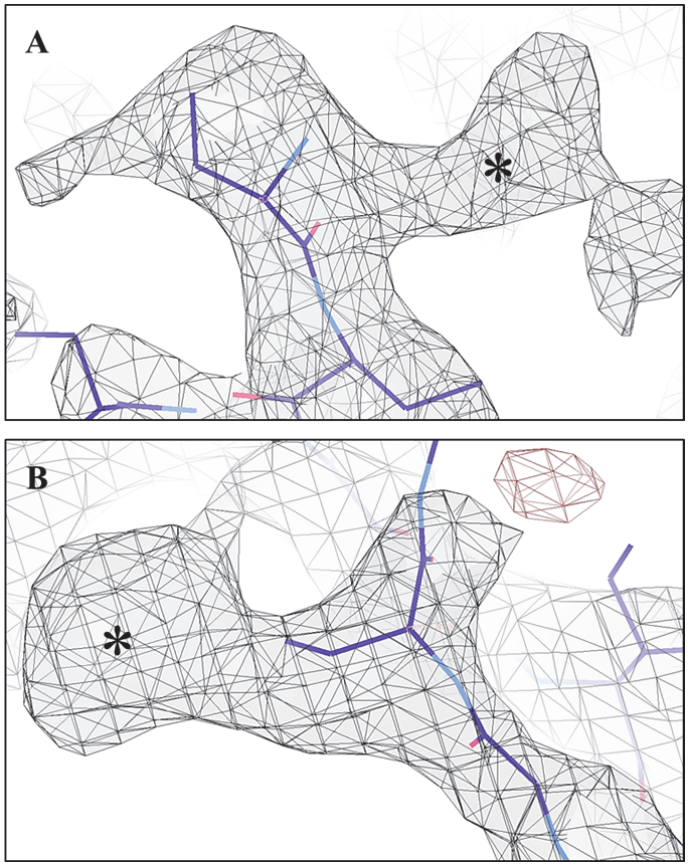


Figure 6.8 *Initial electron density maps for $Spa33_{208\text{-}293}$.*

The output from Phaser after MR with the $FliN_{CHAINSAW}/HrcQ_B\text{-}C_{CHAINSAW}$ trimmed dimer ensemble (**Figure 6.7B**) as the search model is shown in Coot (Emsley *et al.*, 2010). The initial electron density map (black cage) for $Spa33_{208\text{-}293}$ is contoured to 1.04σ. The corresponding initial model for $Spa33_{208\text{-}293}$ is shown by stick representation, with carbon, nitrogen and oxygen atoms shown in purple, blue and red respectively. The regions of the map shown represent additional density not fit by the search model, which can be accounted for by a $Spa33_{208\text{-}293}$ model.

A Density corresponding to I224 at the N-terminus of chain A of $Spa33_{208\text{-}293}$ (asterisk).
B Density corresponding to the side-chain of H271 of chain A of $Spa33_{208\text{-}293}$ (asterisk).

model due to trimming of the N-terminus of the search model (**Figure 6.7A, asterisk**). Similarly, there was density for the H271 side-chain of chain A, which had been pruned back to C^{γ} during construction of FliN and $HrcQ_B$-C Chainsaw models (**Figure 6.7B, asterisk**). These features of the electron density map indicated the MR solution was minimally biased towards the search model, and therefore that this $Spa33_{208-293}$ model could be used to provide an initial estimate of structure factor phases. Furthermore, this solution indicated that $Spa33_{208-293}$ was arranged as a dimer with a 2-fold rotational symmetry axis, as had been previously observed in homologous structures. The solution obtained by MR with the $FliN_{CHAINSAW}/HrcQ_B\text{-}C_{CHAINSAW}$ dimer ensemble was subsequently found to be the same, confirming the likely correct nature of this solution.

An initial improvement of these phases obtained by MR was obtained by density modification of the electron density map. Ten cycles of solvent flattening was carried out with 2-fold rotational NCS averaging applied using Parrot (Cowtan, 2010), which acted to raise the positive electron density within the protein region and smooth the weak electron density in the solvent region to a constant value. The output was a more defined electron density map for $Spa33_{208-293}$, which provided the structure factor phases for use in model building and structure refinement.

Model Building and Structure Refinement

The initial model for $Spa33_{208-293}$ was adjusted to give the best possible fit to the experimental data by successive rounds of real space model building and reciprocal space refinement. Due to trimming of the search model used for MR, the initial model for $Spa33_{208-293}$ was incomplete, comprising residues 225-252 and residues 261-288 for both chains A and B of the dimer. Automatic tracing of the polypeptide chain into the electron density using Buccaneer (Cowtan, 2006) was unsuccessful, therefore the $Spa33_{208-293}$ model was built and modified manually. This was achieved by comparison in Coot of the model and its respective electron density difference map, which subtracts the map calculated using the model-derived structure factor amplitudes (F_C) from that calculated using the experimentally-derived amplitudes (F_O). As the phases used

to calculate F_O were taken from the $Spa33_{208-293}$ model, a $2F_O$-F_C map was visualised to emphasise the experimental data and reduce model bias. Significant regions where the $Spa33_{208-293}$ model and electron density disagreed were identified by F_O-F_C difference map peaks above 4σ intensity, showing negative electron density where the model had features not evidenced by

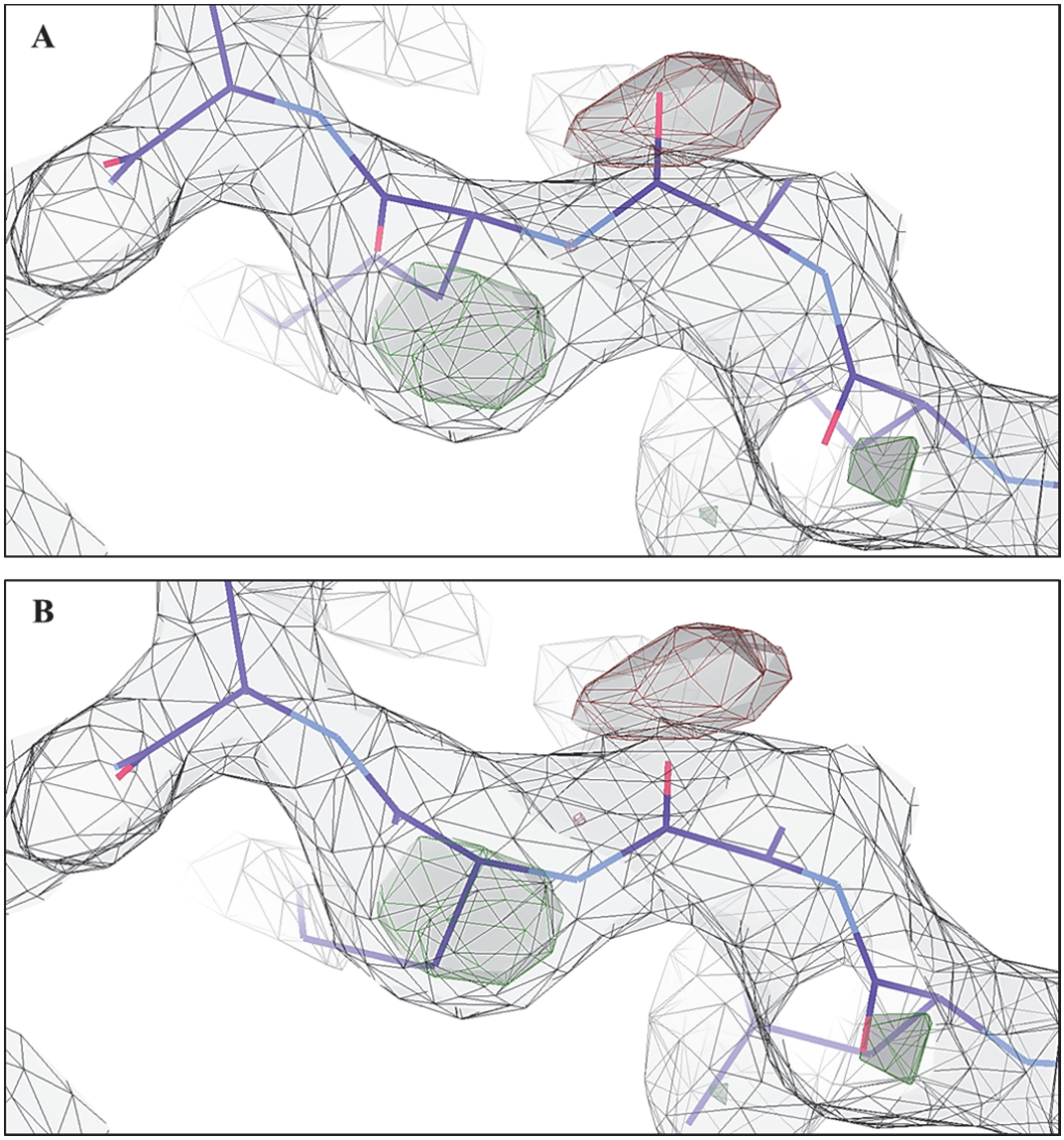


Figure 6.9 *The use of difference electron density maps to modify the $Spa33_{208-293}$ model.*

The $Spa33_{208-293}$ model is shown as a stick representation, with carbon, nitrogen and oxygen atoms in purple, blue and red respectively. The corresponding electron density difference maps are shown as a cage representation contoured at 0.99σ, with $2F_O$-F_C electron density in black and the F_O-F_C positive and negative electron density in green and red respectively. Visualisation in Coot.

A K235 of chain B of the $Spa33_{208-293}$ model occupies regions that are not accounted for by the data (red electron density) and conversely does not occupy regions where the data suggests features are present (green electron density).

B The backbone of the $Spa33_{208-293}$ model has been modified via real space refinement in Coot so that K235 better fits the electron density difference map.

the data and positive features where the data have features not accounted for by the model. These difference map peaks were used to guide both building and modification of the $Spa33_{208-293}$ model so that it better fit the experimental data. For example, significant positive and negative F_O-F_C peaks showed that K235 of chain B of $Spa33_{208-293}$ occupied a region not supported by the data (**Figure 6.9A**), allowing rearrangement of the polypeptide backbone in this region to better account for the $2F_O$-F_C electron density (**Figure 6.9B**). Although adjustments to chain A and chain B in the $Spa33_{208-293}$ dimer model were made separately, display of the NCS ghost within Coot was used a guide where the difference electron density of the chains was sufficiently similar.

Each round of model adjustment was followed by a round of refinement of the model against the $Spa33_{208-293}$ data using autoBUSTER (Blanc *et al.*, 2004). This process adjusted the model to bring the experimentally- and model-derived structure factor amplitudes into closer alignment, whilst restraining bond lengths, bond angles and interatomic distances to those expected from protein chemistry. Specifically, autoBUSTER used a maximum likelihood function to optimise the model parameters (coordinates and the B-factor), performing the default refinement of five BIG cycles per run. In addition, the NCS displayed by the $Spa33_{208-293}$ dimer was used during refinement to apply local structure similarity restraints on atom positions where symmetry was evident.

The progress of structure refinement was assessed by the R-factor (R), defined using the experimentally- ($|F_O|$) and model-derived ($|F_C|$) structure factor amplitudes as follows:

$$R = \frac{\sum_h ||F_O| - |F_C||}{\sum_h |F_O|}$$

During successful refinement, R after each successive run would be expected to decrease as $|F_C|$ agrees more closely with $|F_O|$. The model bias introduced by using the same data for monitoring as for refinement was overcome by also following the free R-factor (R_{free}), calculated as above using 5% of reflections that were not included in the structure refinement with autoBUSTER.

This allowed model bias during refinement to be assessed by the divergence of R and R_{free}. Over 36 runs of refinement of the $Spa33_{208-293}$ model, R/R_{free} can be seen to dramatically decrease in parallel from the initial 43.6%/47.8% after the first refinement of the MR solution (**Figure 6.10**), indicating a marked improvement in the agreement of a relatively unbiased model with the experimental data. A notable improvement in R/R_{free} was obtained throughout initial model building and adjustment, particularly as a result of an increase in the weighting of the experimental data relative to the chemistry restraints during refinement by changing the XrayWeight_start parameter in autoBUSTER from 1.0 to 4.0 (**Figure 6.10, arrow**).

Figure 6.10 *Structure refinement of $Spa33_{208-293}$.*

The values of R and R_{free} after each cycle of refinement using autoBUSTER (Blanc *et al.*, 2004) are shown. A notable improvement in these values were due to an increase in the weighting of the experimental data relative to the chemical restraints during refinement (arrow).

During structure refinement, $Spa33_{208-293}$ models were also validated by checking the protein stereochemistry using MolProbity (Davis *et al.*, 2007). After run 23 of refinement, R/R_{free} indicated a good fit of the $Spa33_{208-293}$ model to the experimental data (**Figure 6.10**), however the model had two cis peptides and 2.26% outliers in the Ramachandran plot. The $Spa33_{208-293}$ model was therefore subjected to 10 further runs of model adjustment and refinement, with

manual correction of stereochemical outliers within Coot. The final $Spa33_{208-293}$ model showed a much better agreement with expected protein chemistry, with no cis peptides or Ramachandran outliers (**Table 6.8**).

Table 6.8 Refinement statistics for the final $Spa33_{208-293}$ model

Agreement with processed data	
R (%)	24.3
R_{free} (%)	27.2
Agreement with protein chemistry[1]	
Residues in favoured regions (%)	96.18
Residues in allowed regions (%)	100
Residues with bad bonds (%)	0.00
Residues with bad angles (%)	0.00
Clashscore[2]	9.42 (94^{th} percentile)
Poor side-chain rotamers (%)	3.91
MolProbity score	2.21 (80^{th} percentile)

[1]Determined using MolProbity (Davis *et al.*, 2007).
[2]Determined from a $Spa33_{208-293}$ model with the hydrogens added by MolProbity. Clashscore defined as the number of steric overlaps >0.4 Å per 1000 atoms.

The final model and $2F_O$-F_C map for $Spa33_{208-293}$ were in good agreement when visualised in Coot (**Figure 6.11**), supported by a final R/R_{free} of 24.3%/27.2% (**Table 6.8**). In addition to the polypeptide chains of the $Spa33_{208-293}$ dimer, 18 H_2O molecules had been fit into the density. However, clear electron density was not present for all the expected residues of the $Spa33_{208-293}$ construct during model building; the final model for the $Spa33_{208-293}$ dimer comprised residues 220-290 in chain A and residues 222-291 in chain B. In addition, D255 in chain A and D280 in chain B could not be modelled due to poor electron density in this region of the map. Nonetheless this model provided a good starting point for analysis of the Spa33-C structure and crystal lattice, as outlined in the following sections.

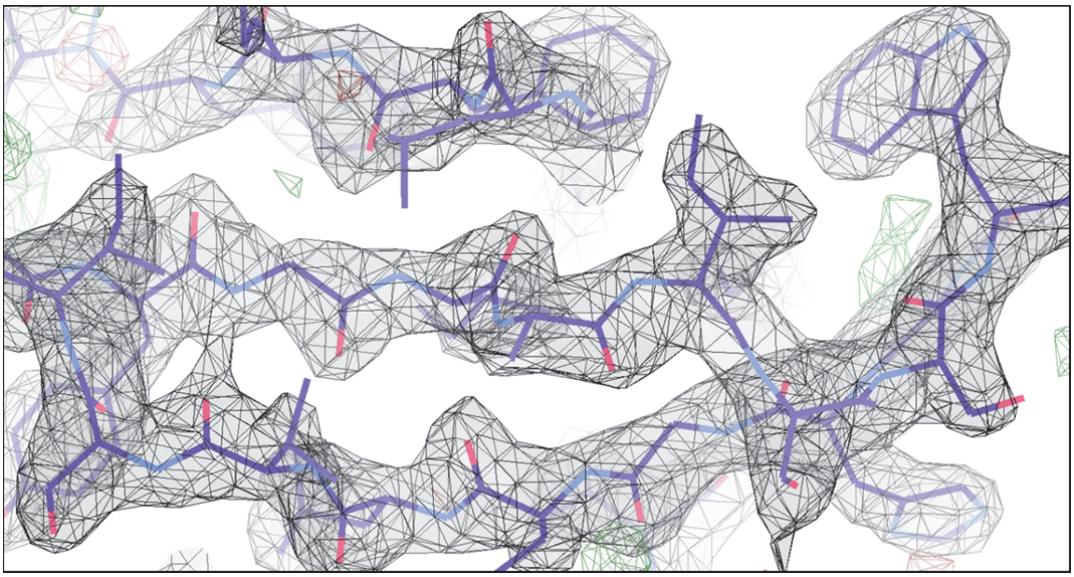


Figure 6.11 *The final model and electron density difference map for Spa33$_{208-293}$.*

The final Spa33$_{208-293}$ model is shown as a stick representation, with carbon, nitrogen and oxygen atoms in purple, blue and red respectively. The corresponding $2F_O$-F_C map is shown as a cage representation (black) contoured at 1.58σ. Visualisation in Coot.

6.4 Structural Analysis of Spa33$_{208-293}$

The Structure of Spa33$_{208-293}$

Spa33$_{208-293}$ forms a saddle-shaped dimer, with an approximate two-fold rotational symmetry axis (**Figure 6.12A**). Each Spa33$_{208-293}$ monomer has five β-strands and two small α-helices (**Figure 6.12E**). The dimer has two β-barrels, each formed of a 6-stranded antiparallel β-sheet comprising β1, β3, β4 and β5 of one chain and β1 and β2 of the other (**Figure 6.12B**). The α-helices are located in the linker regions between β1 and β2 and between β2 and β3 and are packed against the rest of this compact molecule (**Figure 6.12C**). Each copy of Spa33$_{208-293}$ is highly intertwined within the dimer, stabilised by antiparallel β-strand interactions; a β-ribbon formed by β1 of each copy bridges the two β-barrels and β2 of one copy and β5 of the other are hydrogen bonded within each barrel (**Figure 6.12D**).

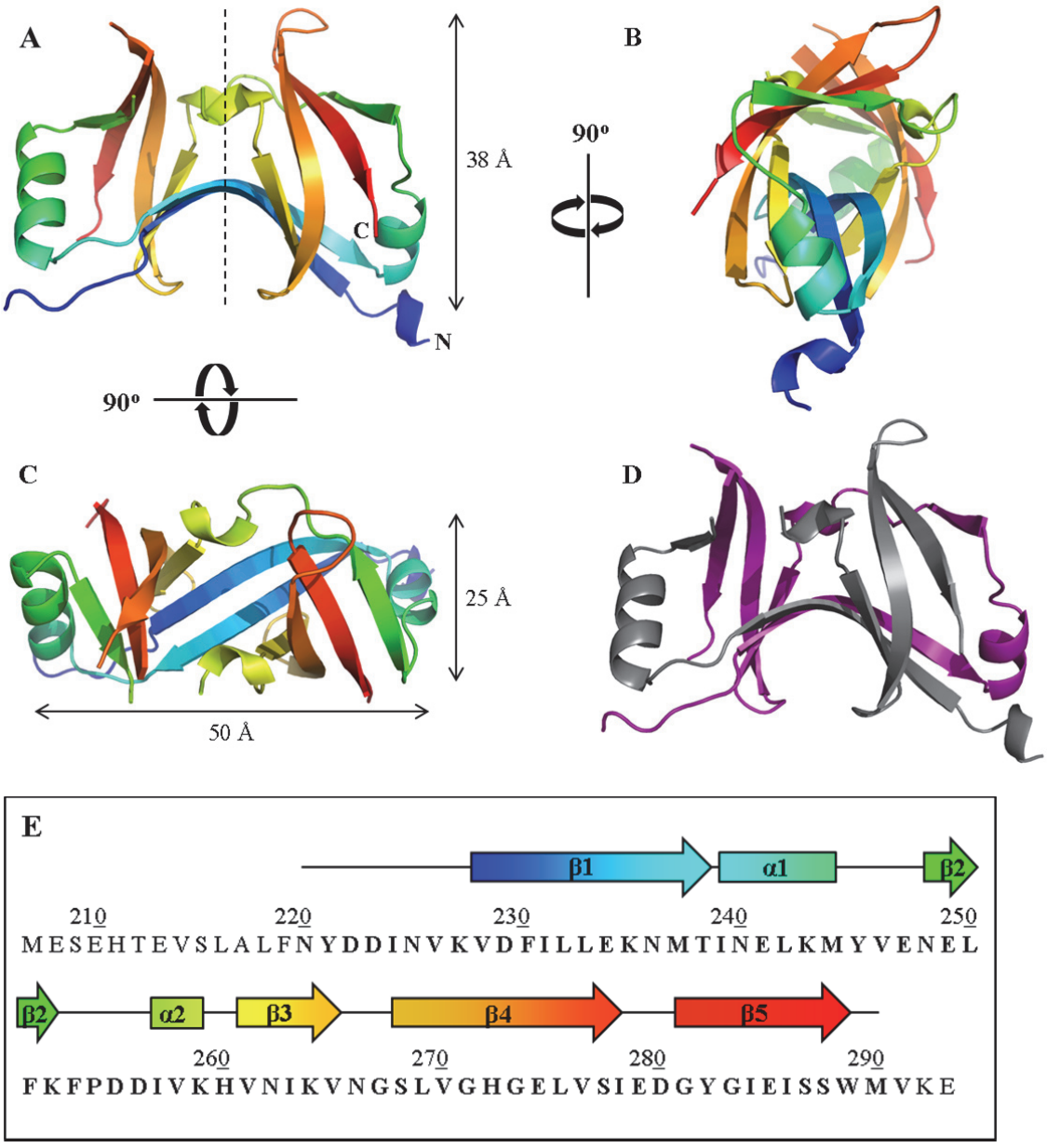


Figure 6.12 *The crystal structure of Spa33*$_{208\text{-}293}$.

A-C Cartoon representations of the Spa33$_{208\text{-}293}$ dimer coloured from the N-terminus (blue) to the C-terminus (red). Approximate dimensions determined from a sphere representation within PyMol are shown.
A View perpendicular to the two-fold symmetry axis for the dimer (dashed line). The termini of chain B are labeled.
B View from the N-terminus of chain B showing the β-barrel arrangement.
C View down the two-fold symmetry axis of the dimer.

D Cartoon representation of the Spa33$_{208\text{-}293}$ dimer showing the intertwining of chain A (purple) and chain B (grey).

E The sequence of Spa33$_{208\text{-}293}$ numbered according to the native *S. flexneri* Spa33 sequence, with residues represented in either chain A or chain B of the Spa$_{208\text{-}293}$ dimer structure shown in bold. The secondary structure is shown for chain B of the dimer structure, with β-strands (arrows) and α-helices (cylinders) coloured according to the cartoon representations shown in **A-C**.

Superimposition of chain A and chain B of the $Spa33_{208-293}$ dimer using SSM in Superpose showed minor structural differences between the monomers, with a C^{α} RMSD of 1.70 Å over 64 atoms (**Figure 6.13A**). Notably, the N-terminus (**Figure 6.13A, 1**), α1 (**Figure 6.13A, 2**) and the loop between β4 and β5 (**Figure 6.13A, 3**) did not align well between the $Spa33_{208-293}$ monomers either as a result of conformational flexibility or dimer asymmetry. The B-factor provides a measurement of the disorder for each atom, accounting for both the dynamic disorder from thermal vibrations and static disorder from conformational differences between molecules in different unit cells. When the C^{α} B-factors were mapped on to the structure of $Spa33_{208-293}$ there was seen to be a correlation between those regions that did not align well between the monomers and those that exhibited B-factors >60 $Å^2$ (**Figure 6.13B**). Specifically the backbone of chain A and chain B appeared to show disorder at the N-terminus and in the loop region between β4 and β5, indicating their structures likely differed between $Spa33_{208-293}$ monomers due to conformational flexibility. Interestingly however, the region immediately C-terminal to α1 was only disordered in chain B, with the equivalent helix in chain A being longer and more rigid. Therefore, true asymmetry within the $Spa33_{208-293}$ dimer could exist in this region.

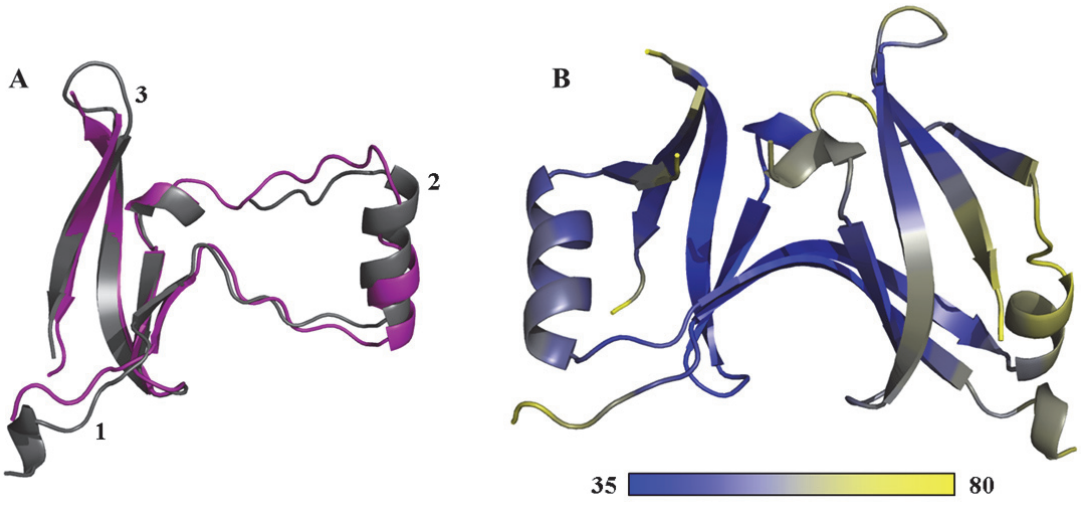


Figure 6.13 *Structural differences between the $Spa33_{208-293}$ monomers*.

A Superimposition of chain A (grey) and chain B (purple) of the $Spa33_{208-293}$ dimer obtained using SSM in Superpose. Overall RMSD for the ensemble is 1.70 Å for 64 C^{α} atoms. The numbers denote regions that do not structurally align: **1** N-terminus, **2** α1 and **3** the loop between β4 and β5.

B Cartoon representation of the $Spa33_{208-293}$ dimer coloured according to the C^{α} B-factor ($Å^2$).

Interestingly, the superimposition of chain B from the $Spa33_{208-293}$ dimer on to chain A showed they were related by the rotation (θ, φ, κ) 170°, 44°, 178°, suggesting that the SRF peak corresponding to the 2-fold rotational NCS was obscured by the 0°, 0°, 180° origin peak in the κ = 180° section (**Figure 6.6, c**). This was confirmed by visualisation of the unit cell and $P2_12_12_1$ space group, showing that the 2-fold rotational NCS axis for the $Spa33_{208-293}$ dimer was almost parallel to the 2-fold crystallographic axis in the c direction (**Figure 6.14B**).

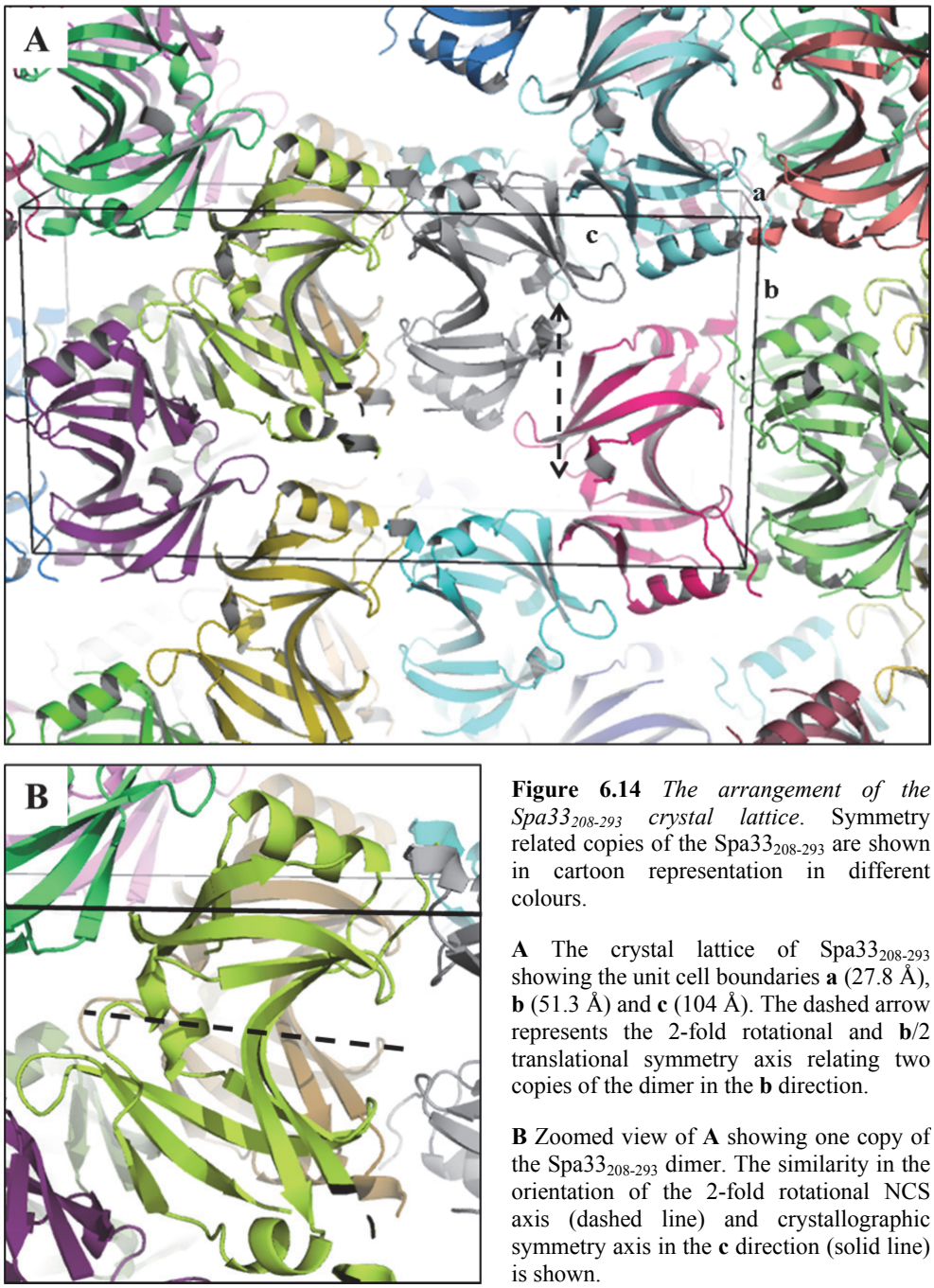


Figure 6.14 *The arrangement of the $Spa33_{208-293}$ crystal lattice*. Symmetry related copies of the $Spa33_{208-293}$ are shown in cartoon representation in different colours.

A The crystal lattice of $Spa33_{208-293}$ showing the unit cell boundaries **a** (27.8 Å), **b** (51.3 Å) and **c** (104 Å). The dashed arrow represents the 2-fold rotational and **b**/2 translational symmetry axis relating two copies of the dimer in the **b** direction.

B Zoomed view of **A** showing one copy of the $Spa33_{208-293}$ dimer. The similarity in the orientation of the 2-fold rotational NCS axis (dashed line) and crystallographic symmetry axis in the **c** direction (solid line) is shown.

Comparison with Homologous Structures

The crystal structures for several proteins homologous to Spa33-C have been determined, including *T. maritima* FliN (Brown *et al.*, 2005) and *P. syringae* $HrcQ_B$-C (Fadouloglou *et al.*, 2004), which were used in the construction of search models for MR (**Figure 6.7**). In addition, a 2.25 Å structure of residues 218-307 of YscQ (YscQ-C) from *Y. pseudotuberculosis* was published (Bzymek *et al.*, 2012) after the $Spa33_{208-293}$ structure was determined, providing a homologous structure from another animal pathogen. These homologous proteins all crystallised as highly intertwined saddle-shaped dimers, as observed for $Spa33_{208-293}$. Subsequent superimposition of the $Spa33_{208-293}$ dimer coordinates with each of these homologous dimer structures in turn showed a remarkably similar arrangement of secondary structure elements (**Figure 6.15**), indicating a conserved core structure. Minor structural variation in loop regions could be seen; most notably the loop between β4 and β5 was arranged differently between

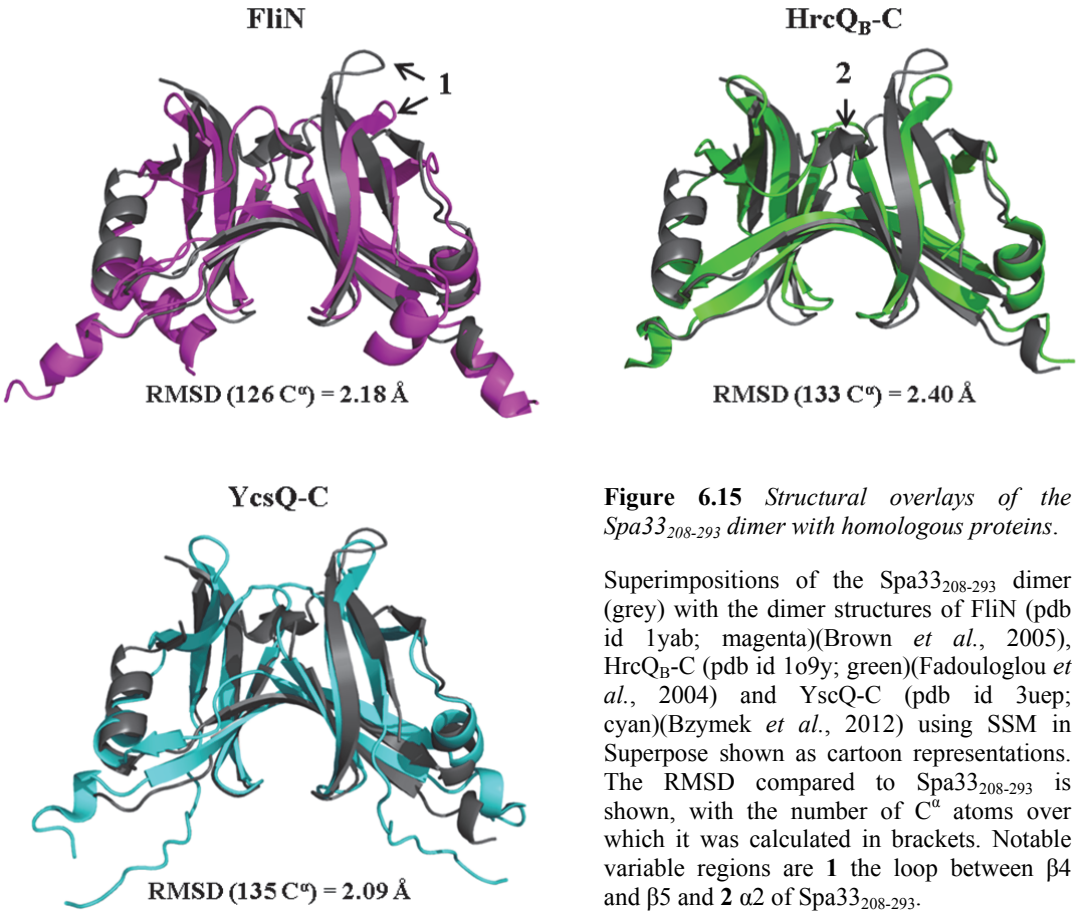


Figure 6.15 *Structural overlays of the $Spa33_{208-293}$ dimer with homologous proteins*.

Superimpositions of the $Spa33_{208-293}$ dimer (grey) with the dimer structures of FliN (pdb id 1yab; magenta)(Brown *et al.*, 2005), $HrcQ_B$-C (pdb id 1o9y; green)(Fadouloglou *et al.*, 2004) and YscQ-C (pdb id 3uep; cyan)(Bzymek *et al.*, 2012) using SSM in Superpose shown as cartoon representations. The RMSD compared to $Spa33_{208-293}$ is shown, with the number of C^{α} atoms over which it was calculated in brackets. Notable variable regions are **1** the loop between β4 and β5 and **2** α2 of $Spa33_{208-293}$.

$Spa33_{208-293}$ chain A and the homologous structures, particularly FliN (**Figure 6.15, arrow 1**), which could be due to the conformational flexibility indicated by the C^{α} B-factors in this region (**Figure 6.13B**). In addition, α2 appeared to be a unique feature of the $Spa33_{208-293}$ dimer (**Figure 6.15, arrow 2**), although the small size of the α-helix implicated a minor structural contribution. The poorest alignment to the $Spa33_{218-307}$ dimer structure was obtained with $HrcQ_B$-C (C^{α} RMSD over 133 atoms = 2.40 Å), in correlation with $HrcQ_B$-C having the lowest sequence identity with $Spa33_{218-307}$ (**Figure 6.1**). YscQ-C was best aligned with the $Spa33_{218-307}$ structure (C^{α} RMSD over 135 atoms = 2.09 Å), which is perhaps unsurprising given both proteins originate from an animal pathogenic T3SS and are likely involved in similar interactions. As the YscQ-C structure was determined using experimental phasing (Bzymek *et al.*, 2012), its high similarity with the $Spa33_{218-307}$ structure indicated significant model bias had not been introduced by using FliN and $HrcQ_B$-C as search models for MR.

Superimposition of the $Spa33_{218-307}$ monomers indicated regions of potential asymmetry in the dimer and gave a C^{α} RMSD of 1.70 Å over 64 atoms, as highlighted previously (**Figure 6.13A**). An equivalent overlay of FliN, $HrcQ_B$-C and YscQ-C yielded values of 0.91 Å over 79 C^{α} atoms, 0.95 Å over 71 C^{α} atoms and 1.31 Å over 84 C^{α} atoms respectively (not shown), indicating that dimer asymmetry was another parameter in which $Spa33_{218-307}$ and YscQ-C were most similar. However, the structural differences between YscQ-C monomers was largely manifested as disruption of β1 in one of the chains (Bzymek *et al.*, 2012), in contrast to differences within α1 and the loop region between β4 and β5 being most prominent between $Spa33_{218-307}$ monomers (**Figure 6.13A**). Therefore the more pronounced and differentially localised asymmetry of the $Spa33_{218-307}$ dimer could be due to flexible regions from each monomer being trapped in different conformations during crystallisation, as suggested by high C^{α} B-factors observed for structurally distinct regions (**Figure 6.13B**).

Whilst *T. maritima* FliN, *P. syringae* $HrcQ_B$-C and *Y. pseudotuberculosis* YscQ-C exhibited <25% sequence identity with $Spa33_{218-307}$, several residues were highly conserved between homologues from flagellar, plant-associated and animal-associated T3SSs (**Figure 6.1**).

Mapping of these residues on to the dimer structure of $Spa33_{218-307}$ showed them to be located in diverse regions (**Figure 6.16, A-C**). However, the majority of conserved residues were localised to the secondary structure elements, likely contributing to the formation of the highly conserved

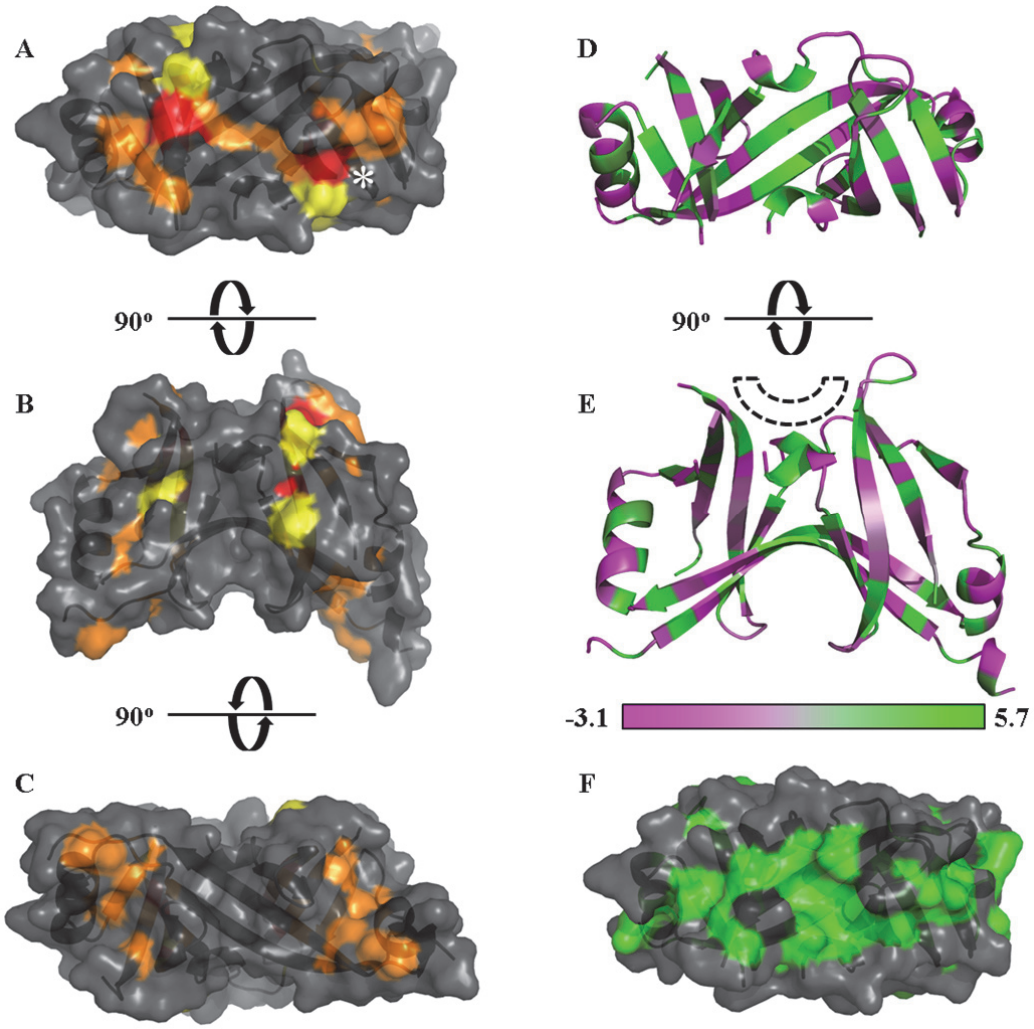


Figure 6.16 *Identification of functionally important residues within the $Spa33_{208-293}$ dimer.*

A-C Identical (red), conserved (orange) and semi-conserved (yellow) residues are mapped on to the surface representation according to the sequence alignments shown in **Figure 6.1**, with the cartoon representation shown underneath.
A View down the two-fold symmetry axis from the concave dimer surface (**Figure 6.12, C**). The asterisk denotes V276 of chain A.
B View perpendicular to the two-fold symmetry axis of the dimer (**Figure 6.12, A**).
C View down the two-fold symmetry axis from the convex dimer surface.

D & E Cartoon representation of the $Spa33_{208-293}$ dimer coloured according to the Cornette hydrophobicity scale (Cornette *et al.*, 1987).
D Equivalent view to **A**. The dashed rainbow denotes the conserved hydrophobic groove.
E Equivalent view to **B**.

F Residues with a hydrophobicity >3.0 (green) mapped on to the surface representation with the cartoon representation shown underneath.

fold found for $Spa33_{208-293}$ homologues (**Figure 6.15**). The absolutely conserved glycines, G273 and G283, were found within β4 and β5 respectively, buried in the core of the structure where large side-chains could not be accommodated. Notably, most identical and conserved residues were hydrophobic in nature (**Figure 6.1**), including the absolutely conserved V276 in $Spa33_{208-293}$ that was seen to contribute to the convex dimer surface (**Figure 6.16A, asterisk**). When the $Spa33_{208-293}$ dimer was coloured according to the Cornette hydrophobicity scale (Cornette *et al.*, 1987), a large number of hydrophobic residues expectedly contributed to the β-barrels (**Figure 6.16D&E**), with their side-chains pointing inwards. However, hydrophobic residues were also found to both make a significant contribution to the dimer interface and form extensive solvent-exposed patches, as had been previously observed in homologous structures (Fadouloglou *et al.*, 2004, Brown *et al.*, 2005, Bzymek *et al.*, 2012). In particular, a groove found at the dimer interface (**Figure 6.16E, dashed rainbow**) was seen to contain solvent-exposed hydrophobic residues (**Figure 6.16F**), several of which were conserved between $Spa33_{208-293}$ homologues (**Figure 6.16A**). As mutation of equivalent residues in *E. coli* FliN led to an impairment in flagellum assembly (Brown *et al.*, 2005), the hydrophobic groove in $Spa33_{208-293}$ could provide a conserved binding site for interaction partners.

The electrostatic surface of $Spa33_{208-293}$ calculated using APBS (Baker *et al.*, 2001) showed there to be a large distribution of negative potential (**Figure 6.17A-C**). In particular, the convex dimer surface was almost entirely negatively charged (**Figure 6.17A**), corresponding to the surface where the hydrophobic groove was identified (**Figure 6.16F**). Calculation of the electrostatic surface potential of FliN (**Figure 6.16D**), $HrcQ_B$-C (**Figure 6.16E**) and YscQ-C (**Figure 6.16F**) showed the equivalent regions to also possess large amounts of negative charge, suggesting this is a conserved feature of $Spa33_{208-293}$. This reinforced a potential role for this face of the $Spa33_{208-293}$ dimer in mediating protein interactions, suggesting both electrostatic and hydrophobic interactions could be important for conferring specific binding.

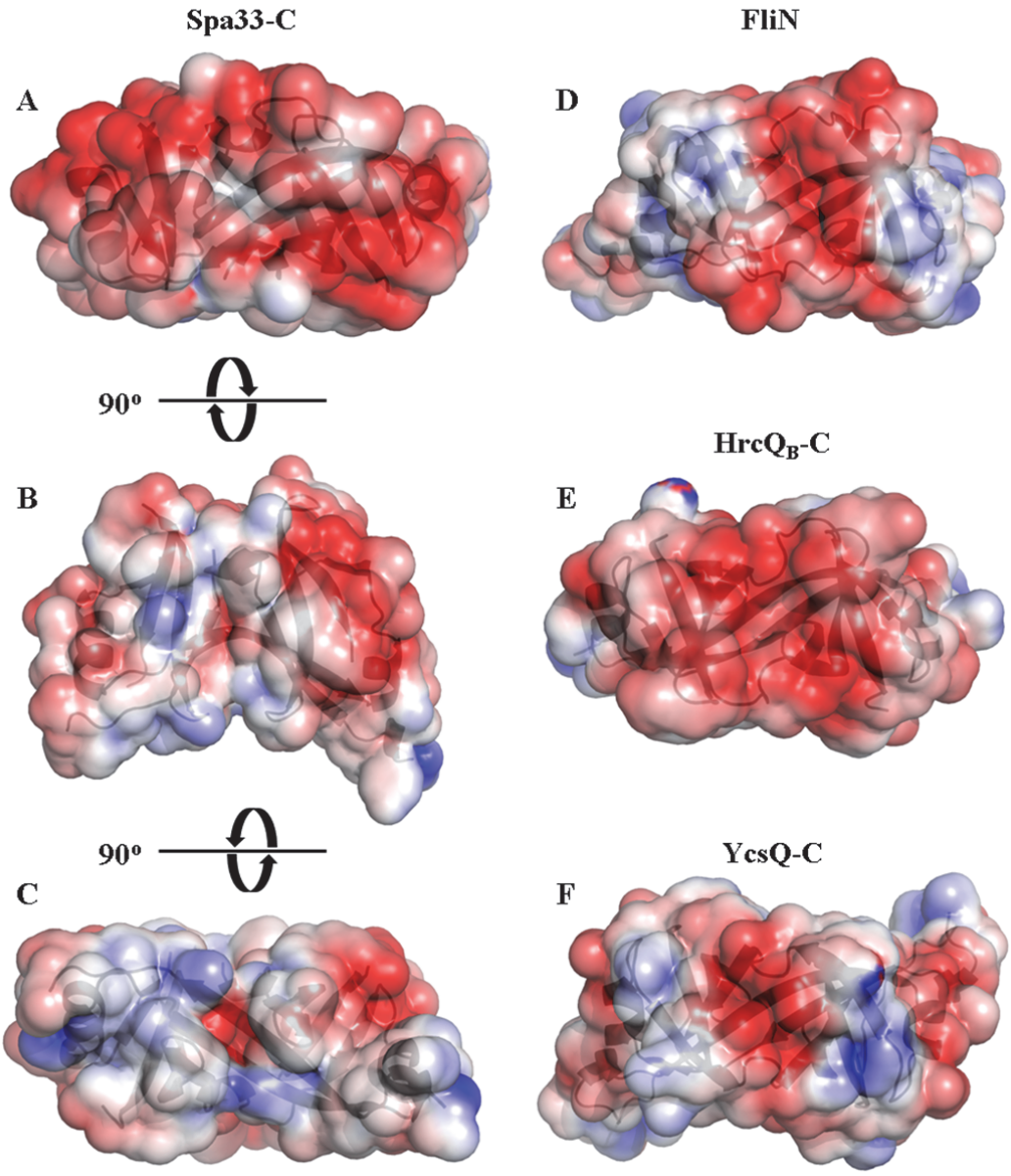


Figure 6.17 *The electrostatic surfaces of Spa33$_{208-293}$ and homologues*.

Positive (blue) and negative (red) regions are mapped on the surface representation, with the cartoon representation shown underneath. Electrostatics were calculated with the APBS plugin for PyMol (Baker *et al.*, 2001) using a physiological ion concentration of 150 mM and contoured to ± 3.

A-C The electrostatic surface of the Spa33$_{208-293}$ dimer.
A View down the two-fold symmetry axis from the concave dimer surface (**Figure 6.16A**)
B View perpendicular to the two-fold symmetry axis of the dimer (**Figure 6.16B**)
C View down the two-fold symmetry axis from the convex dimer surface (**Figure 6.16C**).

D-F The electrostatic surfaces of homologous dimers. In each case, the view is equivalent to that shown for the Spa33$_{208-293}$ dimer in **A**.

Improvement of the Crystallisation Efficiency

It was notable that 13 residues at the N-terminus of $Spa33_{208-293}$ could not be modelled into the electron density, suggesting either that the N-terminus had been proteolysed within the crystallisation condition or that it was too dynamic to make a significant contribution to high resolution scattering, as hinted at by the high C^{α} B-factors for N220-K227 (**Figure 6.13B**). This was also observed during the structure determination of YscQ-C, where the first 11 residues could not be modelled into the electron density (Bzymek *et al.*, 2012). Indeed, the close packing of the N-terminus seen within the $Spa33_{208-293}$ crystal lattice (**Figure 6.14A**) suggested any extension could inhibit crystal contact formation. Furthermore, addition of subtilisin to the $Spa33_{208-293}$ sample was seen to promote crystal formation in Optimisation Screen 1 (**Figure 6.3B**), presumably via removal of flexible termini. Therefore it was hypothesised that the slow and sporadic crystallisation of $Spa33_{208-293}$ was due to the extreme N-terminus and therefore that its removal would accelerate crystallisation. To test this, residues E208-F219 of $Spa33_{208-293}$ (**Figure 6.12E**) were deleted by site-directed mutagenesis (**Appendix 8.2.3**) to create the pET28b-*spa33*$_{220-293}$ construct. This required T_a of 55^{o}C, t_e of 6.5 min and the primers shown in **Table 6.9**.

Table 6.9 Primers used to create the pET28b-*spa33*$_{220-293}$ construct.

Forward	**Reverse**
5'- G AAA TAA TTT TGT TTA ACT TTA AGA AGG AGA TAT ACC ATG AAT TAT GAT GAT ATC AAT GTA AAA GTG GAC TTT ATT C -3'	5'- G AAT AAA GTC CAC TTT TAC ATT GAT ATC ATC ATA ATT CAT GGT ATA TCT CCT TCT TAA AGT TAA ACA AAA TTA TTT C 3'

Residues 220-293 of Spa33 ($Spa33_{220-293}$) were expressed with a LEHHHHHH C-terminal tag ($Spa33_{220-293}$-His) from pET28b-*spa33*$_{220-293}$ transformed into BL21 DE3 (pLysS) cells, as described previously for expression from B834 DE3 (pLysS) pET28b-*spa33*$_{208-293}$. $Spa33_{220-293}$ was subsequently purified in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl in an analogous manner

to unmethylated $Spa33_{208-293}$. As the $Spa33_{208-293}$ datasets used for structure determination were acquired from a crystal grown in the ProPlex screen (**Figure 6.3A**), crystallisation trials were set up using this broad screen and 10 mg/ml $Spa33_{220-293}$, as described for unmethylated and methylated $Spa33_{208-293}$. Within just two days, many crystals had grown within 1:1 drops of condition C7 (20% v/v isopropanol, 0.1 M Na Citrate (pH 5.6), 20% PEG 4000; **Figure 6.18A**) and H7 (15% v/v 2-methyl-2,4-pentanediol (MPD), 0.1 M Na Citrate (pH 5.0), 2% PEG 4000; **Figure 6.18B**). Although no crystals had grown in condition B7 (10% v/v isopropanol, 0.1 M Na HEPES (pH 7.0), 10% PEG 4000) where the highly diffracting $Spa33_{208-293}$ crystal had previously been obtained, it was mimicked by the presence of a precipitating alcohol and PEG-4000 in conditions C7 and H7, suggesting the crystals were likely to be of $Spa33_{220-293}$. The crystals shown in **Figure 6.18** were briefly soaked in cryoprotectant (1:5 v/v ethylene glycol:mother liquor), flash-frozen and stored in liquid N_2.

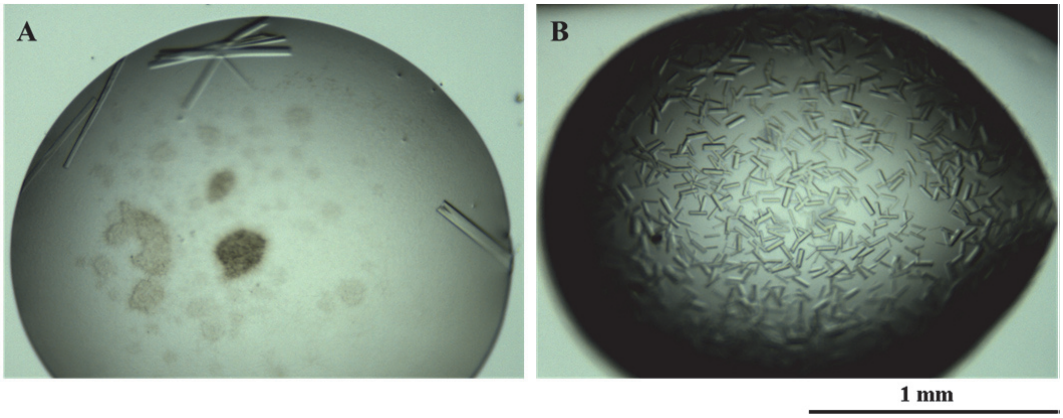


Figure 6.18 *Crystals of $Spa33_{220-293}$*. Visualisation within the drop.

A Rod-shaped crystal grown at 21°C within a 400 nl drop of 10 mg/ml $Spa33_{220-293}$ mixed 1:1 with condition C7 of the ProPlex screen (20% v/v isopropanol, 0.1 M Na Citrate (pH 5.6), 20% PEG 4000).

B Rod-shaped crystals grown at 21°C within a 400 nl drop of 10 mg/ml $Spa33_{220-293}$ mixed 1:1 with condition H7 of the ProPlex screen (15% v/v MPD, 0.1 M Na Citrate (pH 5.0), 2% PEG 4000).

Diffraction images collected by Dr Pietro Roversi at beamline I02 (Diamond Light Source, UK) showed ~4.5 Å resolution and ~3.6 Å resolution scattering for crystals shown in **Figure 6.18A** and **Figure 6.18B** respectively, indicating both were protein crystals. A dataset was collected for a 90° rotation ($\Delta\varphi = 0.3°$, 300 images) of the higher resolution crystal and the images were

processed using the -3dii command in Xia2. Indexing showed the space group to be orthorhombic with similar unit cell dimensions (**Table 6.10**), suggesting the crystals comprised $Spa33_{220-293}$ arranged in a similar lattice to that observed for $Spa33_{208-293}$ (**Figure 6.14A**). Therefore, removal of the extreme N-terminus of $Spa33_{208-293}$ had dramatically improved the efficiency of crystallisation, allowing reproducible crystals to be obtained for further work.

Table 6.10 Comparison of unit cell dimensions of $Spa33_{208-293}$ and $Spa33_{220-293}$ crystals.	
Unit cell dimensions $Spa33_{208-293}$ (Å)	a = 27.8, b = 51.3, c = 104.5
Unit cell dimensions $Spa33_{220-293}$ (Å)	a = 27.6, b = 50.8, c = 107.1

6.5 $Spa33_{220-293}$-His Interacts with the Extreme N-terminus of MxiN *in vitro*

Within the flagellar T3SS, FliN undergoes interactions important for both the assembly and normal rotation of flagella (Tang *et al.*, 1995, Irikura *et al.*, 1993). Systematic mutagenesis of the *E. coli* FliN surface demonstrated that residues within the conserved hydrophobic groove, including the strictly conserved valine (**Figure 6.1**), were essential for flagellum assembly, being required for co-purification of FliN with FliH (Paul *et al.*, 2006). Subsequent characterisation of this FliN-FliH interaction in *S. typhimurium* showed it to have a K_D of 110 nM (McMurry *et al.*, 2006) and to require W7 and W10 at the extreme N-terminus of FliH, presumably for binding within the hydrophobic groove of FliN (Minamino *et al.*, 2009). Dimeric FliH also interacts with the flagellar ATPase FliI via its C-terminal domain (González-Pedrajo *et al.*, 2002), preventing wasteful ATP consumption when flagellar proteins are not being secreted by the T3SS (Minamino & Macnab, 2000). The interaction between FliN and FliH therefore allows localisation of FliI to the C-ring (McMurry *et al.*, 2006). As FliH is also thought to promote association of FliI with the export gate components FlhA and FlhB (Minamino *et al.*, 2003) to allow initial entry of substrates (Minamino & Namba, 2008), it has

been proposed that association of the $(FliH)_2$-FliI complex with the C-ring is a prerequisite for its localisation to the export complex (McMurry *et al.*, 2006).

FliH homologues are also present within animal pathogenic T3SSs, with MxiN from *S. flexneri* exhibiting weak sequence similarity to FliH from *S. typhimurium* (**Figure 6.19**). MxiN has been found to be essential for needle formation during T3SS assembly and co-purifies with the ATPase Spa47 (Jouihri *et al.*, 2003, Johnson & Blocker, 2008). Furthermore an interaction between MxiN and Spa33 homologues has been demonstrated for *S. flexneri* (Jouihri *et al.*, 2003, Morita-Ishihara *et al.*, 2006), *S. typhimurium* (Lara-Tejero *et al.*, 2011), *Y. pestis* (Jackson & Plano, 2000) and *C. pneumoniae* (Johnson *et al.*, 2008) T3SSs. Specifically, residues 217-293 at the C-terminus of Spa33 were found to be sufficient for binding to MxiN (Morita-Ishihara *et al.*, 2006). These data indicated MxiN has equivalent interaction partners to FliH and so could be functioning in an analogous manner within the T3SS. Therefore, the putative interaction between MxiN and Spa33-C was probed further.

```
FliH     1 MSNELPWQVWTPDDLAPPPETFVPVEADNVTLTDDTPEPELTAEQQLEQE     50
           |. ....|..|...........|.::.....|.....:.:..:::::....
MxiN     1 MK-VCNMQKGTLPVSRHHAYDGVVIKRIEKELCKTIKDRDTESKKKAICV     49

FliH    51 LAQLKIQAHEQGYNAGLAEGRQKGHAQGYQEGLAQGLEQGQAQAQTQQAP    100
           :.....:|.....:| :.:|.|.|....: |.:...:.:.:.:....:..
MxiN    50 IKDATKKAESLRIDA-VCDGYQIGIQTAF-EHIIDYICEWKLKQNENRRN     97

FliH   101 IHARMQQLVSEFQNTLDALDSVIASRLMQMALEAARQVIGQTPAVDNSAL    150
           |...:..|:||     :..|..|.|.|::..|.:.|..:.:...|.....
MxiN    98 IEDYITSLLSE-----NLHDERIISTLLEQWLSSLRNTVTELKVVLPKCN    142

FliH   151 IKQIQQLLQQEPLFSGKPQLRVHPDDLQRVEEMLGATLSLHGWRLRGDPT    200
           :...::|  :..|...:.:::    :.:..|............:...|.
MxiN   143 LALRKKL--ELDLHKYRSDVKI----ILKYSEGNNYIFCSGNQVVEFSPQ    186

FliH   201 LHHGGCKVSADEGDLDASVATRWQELC----RLAAPGVL-------        235
           ....|.|:...| .|..:....::||.    |..|..:|
MxiN   187 DVISGVKIELAE-KLTKNDKKYFKELAHKKLRQIAEDLLKENPVND        231
```

Figure 6.19 *Alignment of S. flexneri MxiN with S. typhimurium FliH.*

Global pairwise alignment of sequences using the EMBOSS-Needle server (EMBL-EBI). The alignment shown has 14.2% identity, 32.1% similarity and 10.6% gaps. Dashes, colons and full-stops represent identical, conserved and semi-conserved residues respectively. The Trp residues essential for the FliH interaction with FliH are shown in red. The sequence used to synthesise the $MxiN_{1-18}$ peptide is shown in bold, with residues with a hydrophobicity >3.0 on the Cornette scale in green and positively charged residues in blue.

As residues 2-10 of *S. typhimurium* FliH have been implicated in binding the hydrophobic groove of FliN (Minamino *et al.*, 2009), it was hypothesised that the extreme N-terminus of MxiN would mediate an analogous interaction with Spa33-C. Residues equivalent to W7 and W10 of FliH, shown to be crucial for the interaction with FliN (Minamino *et al.*, 2009), were absent from MxiN, with only 16% identity between the N-terminal 18 residues of both proteins (**Figure 6.19**). However, MxiN was still found to have a reasonable distribution of hydrophobic residues at the N-terminus (**Figure 6.19, green**), which could bind the hydrophobic groove of Spa33-C, performing the equivalent role to the tryptophan residues in FliH. Furthermore, lysine and arginine residues present at the extreme N-terminus (**Figure 6.19, blue**) could undergo additional electrostatic interactions with the negatively charge surface surrounding the hydrophobic groove (**Figure 6.17A**). As an interaction between the N-terminus of MxiN and Spa33-C seemed plausible, a peptide comprising residues 1-18 of MxiN ($MxiN_{1-18}$; **Figure 6.19, bold**) was synthesised (Severn Biotech) for use in interaction studies.

With a view to performing NMR titrations with $MxiN_{1-18}$, the optimal Spa33-C construct was chosen based on the quality of its $^{1}H,^{15}N$-HSQC spectrum. As evidence from crystallisation trials had suggested the N-terminus of $Spa33_{208-293}$ was highly flexible, these constructs would likely lead to more overlapped spectra and were therefore not tested. ^{15}N-labelled $Spa33_{220-293}$-His was expressed in BL21 DE3 (pLysS) pET28b-*spa33*$_{220-293}$ cells, as described in previous chapters. Both ^{15}N-$Spa33_{220-293}$ and ^{15}N-$Spa33_{220-293}$-His were subsequently purified in 100 mM Tris-HCl (pH 7.5), 150 mM NaCl, with the His-tagged construct being purified as for $Spa33_{208-293}$ with omission of the dialysis and carboxypeptidase A digest steps. A $^{1}H,^{15}N$-HSQC spectrum acquired with 788 μM ^{15}N-$Spa33_{220-293}$ had very few peaks above the noise (**Figure 6.20A**). As the protein was confirmed to be 100% ^{15}N-labelled by mass spectrometry and its concentration had not changed during the course of the experiment, the spectrum indicated that ^{15}N-$Spa33_{220-293}$ was tumbling slowly in solution and therefore was likely to be forming high molecular weight complexes. In contrast, the $^{1}H,^{15}N$-HSQC spectrum for 50 μM ^{15}N-$Spa33_{220-293}$-His showed good peak dispersion in both dimensions, indicating the protein was most likely

tumbling as the expected 19.5 kDa dimer (**Figure 6.20B**). Subsequently 71 putative backbone amide and 6 side-chain amide peaks were picked, although it was plausible that a further pair of side-chain amide peaks was present in an overlapping region of the spectrum (**Figure 6.20B, asterisk**). As Spa33$_{220\text{-}293}$ has 83 residues, including 7 asparagine residues, the number of peaks suggested that equivalent residues within chain A and chain B of the dimer were experiencing

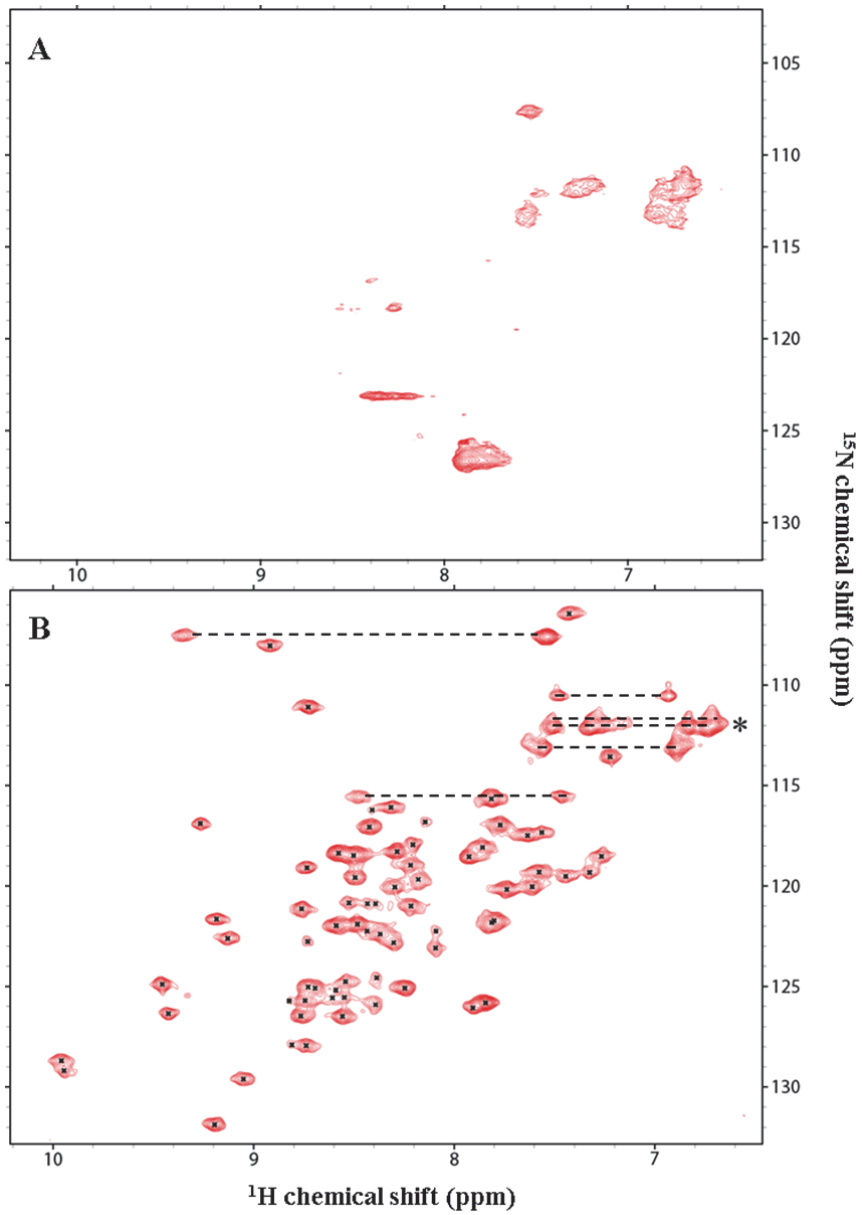


Figure 6.20 *^{1}H,^{15}N-HSQC spectra of ^{15}N-labelled Spa33-C constructs*. Acquired in 95 mM Tris-HCl (pH 7.5), 142 mM NaCl, 5% D_2O. Spectra were processed using NMRPipe (Delaglio *et al.*, 1995).

A Spectrum for 788 µM ^{15}N-Spa33$_{220\text{-}293}$.

B Spectrum for 50 µM ^{15}N-Spa33$_{220\text{-}293}$-His showing putative peaks for backbone (black dots) and Asn side-chain (dashed lines) amides. The asterisk denotes a region of overlapping peaks which could correspond to 2 or 3 Asn side-chain amides.

similar chemical environments. Therefore, asymmetry within the ^{15}N-Spa33$_{220\text{-}293}$-His dimer did not appear to be significant, suggesting structural differences between chain A and chain B seen in the crystal structure (**Figure 6.13A**) were more likely due to flexibility of these regions (**Figure 6.13B**). As a result, ^{15}N-Spa33$_{220\text{-}293}$-His provided an appropriate NMR sample for investigation of the putative interaction with MxiN$_{1\text{-}18}$.

Addition of 1.5 mM MxiN$_{1\text{-}18}$ to a 50 µM sample of ^{15}N-Spa33$_{220\text{-}293}$-His resulted in chemical shift perturbations of at least nine backbone amide peaks (**Figure 6.21, black arrows**), four of which were also evident in the presence of 200 µM MxiN$_{1\text{-}18}$ to a lesser extent (**Figure 6.21, asterisks**). As pH indicator paper was used to confirm there was no pH change associated with addition of peptide, the changes in peak positions were likely due to an *in vitro* interaction between ^{15}N-Spa33220$_{\text{-}293}$-His and MxiN$_{1\text{-}18}$.

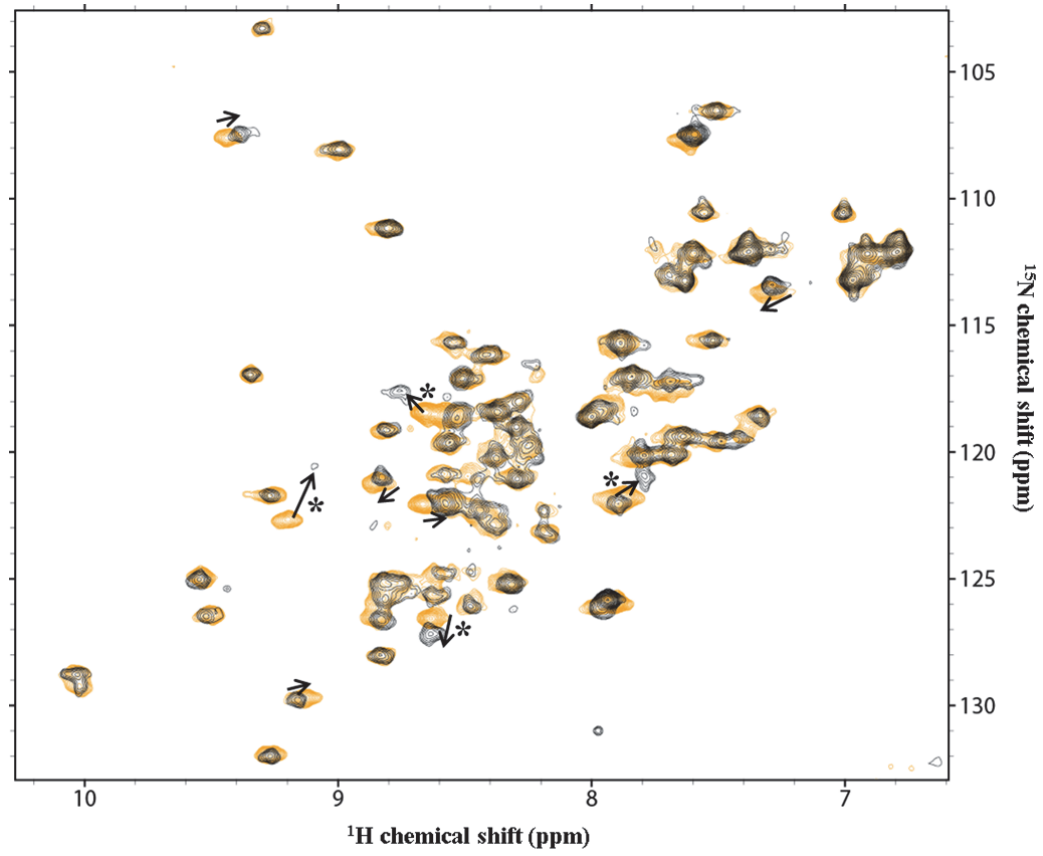


Figure 6.21 *Titration of ^{15}N-Spa33$_{220\text{-}293}$-His with 1.5 mM MixN$_{1\text{-}18}$.*

Overlay of ^{1}H,^{15}N-HSQC spectra of 50 µM ^{15}N-Spa33$_{220\text{-}293}$-His obtained before (orange) and after (black) addition of 1.5 mM MxiN$_{1\text{-}18}$, acquired in 95 mM Tris-HCl (pH 7.5), 142 mM NaCl, 5% D_2O. Spectra were processed using NMRPipe (Delaglio *et al.*, 1995). The arrows show examples of chemical shift perturbations due to the peptide, with asterisks denoting those that were evident after titration of 50 µM ^{15}N-Spa33$_{220\text{-}293}$-His with 200 µM MxiN$_{1\text{-}18}$.

Although these NMR titrations provided preliminary evidence that an analogous FliH-FliN interaction was present in the *Shigella* T3SS, further work is needed to characterise the nature of the interaction between MxiN and Spa33-C. For example, full characterisation of the $MxiN_{1-18}$ binding site within Spa33-C could be achieved by assignment of the $^{1}H,^{15}N$-HSQC for ^{15}N-$Spa33_{220-293}$-His (**Figure 6.20B**). In order to mimic the putative FliN interaction site, $MxiN_{1-18}$ would be expected to bind the conserved hydrophobic groove identified on the $Spa33_{208-293}$ dimer surface (**Figure 6.16F**). Therefore, the specificity of the $MxiN_{1-18}$ interaction could be more simply probed by repeating the NMR titration with a V276A mutant of ^{15}N-$Spa33_{220-293}$; the prediction is that a decreased interaction affinity would be observed, as was achieved through mutation of the conserved valine in *S. typhimurium* FliN (McMurry *et al.*, 2006).

The occurrence of significant peak shifts only after the addition of 1.5 mM $MxiN_{1-18}$ indicated a much lower K_D for the interaction with Spa33-C than the 110 nM observed for *S. typhimurium* FliH binding to FliN (McMurry *et al.*, 2006). However, this affinity was obtained via quantitative immunoblotting of full-length FliH, suggesting other regions of the molecule could be contributing to the interaction with FliN. Athough FliH does not appear to interact productively with FliN in the absence of the first 10 residues (Minamino *et al.*, 2009), the interaction with the extreme N-terminus of FliH in isolation has never been quantified. Furthermore, the interaction with FliI requires the dimerisation domain of FliH (González-Pedrajo *et al.*, 2002), suggesting FliH is likely to interact with FliN as a dimer. Given that the hydrophobic groove lies at the $Spa33_{208-293}$ dimer interface and exhibits two-fold rotational symmetry (**Figure 6.16E**), it is plausible that the Spa33-C dimer could bind two MxiN N-termini simultaneously, resulting in an increased interaction affinity. Therefore, binding of full-length MxiN to Spa33-C would have to be investigated to recapitulate the *in vivo* interaction.

Given the reliability of $Spa33_{220-293}$ crystallisation discussed previously, this construct could be amenable to co-crystallisation with $MxiN_{1-18}$. If successful, structure determination of the complex by MR with $Spa33_{208-293}$ would allow detailed characterisation of the MxiN-Spa33-C interaction interface. However, the formation of a $Spa33_{220-293}$-$MxiN_{1-18}$ co-crystal may not be

trivial. Firstly, the low interaction affinity shown by the NMR titration indicated that a >10-fold molar excess of $MxiN_{1-18}$ would need to be present to achieve the high occupancy of the $Spa33_{220-293}$ binding site required for the peptide to make a significant contribution to the scattering. Secondly, as the conserved hydrophobic groove was in a reasonably packed region of the $Spa33_{220-293}$ crystal lattice (**Figure 6.14A**), one or two copies of $MxiN_{1-18}$ may not be accommodated in the current crystal form. Therefore, a $Spa33_{220-293}$-$MxiN_{1-18}$ complex may need to form novel crystal contacts and pack in an alternative space group to $Spa33_{220-293}$.

Finally, the oligomerisation of the $Spa33_{208-293}$ construct indicated by the $^{1}H,^{15}N$-HSQC (**Figure 6.20A**) was interesting, suggesting Spa33-C may have the propensity to form higher order structures. Although *S. typhimurium* SpaO has been found to form a large molecular weight complex (Lara-Tejero *et al.*, 2011), a cytoplasmic C-ring similar to that found in the flagellar T3SS has never been directly visualised in translocation-associated T3SSs. As both FliN (Brown *et al.*, 2005, Paul *et al.*, 2006) and $HrcQ_B$-C (Fadouloglou *et al.*, 2004, Fadouloglou *et al.*, 2009) have been observed to form dimers of dimers in solution, it was plausible that Spa33 undergoes Spa33-C mediated oligomerisation within the *S. flexneri* T3SS C-ring. With a view to gaining an insight into the assembly of the putative C-ring in *S. flexneri*, the oligomerisation of both Spa33 and Spa33-C was explored, as outlined in the next chapter.

6.6 Perspectives

The 2.3 Å crystal structure of $Spa33_{208-220}$ was determined, showing the protein forms a highly intertwined dimer that closely resembled homologous structures. The concave surface of the $Spa33_{208-220}$ dimer had a conserved hydrophobic groove and negative electrostatic potential, implicating it as an important interaction interface. An NMR titration subsequently provided preliminary evidence for an interaction between Spa33-C and the extreme N-terminus of MxiN. Furthermore, the crystallisation efficiency was drastically improved by an N-terminal truncation of $Spa33_{208-220}$, providing a construct amenable to co-crystallisation trials with $MxiN_{1-18}$.

7. Characterisation of the *Shigella* T3SS C-ring

The C-ring at the base of flagellar T3SSs has been extensively characterised, with EM visualisation of the intact complex from *S. typhimurium* showing it to be a ring structure of ~40 nm diameter and 32-36 fold symmetry (Thomas *et al.*, 2006). FliM and FliN comprise most of the C-ring, with current evidence suggesting a FliM-$(FliN)_4$ complex is the basic building block for assembly (Brown *et al.*, 2005, Sarkar *et al.*, 2010b). In contrast, intact C-rings have never been isolated or directly visualised in translocation-associated T3SSs, although low resolution EM analysis of osmotically-shocked *S. flexneri* cells identified a large bulb beneath the NC with dimensions closely matching the flagellar C-ring (Blocker *et al.*, 1999). In addition, the $HrcQ_A$ and $HrcQ_B$ proteins from the plant pathogen *P. syringae* are homologues of FliM and FliN respectively and appear to interact in an analogous manner (Fadouloglou *et al.*, 2004), indicating C-ring architecture may also be conserved. Intriguingly, T3SSs associated with animal pathogens possess a single protein that is both an equivalent size and homologous to $FliM/HrcQ_A$, particularly at the C-terminus (**Figure 7.1**). However, the C-terminal ~100 residues display higher sequence homology with $FliN/HrcQ_B$ (**Figure 7.1, arrow**) and share three strictly conserved residues (**Figure 7.1, orange**). Furthermore, the highly intertwined *S. flexneri* Spa33-C (previous chapter) and *Y. pseudotuberculosis* YscQ-C (Bzymek *et al.*, 2012) dimers are structurally similar to *T. maritima* FliN (Brown *et al.*, 2005) and *P. syringae* $HrcQ_B$-C (Fadouloglou *et al.*, 2004)(**Figure 6.15**), implicating the proteins in similar roles. Indeed, an *in vitro* interaction between Spa33-C and the extreme N-terminus of MxiN (**Figure 6.21**) mimicked the observed interaction between FliN and FliH in the flagellar T3SS (Minamino *et al.*, 2009). Given these sequence, structural and functional similarities, proteins such as *S. flexneri* Spa33 could be predicted to fulfil the roles of both FliM and FliN within the T3SS. In particular, it was unknown how Spa33 alone could form a flagellar-like C-ring. Therefore, *S. flexneri* Spa33 was purified and characterised with a view to understanding its intermolecular interactions, as discussed in this chapter.

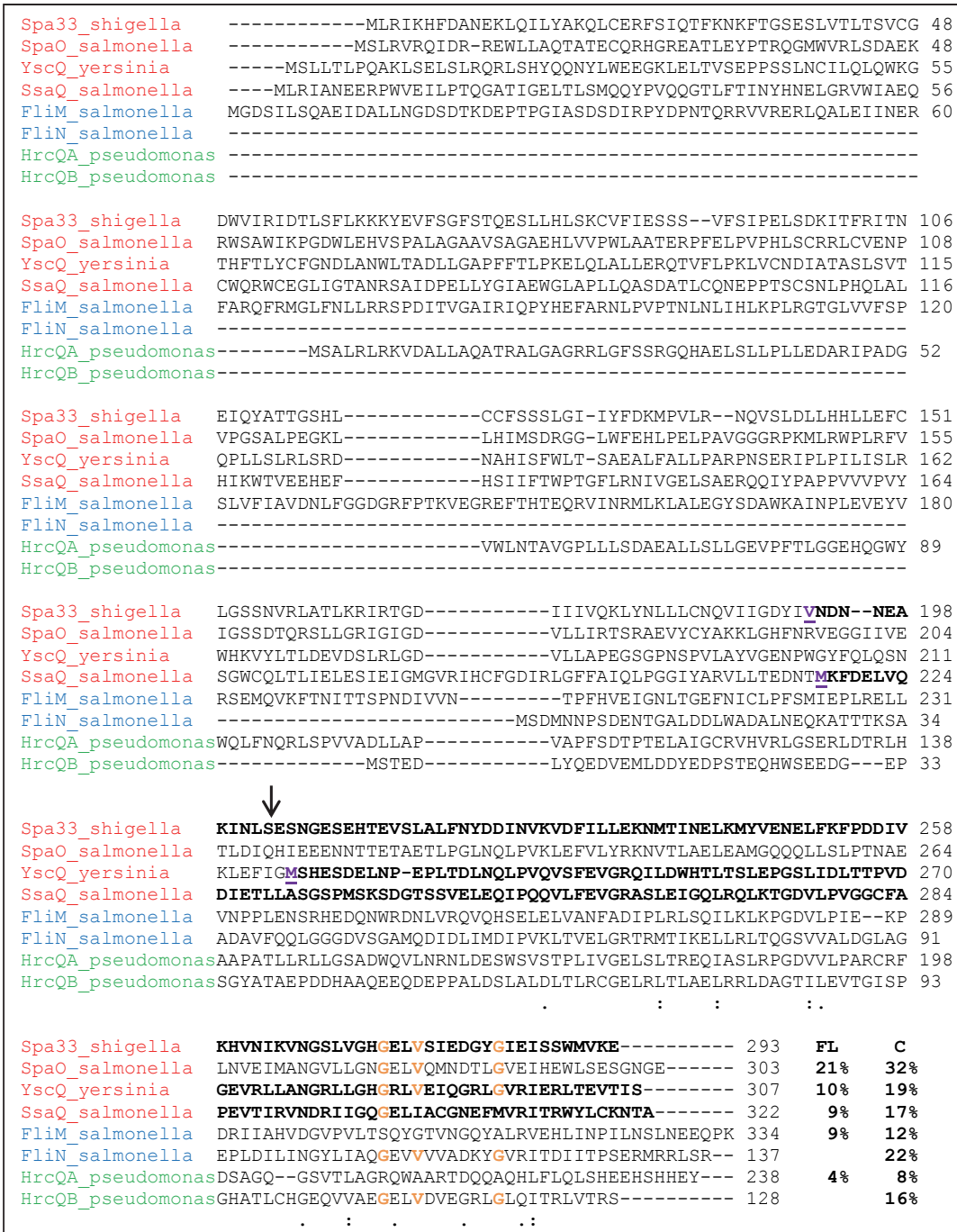


Figure 7.1 *Sequence alignments of Spa33 homologues*.

Sequence alignments of Spa33_shigella (*S. flexneri*), SpaO_salmonella (*S. typhimurium*, SPI1), YscQ_yersinia (*Y. pseudotuberculosis*), SsaQ_salmonella (*S. typhimurium*, SPI2), FliM_salmonella (*S. typhimurium*), FliN_salmonella (*S. typhimurium*), HrcQA-C_pseudomonas (*P. syringae*) and HrcQB-C_pseudomonas (*P. syringae*) were performed using ClustalW2 (Chenna *et al.*, 2003), with identical, conserved and semi-conserved residues being indicated by asterisks, colons and full-stops respectively. Sequence numbers are shown relative to the full-length protein. The pairwise identity with the full-length Spa33 sequence (FL) and with the C-terminal Spa33 sequence (C; defined as sequence downstream of the arrow) are indicated in bold after each sequence. Invariant residues identified from alignment of $Spa33_{208\text{-}293}$ with FliN/$HrcQ_B$ homologues (**Figure 6.1**) are shown in **orange**. Identified alternative translation products are shown in **bold**, with their start side in **purple**. Proteins from animal-associated, flagellar and plant-associated T3SSs are shown in red, blue and green respectively.

7.1 Spa33$_{192-293}$ is an Alternative Translation Product of Spa33

Purification of Spa33

Full-length Spa33 with a MGSSHHHHHHSSGLVPRGSH N-terminal tag (His-Spa33) was expressed from a pET28b-*spa33* construct (donated by J. Deane) in B834 DE3 (pLysS) cells, as described for expression of Spa33$_{208-293}$-His in the previous chapter. SDS-PAGE analysis confirmed overexpression of His-Spa33 at the expected molecular weight (35.6 kDa) in the final cultures (**Figure 7.2**).

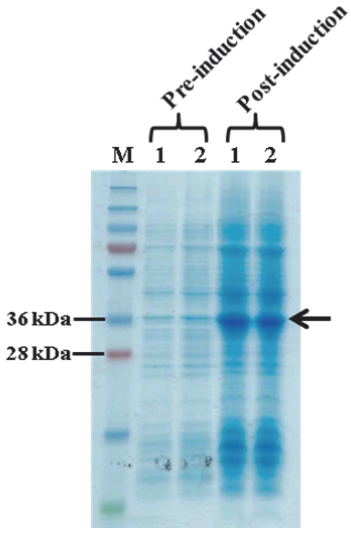


Figure 7.2 *15% SDS-PA gel showing over-expression of His-Spa33 in B834 (DE3) pLysS.*

Samples taken from repeat cultures 1 and 2 pre- and post-induction with 1 mM IPTG are labelled. The arrow indicates the band corresponding to His-Spa33. M: PageRuler Plus Prestained Protein Ladder (Fermentas).

B834 DE3 (pLysS) pET28b-*spa33* cells were resuspended in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl and supplemented with lysozyme, DNase I and an EDTA-free protease inhibitor. Cells were subsequently lysed, fractionated and His-Spa33 purified from the soluble fraction by Ni-affinity chromatography, as described for the purification of MxiG-N$_{1-126}$ in Chapter 2. His-Spa33 was seen to elute in the first fraction from the Ni-cartridge, however several contaminants were present (**Figure 7.3A, asterisk**). In an attempt to separate His-Spa33 from these variously sized proteins, this fraction was passed through a HiLoad 16/60 Superdex 200 pg column pre-equilibrated in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Although SDS-PAGE analysis of eluted fractions showed that some species had been successfully separated, a high

and low molecular weight contaminant had still eluted in fractions containing His-Spa33 (**Figure 7.3A**). Fractions taken across the main peak in the elution profile (**Figure 7.3B**) were subsequently pooled and concentrated using an Amicon Ultra centrifugal device with a 30 kDa cut-off (Millipore), resulting in simultaneous concentration of the contaminating species (**Figure 7.4**).

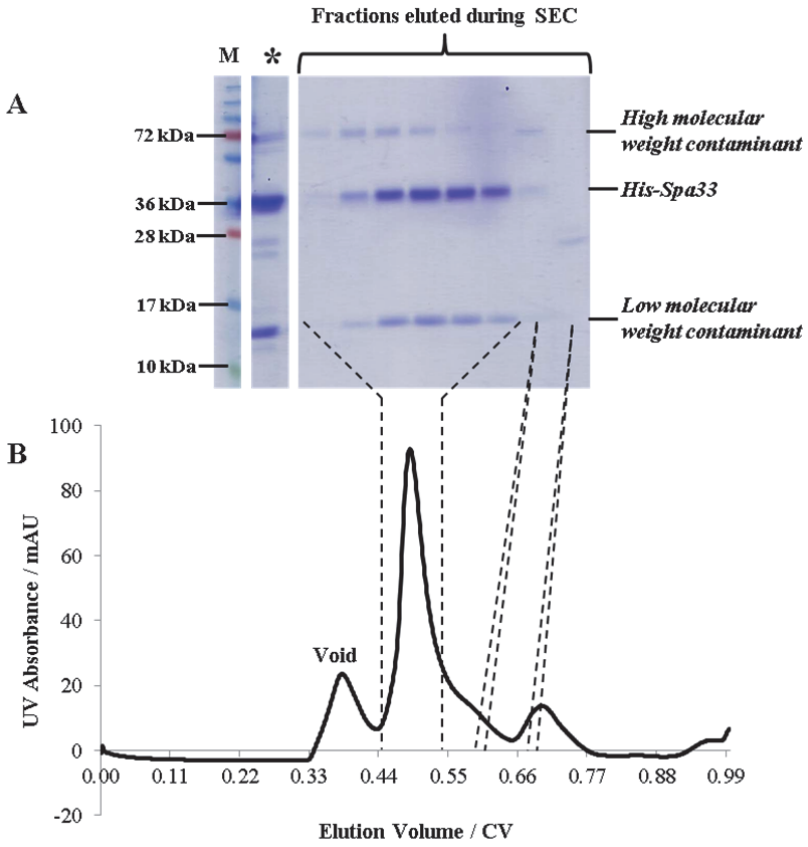


Figure 7.3 *Co-purification of His-Spa33 with contaminants.*

A 15% SDS-PA gel showing eluted fractions containing His-Spa33 and contaminants from Ni-affinity and SEC. The asterisk shows a sample taken from fraction 1 eluted with 300 mM imidazole from a Ni-NTA cartridge. Samples taken during SEC are shown relative to the position of the fraction in the His-Spa33 elution profile. M: PageRuler Plus Prestained Protein Ladder (Fermentas).

B Elution profile for His-Spa33 from a HiLoad 16/60 Superdex 200 pg column. Elution with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl.

Identification of Contaminating Fragments

In order to characterise and optimise the purification protocol for His-Spa33, the high and low molecular weight contaminants had to be identified. Given that His-Spa33 had seven cysteine residues and no reducing agent was present in the buffer, it was questioned whether the high molecular weight contaminant was a disulphide-bonded His-Spa33 dimer. However, SDS-PAGE analysis indicated that this was unlikely to be the case, as the abundance of this species did not change in the presence of 20 mM DTT (**Figure 7.4**). Subsequently, the gel bands corresponding to the high and low molecular weight contaminants were excised and subjected to trypsin digestion and mass spectrometry identification (Central Proteomics Facility, University of Oxford). The high molecular weight contaminant was unambiguously identified as the *E. coli* formyl transferase YfbG (mascot score = 2058), a 74.9 kDa protein with histidine clusters that afford natural affinity for Ni^{2+} resin (Bolanos-Garcia & Davies, 2006). In contrast, there were several candidates for the low molecular weight contaminant, most significantly YfbG (mascot score = 370) and Spa33 (mascot score = 240). Given that the elution of the low and high molecular weight contaminants was not exactly coincident (**Figure 7.3A**), it was predicted that they were not different fragments of the same species and therefore that the low molecular weight contaminant was more likely to be a fragment of Spa33.

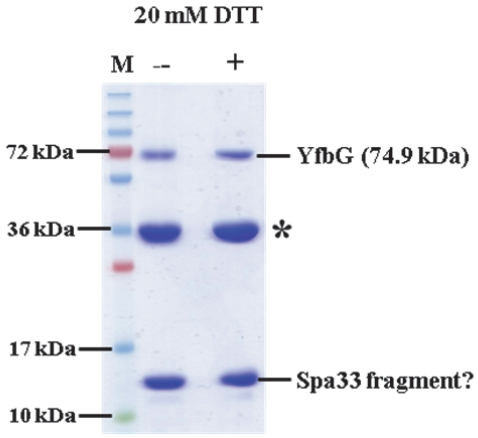


Figure 7.4 *Identification of the contaminants from His-Spa33 purification*.

15% SDS-PA gel comparing non-reduced and reduced samples of concentrated His-Spa33 after SEC. Either no DTT (--) or 20 mM DTT (+) added to the 1 x loading dye prior to sample boiling. The asterisk denotes the band for His-Spa33. The identification of the contaminants as suggested by mass spectrometry is shown. M: PageRuler Plus Prestained Protein Ladder (Fermentas).

The nature of this putative Spa33 fragment was then probed further. Given the low molecular weight of the fragment, its retention with His-Spa33 during SEC and concentrating with a 30 kDa cut-off implied that it was forming a complex with His-Spa33. As Spa33-C constructs were shown to undergo dimerisation and perhaps further oligomerisation in the previous chapter, it was hypothesised that His-Spa33 was most likely forming a complex with a C-terminal fragment of the protein. Therefore, purified $Spa33_{208-293}$ (protocol in previous chapter) in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10% glycerol was used as an antigen both for the production of polyclonal α-Spa33-C antibodies in rabbit hosts and their subsequent affinity purification (Eurogentec). This allowed the concentrated SEC fractions to be separated on a 15% SDS-PA gel and analysed by western blotting (**Appendix 8.2.7**), using a 1:5000 dilution of the α-Spa33-C primary antibody, a 1:2500 dilution of the anti-rabbit IgG HRP-conjugated secondary antibody (Promega) and a 4 s exposure to the Amersham Hyperfilm ECL (GE Healthcare). The blot shown in **Figure 7.5** demonstrated that α-Spa33-C could detect His-Spa33, verifying both the identity of the gel band and the specificity of the antibody. In addition, α-Spa33-C detected the band corresponding to the low molecular weight contaminant, indicating His-Spa33 had co-purified with a fragment encompassing residues 208-293 of Spa33. To further characterise this C-terminal Spa33 fragment, the concentrated SEC fractions were analysed by mass spectrometry, as outlined in Chapter 2. This showed a single low molecular weight species of 11608 ± 4 Da, which corresponded best with residues 192-293 of Spa33 ($Spa33_{192-293}$) when compared to the sequence.

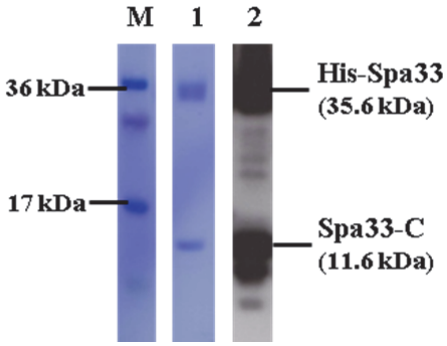


Figure 7.5 *Identification of Spa33-C.*

Western blot of the concentrated His-Spa33 after SEC with α-Spa33-C. M: PageRuler Plus Prestained Protein Ladder (Fermentas), 1: Coomassie-stained sample on the the Amersham Hybond-P membrane and 2: corresponding section from the Amersham Hyperfilm ECL. His-Spa33 and Spa33-C bands detected by α-Spa33-C are detected.

There were two likely origins of $Spa33_{192-293}$: the fragment was a product either of His-Spa33 proteolysis or of an alternative translation start site within the *spa33* gene. Given that the observed molecular weight correlated more closely with the presence of a methionine (11606 Da) at the N-terminus of $Spa33_{192-293}$ as opposed to a valine (11574 Da) and a putative purine-rich ribosome binding site (RBS) was present in the upstream DNA sequence (**Figure 7.6, blue**), the latter seemed more likely. A timely publication revealed the C-terminus of SsaQ (SsaQ-C) from *S. typhimurium* SPI2 to be an alternative translation product of the *ssaQ* gene *in vivo* and to co-immunoprecipitate with full-length SsaQ (Yu *et al.*, 2011). Subsequently, alternative translation of the C-terminus of YscQ (YscQ-C) was found to occur in *Y. pseudotuberculosis*, leading to the formation of a YscQ-C/YscQ complex *in vivo* (Bzymek *et al.*, 2012). Comparison of the Spa33 sequence to those of SsaQ and YscQ showed 17% and 19% identity respectively at their C-termini, with the translation start site of Ysc-Q being closely aligned to Spa33 V192 (**Figure 7.1**). Therefore, $Spa33_{192-293}$ was likely to be a biologically relevant alternative translation product of *spa33* mRNA, utilising a GTG start codon (**Figure 7.6, red**) as for several other genes in the *Shigella* virulence plasmid (Penno *et al.*, 2005, Penno *et al.*, 2006). Furthermore, $Spa33_{192-293}$ was only ~28 residues longer than the C-terminal fragment characterised by X-ray crystallography (previous chapter), providing an appropriate structural model for further analysis.

```
ggggattatattgtgaatgataataatgaggcaaaaattaatctgtcagaaagtaatggtgagtca
            M  N  D  N  N  E  A  K  I  N  L  S  E  S  N  G  E  S
gaacacacagaagtttctttggcattattcaattatgatgatatcaatgtaaaagtggactttatt
 E  H  T  E  V  S  L  A  L  F  N  Y  D  D  I  N  V  K  V  D  F  I
cttttagaaaaaaatatgacaatcaatgaactaaaaatgtatgtagaaaacgaattattcaagttt
 L  L  E  K  N  M  T  I  N  E  L  K  M  Y  V  E  N  E  L  F  K  F
cccgatgacatagttaaacatgtaaatattaaagtaaatggttctttggttgggcatggggaactt
 P  D  D  I  V  K  H  V  N  I  K  V  N  G  S  L  V  G  H  G  E  L
gtttctattgaggatggttatggtatcgagattagttcttggatggtaaaggagta
V  S  I  E  D  G  Y  G  I  E  I  S  S  W  M  V  K  E  -
```

Figure 7.6 *The alternative translation product of the spa33 gene*.

The protein sequence of $Spa33_{192-293}$ is shown relative to the coding strand, with the RBS (blue) and GTG initiation codon (red) highlighted in the DNA sequence.

As mutation of the alternative initiation site indicated that the function of the *S. typhimurium* SPI2 T3SS was not fully impaired in the absence of SsaQ-C and could be fully restored by excess full-length SsaQ, a role for SsaQ-C as a chaperone for SsaQ was proposed (Yu *et al.*, 2011). In contrast, YscQ-C was found to be essential for secretion by the *Y. pseudotuberculosis* T3SS, suggesting a more integral role in the apparatus (Bzymek *et al.*, 2012). Indeed, size and sequence similarities hinted that Spa33 and $Spa33_{192\text{-}293}$ could be the counterparts of FliM and FliN respectively (**Figure 7.1**), suggesting how a single *spa33* gene in *S. flexneri* could fulfil an equivalent role to two genes in flagellated bacteria. This would pose interesting questions on the evolutionary relationship between flagellar and translocation-associated T3SSs, which are thought to have evolved independently from a common ancestor (Gophna *et al.*, 2003). Interestingly, *Chlamydia* spp., which some believe possess the most ancient example of a translocation-associated T3SS, only has a single FliM/FliN homologue encoded in the genome (Kim, 2001). This could suggest that *fliM* and *fliN* arose by a gene duplication event, with sequence divergence driven by their differential functionality in the flagellar T3SS. Indeed, whilst *S. typhimurium* FliM and FliN possess 11% sequence identity at their C-termini, the absence of the absolutely conserved residues of FliN homologues from FliM homologues could reflect the alternative roles for the domains (**Figure 7.1**). Intriguingly the C-ring of *Bacillus subtilus* lacks FliN, utilising instead the FliY protein that is equivalent in size to FliM and has a 33% sequence identity to the N-terminus of *E. coli* FliM and a 52% sequence identity to the extreme C-terminus of *E. coli* FliN. Furthermore, FliY could complement the assembly defect in a *S. typhimurium fliN* null mutant, suggesting an analogous role. (Bischoff & Ordal, 1992) It could therefore be speculated that *B. subtilis fliM* and *fliY* are duplicated genes, with FliN arising from a truncation of a FliY-like protein. The key to understanding the evolution of genes encoding C-ring constituents could lie in exploring the evolutionary history of the plant-pathogenic T3SSs further, as they display an intermediate situation: *Erwinia*, *Xanthomonas* and *Ralstonia* spp. have a single *hrcQ* gene, whilst species such as *P. syringae* have separate $hrcQ_A$ and $hrcQ_B$ genes.

Further investigation was therefore required to determine the nature of this His-Spa33/Spa33$_{192-293}$ complex, to ascertain whether it was likely to be the structural counterpart of a FliM/FliN complex within the putative C-ring of *S. flexneri*.

Optimisation of the Purification Protocol

As the majority of His-Spa33 was seen to be incorporated into inclusion bodies upon expression (**Figure 7.7A, lane 4**), 6 x 1 l B834 DE3 (pLysS) pET28b-*spa33* cells were grown to increase the yield of the His-Spa33/Spa33$_{192-293}$ complex purified from the soluble fraction (**Figure 7.7A, lane 3**). Furthermore, elution of YsbG during SEC was not exactly coincident with the His-Spa33/Spa33$_{192-293}$ complex (**Figure 7.3A**), suggesting it was not associated with the complex and could likely be separated from it. As it has been shown previously that YsbG can be eluted from Ni^{2+} resin with 55-80 mM imidazole (Bolanos-Garcia & Davies, 2006), more stringent wash-steps during Ni-affinity chromatography were carried out to remove the contaminant; the Ni-NTA cartridge was attached to a ÄKTApurifier (GE Healthcare) system and washed with 5 CV each of 50 mM imidazole and 125 mM imidazole added to lysis buffer, before elution with 300 mM imidazole. The elution profile showed that different proteins had been eluted with the different imidazole concentrations (**Figure 7.7A**), with SDS-PAGE analysis of the corresponding fractions showing selective elution of YsbG and other contaminants with 125 mM imidazole (**Figure 7.7A, arrow & bracket 6**) and the His-Spa33/Spa33$_{192-293}$ complex with 300 mM imidazole (**Figure 7.7A, bracket 7**). Subsequent SEC of the pooled fractions from the 300 mM imidazole elution yielded a single main elution peak (**Figure 7.7B**), which corresponded to the His-Spa33/Spa33$_{192-293}$ complex (**Figure 7.7B, inset**). SDS-PAGE analysis of these pooled fractions concentrated 10-fold demonstrated the high purity of the His-Spa33/Spa33$_{192-293}$ complex, indicating that YsbG was no longer a detectable contaminant and an optimal protocol for purification had been designed (**Figure 7.7B, asterisk**).

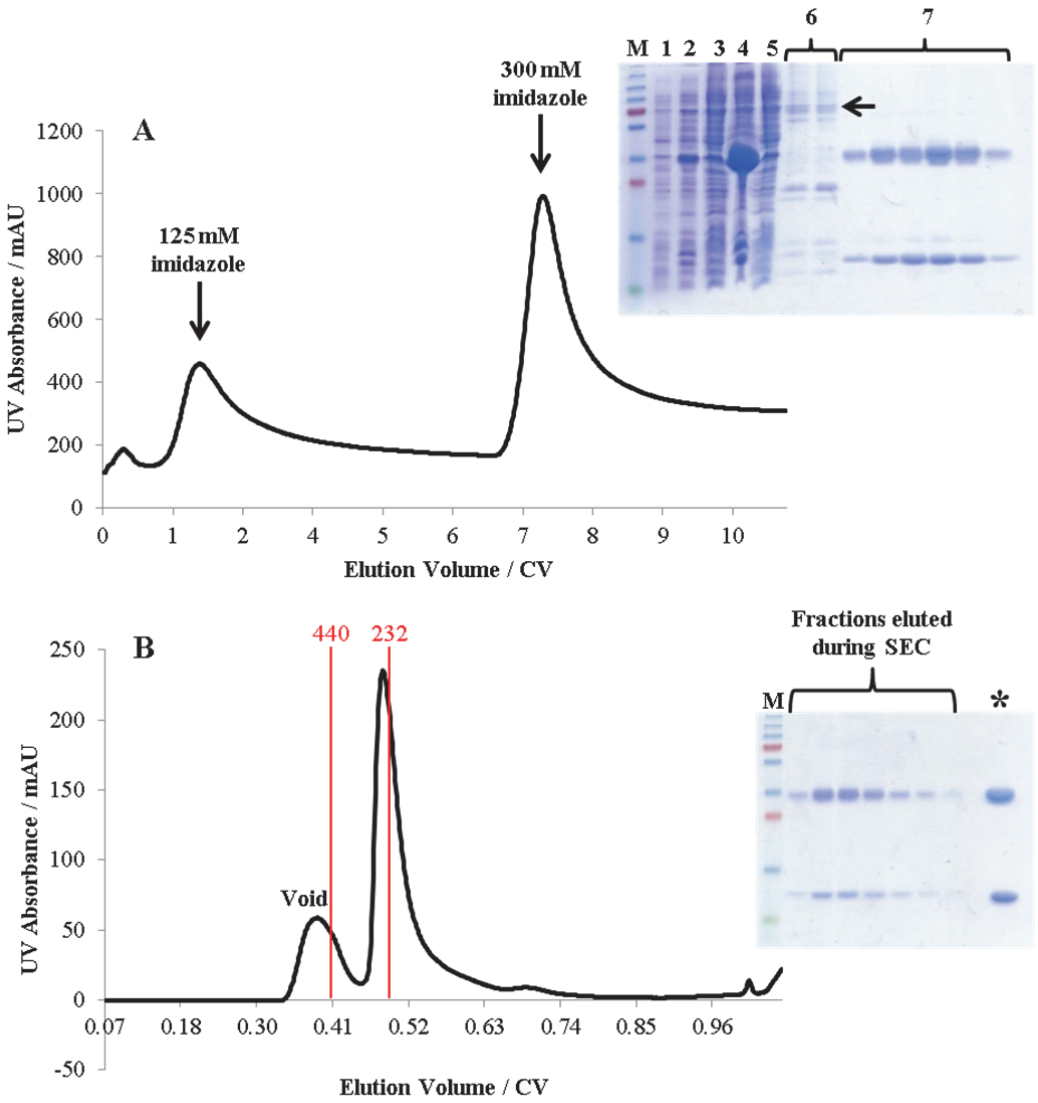


Figure 7.7 *Optimised purification of the His-Spa33/Spa33$_{192-293}$ complex.*

A Ni-affinity chromatography of the His-Spa33/Spa33$_{192-293}$ complex. Elution profile from a 5 ml Ni-NTA Superflow Cartridge equilibrated in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl. Peaks corresponding to elution with 125 mM imidazole and 300 mM imidazole added to the buffer are indicated. The 15% SDS-PA gel in the inset shows samples from the initial part of the purification. M: PageRuler Plus Prestained Protein Ladder (Fermentas), 1: pre-induction sample from B834 (DE3) pLysS pET28b-*spa33$_{192-293}$*, 2: post-induction sample from B834 (DE3) pLysS pET28b-*spa33$_{192-293}$*,, 3: supernatant after 26,000 x g centrifugation, 4: pellet after 26,000 x g centrifugation, 5: flow-through from Ni-cartridge loading, 6: fractions eluted during 125 mM imidazole wash and 7: fractions eluted with 300 mM imidazole. The arrow denotes the likely band corresponding to YsbG.

B SEC of the His-Spa33/Spa33$_{192-293}$ complex. Elution profile from a HiLoad 16/60 Superdex 200 pg column with 20 mM Tris-HCl (pH 7.5), 150 mM NaCl as the running buffer. The red markers indicate the the expected elution volume of calibrated molecular weight markers (kDa) from EMBL (http://www.embl.de/pepcore/pepcore_services/protein_purification/chromatography/hiload26-60_superdex200/). The 15% SDS-PA gel in the inset shows samples taken across the main eluted peak. M: PageRuler Plus Prestained Protein Ladder (Fermentas). The asterisk denotes a sample taken from the concentrated His-Spa33$_{192-293}$ complex.

7.2 Spa33-(Spa33$_{192-293}$)$_2$ is the Building Block of the C-Ring

A His-Spa33-(Spa33$_{192-293}$)$_2$ Complex Undergoes Further Oligomerisation

The stoichiometry and oligomeric state of the purified His-Spa33/Spa33$_{192-293}$ complex was next probed further. Given that no His-Spa33 was eluted with 120 mM imidazole (**Figure 7.7A, bracket 6**), it was possible that several copies of His-Spa33 were present within an oligomer. Furthermore, the main elution peak during SEC was asymmetric and in a position expected for a globular complex >230 kDa (**Figure 7.7B**), suggesting a large, heterogeneous His-Spa33/Spa33$_{192-293}$ complex was present. Indeed, when various concentrations of the purified complex were subjected to MALS by Dr Steven Johnson the main elution peak was seen to elute earlier with increasing concentration, suggesting the His-Spa33/Spa33$_{192-293}$ complex was undergoing dynamic oligomerisation (**Figure 7.8**). The in-line multi-angle scattering across these peaks confirmed this concentration-dependent variability in molecular weight, indicating an increase from ~60 kDa to ~150 kDa over the concentrations tested. Additionally, the molecular weight of the His-Spa33/Spa33$_{192-293}$ complex was variable across each elution peak and particularly noisy for the lowest concentrations. Therefore, it could not be ascertained from the data whether His-Spa33 and Spa33$_{192-293}$ were associating with a defined stoichiometry.

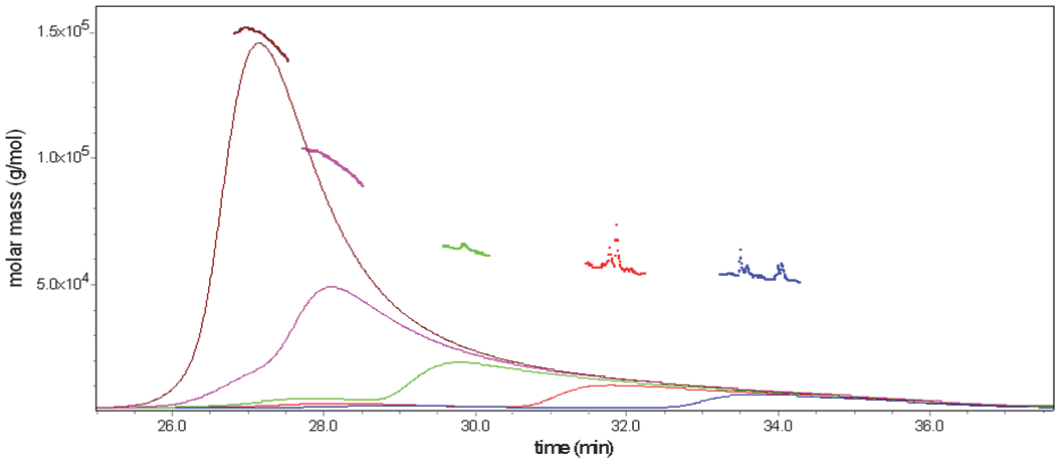


Figure 7.8 *Analysis of the oligomeric state of the His-Spa33/Spa33$_{192-293}$ complex by SEC and in-line multi-angle light scattering.*

The left axis represents the molecular mass at any given point in the elution profile. Elution profiles are shown for the main elution peak at 3.4 mg/ml (brown), 1.7 mg/ml (magenta), 0.85 mg/ml (green). 0.43 mg/ml (red) and 0.20 mg/ml (blue) His-Spa33-(Spa33$_{192-293}$)$_2$.

In order to characterise the His-Spa33/Spa33$_{192-293}$ complex more precisely, native mass spectrometry was carried out by Dr Julien Marcoux (Department of Chemistry, University of Oxford). Spectra of intact complexes were optimally acquired on a Q2B spectrometer, applying low collision energy and no collision gas within the collision cell (**Figure 7.9**). All detectable His-Spa33 was found to be associated with Spa33$_{192-293}$, with the smallest, most abundant complex being accounted for by one copy of His-Spa33 and two copies of Spa33$_{192-293}$ (**Figure 7.9, FL-2C**). Given that this 58757 Da species also correlated with the smallest observed species from MALS measurements (**Figure 7.8, blue**), the core, stable complex formed was likely to be His-Spa33-(Spa33$_{192-293}$)$_2$.

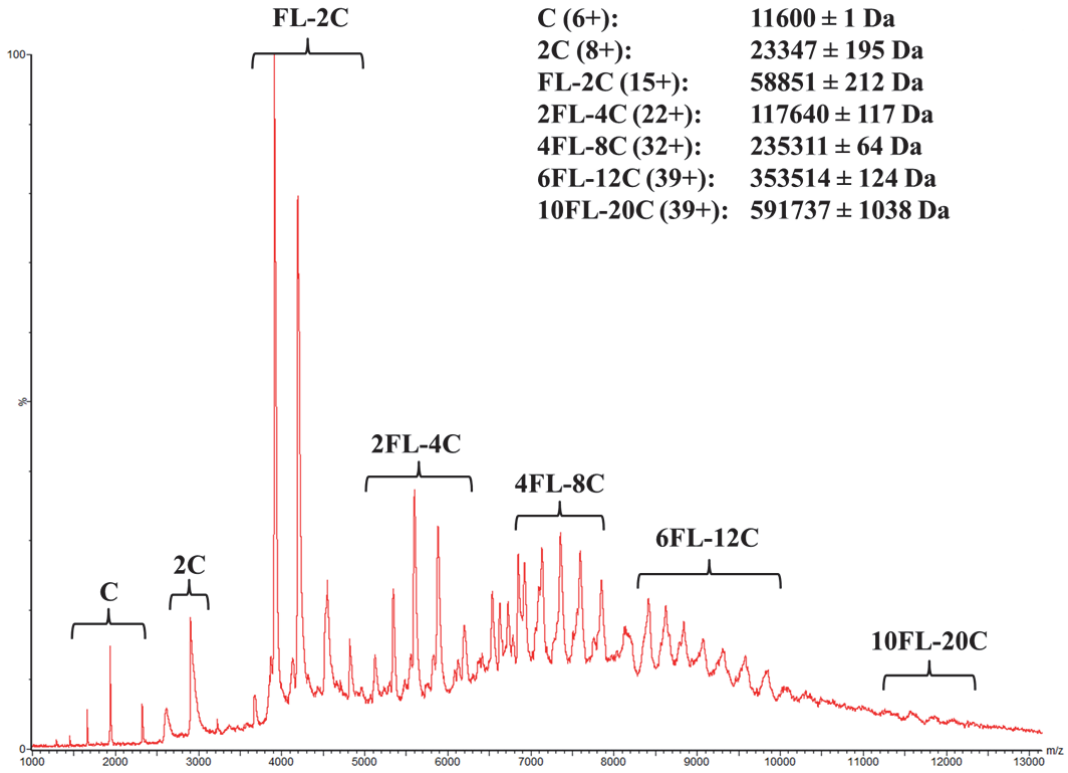


Figure 7.9 *Native Mass Spectrometry of the His-Spa33/Spa33$_{192-293}$ complex.*

Spectrum (provided by Dr Julien Marcoux) obtained for 5.7 mg/ml (97 µM) His-Spa33/Spa33$_{192-293}$ in 200 mM ammonium acetate (pH 7.0) on a Q2B spectrometer (Quadrapole-TOF with collision cell) at 50 V collision energy. The stoichiometry of complexes comprising His-Spa33 (FL) and Spa33$_{192-293}$ (C) are indicated with respect to their corresponding peaks in the spectrum and with their representative molecular weight range and charge state in the inset.

Subsequently it was shown by MALS that YscQ and YscQ-C formed a stable complex with a 1:2 stoichiometry in solution (Bzymek *et al.*, 2012), in agreement with the His-Spa33-$(Spa33_{192-293})_2$ complex present in the gas phase during native mass spectrometry. These results are not in agreement with the consensus view that FliM and FliN associate with a 1:4 stoichiometry, suggesting either that the equivalent C-ring proteins are arranged differently within flagellar and translocation-associated T3SSs or that one of the models is wrong. Initial evidence for the FliM-$(FliN)_4$ model was provided by analytical ultracentrifugation and gel band intensity measurements, with the experimental error on both measurements indicating additional possible stoichiometries (Brown *et al.*, 2005). Although the FliM-FliN interface implicated by cross-linking and mutagenesis allowed a $(FliN)_4$-FliM-$(FliN)_4$ model to be proposed subsequently, assumptions from previous stoichiometry estimates and homology modelling of structurally uncharacterised FliM were also required (Sarkar *et al.*, 2010b). As comparable MALS and native mass spectrometry characterisation demonstrating the accepted FliM-$(FliN)_4$ model are currently unavailable, other stoichiometries could still be possible given the current data. However, a FliM-$(FliN)_4$ complex would agree better than a FliM-$(FliN)_2$ complex with estimated copy numbers of ~34 FliM (Zhao *et al.*, 1996a, Thomas *et al.*, 1999, Thomas *et al.*, 2006) and ~110 FliN (Zhao *et al.*, 1996b) in the entire C-ring. Therefore, further work would be required for both flagellar and translocation-associated T3SSs to unequivocally determine the core structural unit of the C-ring.

Excitingly, the native mass spectrum also implied further oligomerisation of the core His-Spa33-$(Spa33_{192-293})_2$ complex in the gas phase up to 10 copies of His-Spa33 and 20 copies of $Spa33_{192-293}$ (**Figure 7.9, 10FL-20C**). This was in agreement with the variability in the molecular weight of the His-Spa33-$(Spa33_{192-293})_2$ complex observed by MALS (**Figure 7.8**). Together, this demonstrated for the first time that Spa33 homologues from translocation-associated T3SSs have the propensity to form ordered, high molecular weight complexes, as would be required for assembly into a flagellar-like C-ring. Furthermore, although a stable FliM-$(FliN)_4$ complex has been observed during analytical ultracentrifugation (Brown *et al.*,

2005) and intact C-rings from the flagellar T3SS have been visualised by EM (Thomas *et al.*, 2006), there is currently no equivalent bridging data to show how these building blocks form the flagellar C-ring. Therefore, native mass spectrometry provided a novel and interesting means by which to further characterise C-ring assembly across T3SSs.

Intriguingly, only complexes containing an even number of His-Spa33 molecules were present in the spectrum, implying increased stability of these complexes and preferential association of (His-Spa33)$_2$-(Spa33$_{192-293}$)$_4$ oligomers. This could suggest either that His-Spa33 was more stable as a dimer or Spa33$_{192-293}$ was more stable in the gas phase as a tetramer rather than as a dimer, as could be invoked from the characterisation of tetrameric HrcQ$_B$-C (Fadouloglou *et al.*, 2004, Fadouloglou *et al.*, 2009) and FliN (Park *et al.*, 2006, Paul & Blair, 2006). Therefore, the intermolecular interactions within the putative C-ring subcomplexes were likely to be largely dependent on the structure and arrangement of Spa33-C, resulting in further investigation of this domain as discussed in the next sections.

The Interactions and Alternative Conformations of Spa33-C

Whilst two copies of Spa33-C within the His-Spa33-(Spa33$_{192-293}$)$_2$ complex were likely to be arranged as the highly intertwined dimer observed in the crystal structure of Spa33$_{208-293}$ (**Figure 6.12**), it was unknown whether this was a His-Spa33-Spa33$_{192-293}$ heterodimer or a Spa33$_{192-293}$ homodimer. To investigate this, Dr Julien Marcoux carried out further native mass spectrometry, using a LCT instrument with increased collision energy. The spectrum showed that partial dissociation of two high molecular weight His-Spa33/Spa33$_{192-293}$ complexes had occurred (**Figure 7.10**):

2FL-4C (22+) → FL-4C(7+) + FL (15+)

4FL-8C (32+) → 3FL-8C (12+) + FL (20+)

In each case, His-Spa33 had been ejected from the original His-Spa33/Spa33$_{192-293}$ complex, indicating that Spa33$_{192-293}$ molecules were interacting more strongly with each other than with

the full-length protein. Therefore, His-Spa33 was likely to be associating with a stable homodimer of $Spa33_{192-293}$. This result was corroborated by a subsequent analysis of exchange of His-tagged proteins into the untagged YscQ-$(YscQ\text{-}C)_2$ complex, which indicated that YscQ does not form an intertwined dimer with YscQ-C (Bzymek *et al.*, 2012).

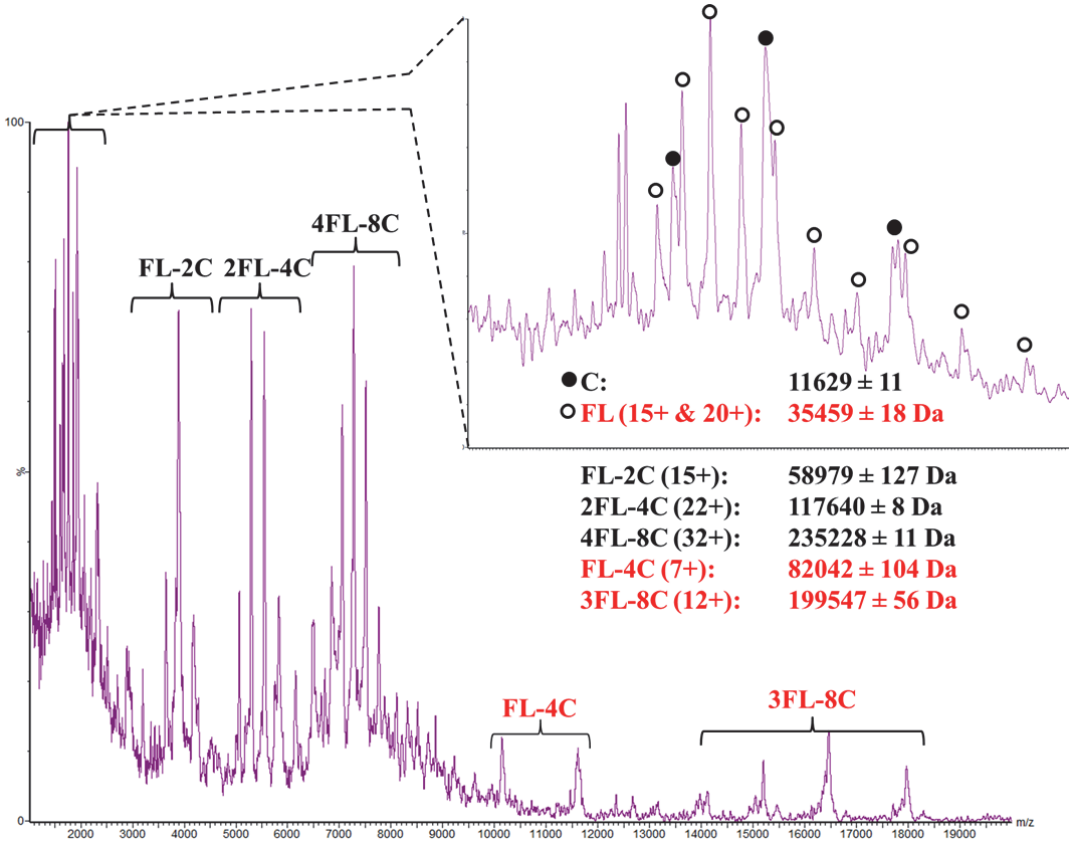


Figure 7.10 *Dissociation of His-Spa33/$Spa33_{192-293}$ complexes during native mass spectrometry*.

Spectrum (provided by Dr Julien Marcoux) obtained for 1.53 mg/ml (27 µM) His-Spa33/$Spa33_{192-293}$ in 200 mM ammonium acetate (pH 7.0) on a LCT spectrometer (Liquid chromatography-TOF). The stoichiometry of complexes comprising His-Spa33 (FL) and $Spa33_{192-293}$ (C) are indicated with respect to their corresponding peaks in the spectrum and with their representative molecular weight range and charge state in the inset. Species shown in red correspond to the products of high energy dissociation. Overlapping peaks corresponding to C and two FL species with different charge states are shown in the zoomed region of the spectrum in the inset.

These findings implied that Spa33-C could exist as a monomer within the context of His-Spa33. Although the individual chains of the $Spa33_{208-293}$ crystal structure could represent the monomer structure, the absence of stabilising antiparallel β-strand interactions between regions corresponding to β1b and β2 in the dimer structure (**Figure 7.11A**) and the solvent exposure of

a large number of hydrophobic side-chains (**Figure 7.11B**) suggested this was unlikely. Alternatively, the region encompassing β1b, α1 and β2 could rearrange within the Spa33-C monomer to form equivalent antiparallel β-strand interactions to those observed within the $Spa33_{208-293}$ dimer, effectively resulting in a monomer model that represents one of the two-fold symmetry-related halves of the dimer structure (**Figure 7.11C**). As a structural rearrangement of either chain from the $Spa33_{208-293}$ dimer structure would be required in the middle of β1 and in the loop region between β2 and β3, the structure of these regions could not be predicted in the Spa33-C monomer model (**Figure 7.11C, dashed lines**). In addition, many of the previously solvent-exposed hydrophobic side-chains (**Figure 7.11B**) were pointing inwards in the monomer model (**Figure 7.11D, inset**), further stabilising this arrangement of Spa33-C.

Such a 'snap-back' mechanism for Spa33-C monomer formation has also been proposed for the monomeric C-terminal domains of FliM (FliM-C)(Sarkar *et al.*, 2010b) and YscQ (Bzymek *et al.*, 2012). Together these models provide the best representation of homologous Spa33-C monomers, as such arrangements are structurally uncharacterised. However, structural differences between monomeric FliM-C and dimeric FliN presumably exist given the low sequence identity of only 11% between *S. typhimurium* domains. Notably, the glycine residues that are absolutely conserved between FliN homologues are not found within FliM/$HrcQ_A$ proteins (**Figure 7.1**). In particular, the side-chain of the residue equivalent to G273 of Spa33 would be expected to point towards the protein core. As these correspond to hydrophilic serine and arginine residues in *S. typhimurium* FliM and $HrcQ_A$, a rearrangement in their C-terminal monomer structure relative to the model (**Figure 7.11C**) could be expected to allow accommodation of these side-chains. Differences in the FliM-C and FliN sequences could therefore be optimal for monomer and dimer formation respectively. Although *S. typhimurium* SPI2 SsaQ has a methionine residue at the position equivalent to G283 (**Figure 7.1**), this side-chain would be expected to point into solution so this substitution does not have significant implications for the structure of the monomer model.

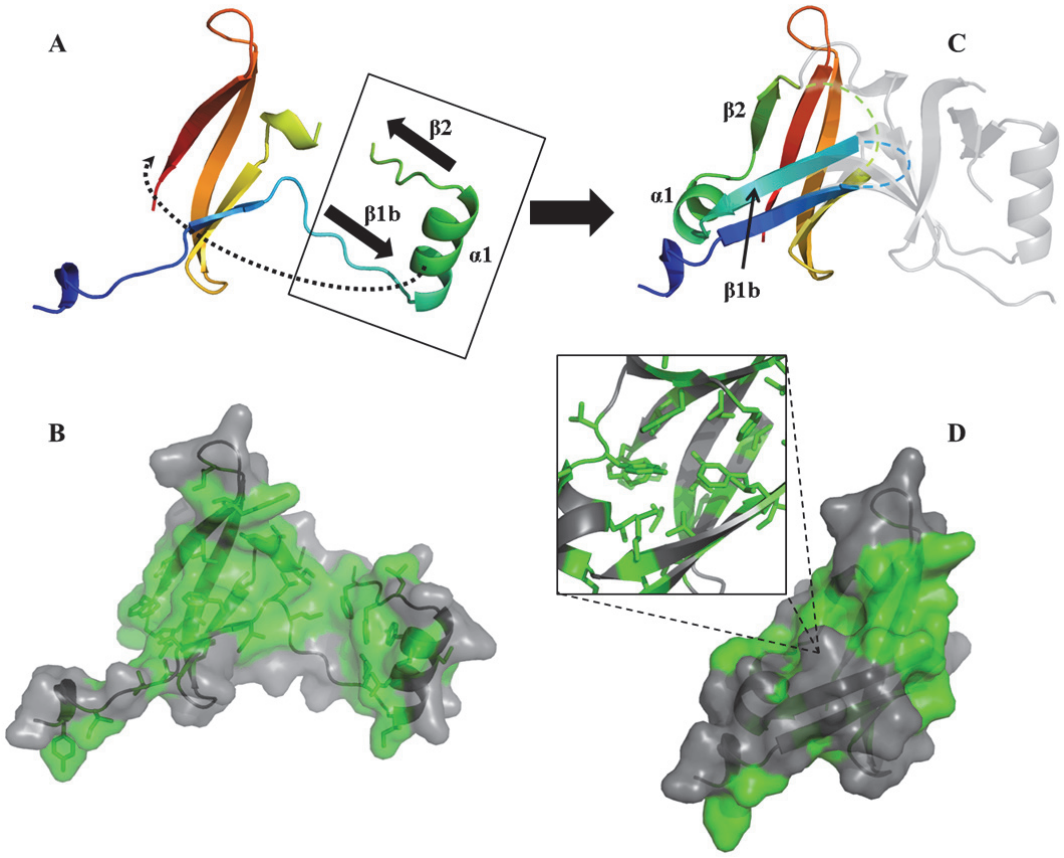


Figure 7.11 *Models for monomeric Spa33-C.*

A Cartoon representation of chain A from the $Spa33_{208-293}$ crystal structure coloured from the N-terminus (blue) to the C-terminus (red). Regions of the sequence which form antiparallel β-strand interactions with β-strands in chain B are labeled as β1b and β2. The dashed arrow indicates the proposed 'snap-back' movement of the boxed region required to form the Spa33-C monomer model.

B Surface representation of chain A from the $Spa33_{208-293}$ crystal structure showing the cartoon representation underneath. Residues with a hydrophobicity >3.0 on the Cornette hydrophobicity scale (Cornette *et al.*, 1987) are shown in green along with their side-chains in stick representation.

C Cartoon representation of the Spa33-C monomer model coloured from the N-terminus (blue) to the C-terminus (red). Residues N220-I231 and H260-M290 of chain A and L232-F253 of chain B from the $Spa33_{208-293}$ structure were used to construct the model. The labeled secondary structure elements correspond to those that are in a different position in the monomer relative to in the dimer subunit shown in **A**. The dashed lines represent regions where the structure cannot be easily inferred from the $Spa33_{208-293}$ dimer structure. The monomer model is superimposed on the $Spa33_{208-293}$ dimer structure (grey).

D Surface representation of the Spa33-C monomer model shown in **C** with the cartoon representation underneath. The zoomed inset shows the side-chains of residues with a hydrophobicity >3.0 (green) pointing in to the centre of the β-barrel.

Within translocation-associated C-rings, additional interactions would presumably be required to dictate whether dimers or monomers form from identical protein sequences. The observation of a $Spa33_{192-293}$ dimer within the His-Spa33/$Spa33_{192-293}$ complex (**Figure 7.10**) and of a $Spa33_{208-293}$ dimer within the crystal structure (**Figure 6.12**), indicated that Spa33-C could dimerise in the absence of the N-terminal region of Spa33. This suggested that the N-terminus of His-Spa33 could be interacting with Spa33-C in a manner that inhibited dimer formation. Indeed, the crystal structure of the middle portion of FliM (FliM-M) from *T. maritima*, which encompassed residues 44-226 of the 328 residue protein, showed the N- and C-termini to be in close proximity (Park *et al.*, 2006). Therefore, the positioning of FliM-M and/or the 44 structurally uncharacterised residues at the N-terminus of FliM (FliM-N) adjacent to FliM-C could inhibit dimerisation of the domain. Although Spa33 only has 10% sequence identity in the region equivalent to FliM-M, a similar general mechanism could be presumed to apply. In future work, it would be interesting to probe these putative interactions between Spa33-C and the currently uncharacterised N-terminal domains of Spa33, to assess their affect on oligomerisation of the former.

Previous work has indicated that the interaction between proteins equivalent to FliN and FliM is mediated by domains equivalent to FliM-C in both flagellar (Tang *et al.*, 1996, Sarkar *et al.*, 2010b) and translocation-associated (Yu *et al.*, 2011) T3SSs. Furthermore, overexpressed full-length *T. maritima* FliM (Brown *et al.*, 2005) and a M218A mutant of YscQ that could not be produced in tandem with YscQ-C (Bzymek *et al.*, 2012) were found to be unstable in solution, indicating complex formation with the FliN or Ysc-Q dimer respectively was likely required for the stability of the monomeric C-terminal domain. Indeed, the Spa33-C monomer model had large patches of surface-exposed hydrophobic residues (**Figure 7.11D**), which would likely decrease the solubility of the domain unless involved in intermolecular interactions. The main insight into the molecular nature of the FliM-FliN interaction comes from a combination of FliN mutants that are no longer pulled down with GST-FliM and cross-links formed between engineered cysteine mutants of FliN and FliM-C (Sarkar *et al.*, 2010b). Sequence alignment

showed that residues implicated in the *E. coli* FliM-FliN interaction are generally not conserved in Spa33, suggesting they may not be involved in an equivalent Spa33-Spa33-C interaction (**Figure 7.12A**). For example, R80 and R121 of FliN are non-conservatively substituted with a methionine and glutamate residue respectively in the Spa33-C sequence. When equivalent residues were mapped on to the structure of $Spa33_{208-293}$, they were mostly found to be distributed away from the dimer symmetry axis (**Figure 7.12B**) and clustered at the side surface in the region of α1 (**Figure 7.12C**). Mapping of equivalent residues to those involved in FliM-FliN cross-linking on to the Spa33-C monomer model also showed clustering on opposite faces of the molecular surface (**Figure 7.12, D&E**).

These positions within the structures correlated with those observed for the FliN and FliM homology models, which had been reconciled by proposing that FliM acts as a bridging subunit between FliN tetramer interfaces in a $(FliN)_4$-FliM-$(FliN)_4$ model (Sarkar *et al.*, 2010b). However, the observation of a His-Spa33-$(Spa33_{192-293})_2$ complex means an analogous interaction interface between Spa33 and $Spa33_{192-293}$ would not be explained with the same model. The stable association of just one copy of a Spa33-C monomer with a $Spa33_{192-293}$ dimer, implies the presence of a unique interaction site within the $Spa33_{208-293}$ dimer structure. Given the 2-fold symmetry of the $Spa33_{208-293}$ dimer, only a binding site along the rotational symmetry axis would not be duplicated. The convex surface of FliN was previously not shown to be involved in the interaction with FliM-C (Sarkar *et al.*, 2010b) and its conserved hydrophobic groove has been implicated in an interaction with FliH (Paul *et al.*, 2006). Therefore the concave surface of the $Spa33_{208-293}$ dimer would be the most plausible location for a monomeric Spa33-C binding site, which would be in the vicinity of observed cross-links between FliM and FliN (**Figure 7.12B, dashed rainbow**). Potentially, a single binding site for a Spa33-C monomer could also be present if the $Spa33_{192-293}$ dimer exhibits asymmetry, as observed for the α1 region in the crystal structure (**Figure 6.13**) where a large number of FliN mutations and cross-links were localised (**Figure 7.12C**). However, the distribution of B factors (**Figure 6.13B**) and the number of peaks observed in the $Spa33_{220-293}$-His $^{1}H,^{15}N$-HSQC spectrum

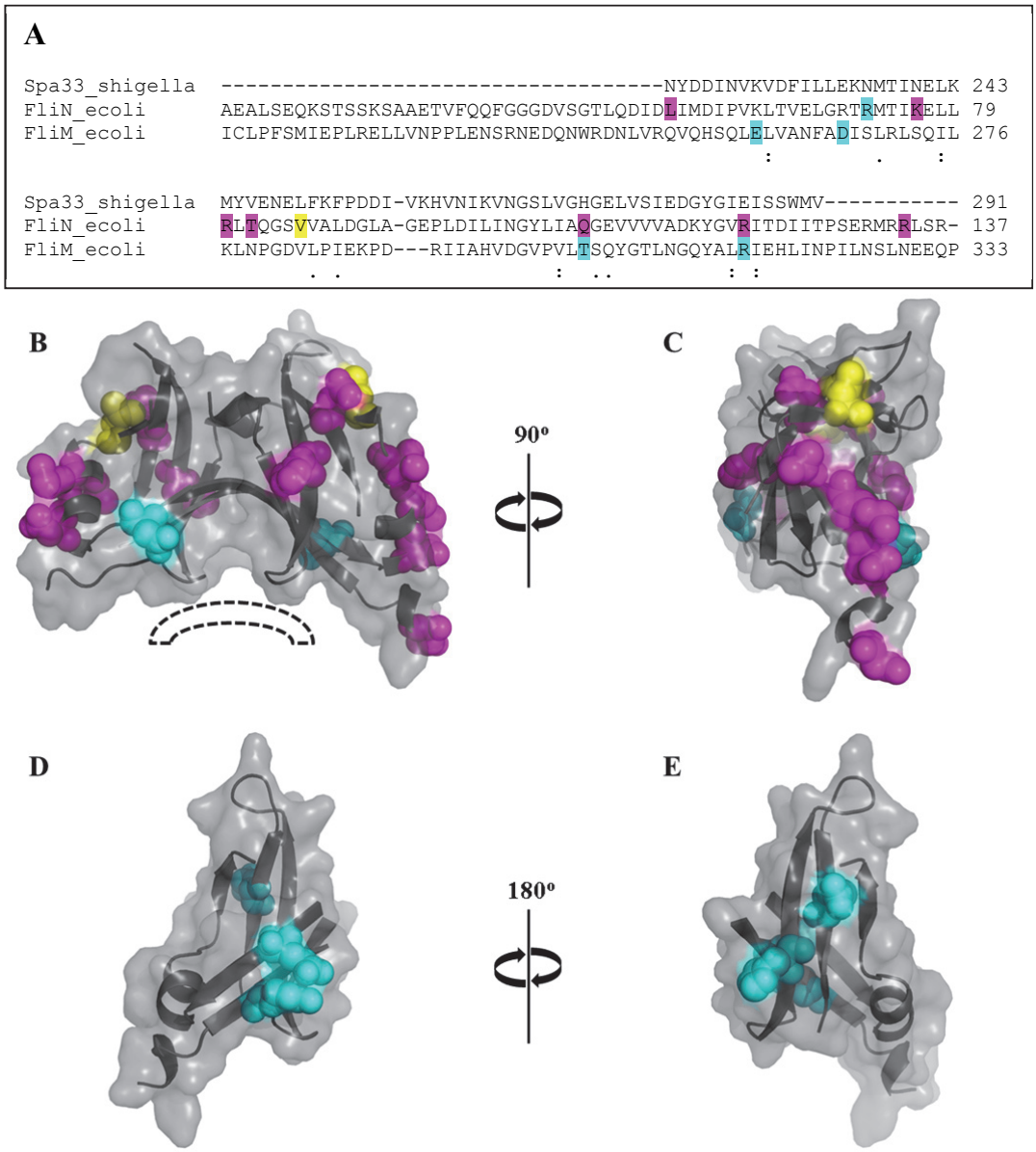


Figure 7.12 *Residues implicated in Spa33-(Spa33$_{192-293}$)$_2$ complex formation from FliM-FliN interaction analysis*.

A Sequence alignment of the C-termini of Spa33_shigella (*S. flexneri*), FliN_ecoli (*E. coli*) and FliM_ecoli (*E. coli*) were performed using Clustal W2, with conserved and semi-conserved residues being indicated by colons and full-stops respectively. Sequence numbers are shown relative to the full-length protein. Only the sequence represented by the structure of Spa33$_{208-293}$ was aligned. The pairwise identity with the Spa33 sequence is 29% for FliN and 16% for FliM. Residues of residues shown to be involved in the FliN-FliM interaction through systematic mutagenesis of FliN (magenta), cross-linking of cysteine engineered FliM and FliN (cyan) or both methods (yellow) are shown.

B-E The molecular surfaces are shown with the cartoon representation underneath. Equivalent residues in Spa33-C to those implicated in the *E. coli* FliM-FliN interaction interface were identified by the alignment in **A**. Side-chains of residues shown to be involved are shown by spheres and coloured according to **A**. The dashed rainbow shows a potential unique binding surface for the Spa33-C monomer.
B The Spa33$_{208-293}$ dimer viewed perpendicular to the 2-fold symmetry axis.
C The Spa33$_{208-293}$ dimer viewed from the N-terminus of chain B. Most residues implicated in Spa33 binding are localised to this face.
D & E Views of the Spa33-C monomer model. Residues implicated in (Spa33$_{192-293}$)$_2$ binding are localised on opposing faces.

(**Figure 6.21**) indicated that regions of asymmetry observed in the $Spa33_{208-293}$ dimer structure were more likely due to backbone flexibility. Alternatively, asymmetry could be induced by binding of monomeric Spa33-C to one of two potential binding sites within the $Spa33_{192-293}$ dimer, with a subsequent conformational change in the second binding site preventing another binding event. Similarly, steric hindrance from bound Spa33 could inhibit binding of another Spa33 molecule to $(Spa33_{192-293})_2$ if the equivalent binding sites within the dimer are close.

To ascertain whether the Spa33-C monomer model had a preferential binding site within the $Spa33_{192-293}$ dimer structure, docking was carried out using the HADDOCK web server (de Vries *et al.*, 2010), specifying residues implicated in FliN-FliM binding as active residues for the interaction. The top cluster of structures obtained from the docking was not outstanding, with similar HADDOCK and Z-scores for the top three clusters indicating they could all represent valid complexes between the Spa33-C monomer model and the $Spa33_{208-293}$ dimer structure (**Table 7.1**). Inspection of these results showed that Spa33-C monomers from the top two clusters were docked in opposite orientations at the dimer interface perpendicular to the two-fold symmetry axis, whilst the Spa33-C monomers from the third cluster were localised to one half of the dimer on the same face (**Figure 7.13A**). Furthermore, none of these interaction interfaces satisfactorily involved the residues implicated in FliM-FliN binding (**Figure 7.13B**). Therefore, the interaction interface between Spa33 and $(Spa33_{192-293})_2$ was unable to be unambiguously predicted from this existing data, requiring further research to structurally define this perplexing 1:2 complex.

Table 7.1 HADDOCK statistics from the docking of the Spa33-C monomer model and $Spa33_{208-293}$ dimer structure

Statistic	**Cluster 1**	**Cluster 2**	**Cluster 3**
HADDOCK score	-48.0 ± 6.5	-44.8 ± 3.2	-43.2 ± 3.6
Cluster size	8	9	8
RMSD from overall lowest energy structure / Å	10.4 ± 0.1	9.4 ± 0.5	8.8 ± 0.9
Buried surface area / $Å^2$	1217 ± 121	1134 ± 118	1049 ± 83
Z-score	-1.6	-1.3	-1.2

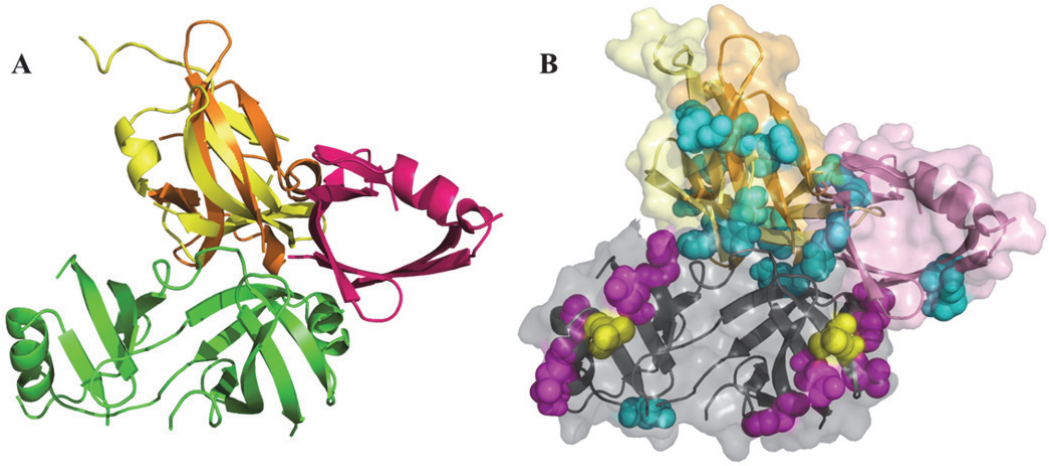


Figure 7.13 *Docking of the Spa33-C monomer model and the Spa33$_{208-293}$ dimer structure*.

A Cartoon representation of the Spa33$_{208-293}$ dimer (green) and representative Spa33-C monomer models from the first (yellow), second (orange) and third (magenta) best docking result from HADDOCK (de Vries *et al.*, 2010) viewed along the dimer two-fold rotational symmetry axis.

B Surface representation of the Spa33$_{208-293}$ dimer (grey) and representative Spa33-C monomer models from the first (pale yellow), second (pale orange) and third (pale magenta) best docking result from HADDOCK with the cartoon representation shown underneath. Side-chains of residues defined as active residues in the interaction are shown by spheres and coloured according to whether they were implicated in the FliM-FliN interaction interface by FliN mutagenesis (magenta), cross-linking of engineered FliM and FliN (cyan) or both (yellow).

Oligomerisation of (Spa33$_{220-293}$)$_2$

In addition to the evidence for a FliM-(FliN)$_4$ complex, isolated FliN and HrcQ$_B$-C have been observed to form stable tetramers both within their crystal lattices and in solution, suggesting a dimer of dimers could be the functionally relevant oligomerisation state of these proteins. However, there is a discrepancy over the molecular nature of dimer association, with two models for the putative tetramer structure. HrcQ$_B$-C was present as an elongated tetramer within the asymmetric unit of the crystal lattice, burying a surface area of 1228 Å^2 and forming stabilising hydrogen bonds at the interface (Fadouloglou *et al.*, 2004). Subsequently, small-angle X-ray scattering and mutagenesis of the dimer-dimer interface indicated the elongated HrcQ$_B$-C tetramer was also present in solution (Fadouloglou *et al.*, 2009). Although analytical ultracentrifugation experiments initially suggested that FliN also formed an elongated tetramer in solution, closer analysis of the crystal contacts revealed a ring tetramer between symmetry-related copies of the protein (Brown *et al.*, 2005). Subsequent cross-linking studies indicated

FliN formed this ring tetramer in solution, predominantly via interactions between the extreme N-terminal and C-terminal helices of each dimer (Paul & Blair, 2006). Interestingly, an analogous ring tetramer of $Spa33_{208-293}$ could be visualised down one of the crystallographic symmetry axes of the lattice (**Figure 7.14A**), providing one possible model for higher order assembly of the domain. However, symmetry related copies were offset by a translation of 13.9 Å when viewed perpendicular to this screw axis (**Figure 7.14B**), indicating the crystal contacts between $Spa33_{208-293}$ dimers were not equivalent to those observed within the FliN tetramer. A crude model for an elongated tetramer of $Spa33_{208-293}$ was also constructed via superimposition of the dimer structure on to each of the $HrcQ_B$-C dimers within the tetramer (**Figure 7.14C**). Although the secondary structure elements involved in the end-to-end association of $Spa33_{208-293}$ dimers were not in equivalent positions to those of $HrcQ_B$-C, the α1-β2 and β4-β5 loop regions that would require minor structural movements were observed to have B-factors >60 $Å^2$ and therefore likely show flexibility (**Figure 6.13B**). Therefore, tetramerisation of Spa33-C was plausible via one of the proposed models and was investigated further.

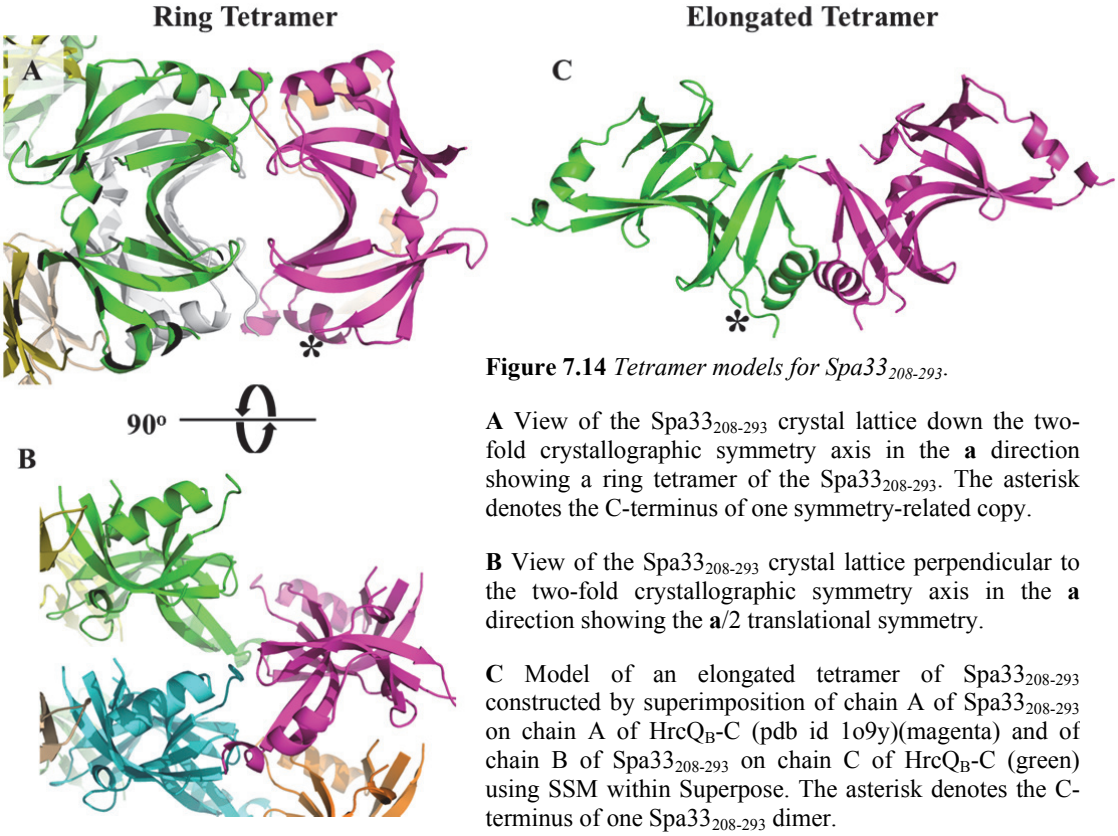


Figure 7.14 *Tetramer models for $Spa33_{208-293}$.*

A View of the $Spa33_{208-293}$ crystal lattice down the two-fold crystallographic symmetry axis in the **a** direction showing a ring tetramer of the $Spa33_{208-293}$. The asterisk denotes the C-terminus of one symmetry-related copy.

B View of the $Spa33_{208-293}$ crystal lattice perpendicular to the two-fold crystallographic symmetry axis in the **a** direction showing the **a**/2 translational symmetry.

C Model of an elongated tetramer of $Spa33_{208-293}$ constructed by superimposition of chain A of $Spa33_{208-293}$ on chain A of $HrcQ_B$-C (pdb id 1o9y)(magenta) and of chain B of $Spa33_{208-293}$ on chain C of $HrcQ_B$-C (green) using SSM within Superpose. The asterisk denotes the C-terminus of one $Spa33_{208-293}$ dimer.

As discussed in the previous chapter, significant line broadening in the $^{1}H,^{15}N$-HSQC spectrum of $Spa33_{220-293}$ suggested further oligomerisation of the dimer in solution (**Figure 6.20A**). Interestingly, the equivalent spectrum for $Spa33_{220-293}$-His showed almost the expected number of backbone amide peaks for the protein, indicating this putative oligomerisation had not occurred to the same extent (**Figure 6.20B**). To probe this further, varying concentrations of both constructs were subjected to MALS by Dr Steven Johnson. As predicted from NMR spectroscopy, the multi-angle scattering across each peak for the $Spa33_{220-293}$ construct demonstrated concentration dependent oligomerisation in solution forming complexes >60 kDa at the highest concentration tested (**Figure 7.15A**). Intriguingly, although the SEC peak for the $Spa33_{220-293}$-His construct was also eluted slightly earlier at higher protein concentrations, the

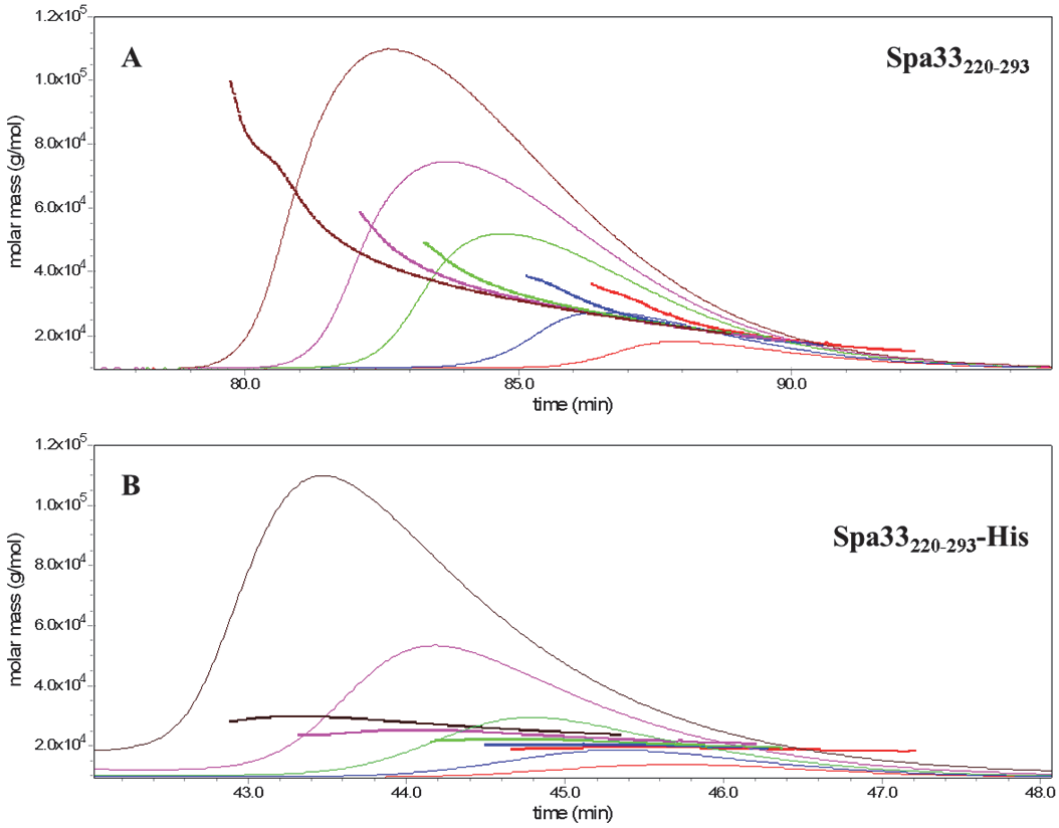


Figure 7.15 *Analysis of the oligomeric state of $Spa33_{220-293}$ constructs by SEC and in-line multi-angle light scattering.*

The left axis represents the molecular mass at any given point in the elution profile.

A Elution profiles for $Spa33_{220-293}$ are shown for the main elution peak at 17.4 mg/ml (brown), 11.7 mg/ml (magenta), 8.3 mg/ml (green), 3.9 mg/ml (blue) and 1.9 mg/ml (red).

B Elution profiles for $Spa33_{220-293}$-His are shown for the main elution peak at 15.6 mg/ml (brown), 8.0 mg/ml (magenta), 4.1 mg/ml (green), 2.0 mg/ml (blue) and 1.0 mg/ml (red).

multi-angle scattering across these peaks showed concentration-dependent oligomerisation had occurred to a lesser extent (**Figure 7.15B**). Furthermore, at the lowest concentrations $Spa33_{220-293}$-His was predominantly dimeric (19.5 kDa), explaining why a good signal-to-noise ratio was observed during acquisition of the $^{1}H,^{15}N$-HSQC spectrum for a 50 μM (0.5 mg/ml) sample. These results suggested that a LEHHHHHH C-terminal tag either sterically or chemically occluded the molecular surface required for $Spa33_{220-293}$ dimer self-association. Indeed, the C-terminus of $Spa33_{208-293}$ can be seen to be proximal to the putative interaction interfaces involved in formation of both the ring and elongated tetramer, requiring further data to discriminate between the possible models for dimer self-association (**Figure 7.14, asterisk**).

To better ascertain the nature of complexes formed by the $Spa33_{220-293}$ complex, Dr Julien Marcoux carried out further native mass spectrometry on a purified sample of the protein (protocol in previous chapter). Under 10 V collision energy on a Q2B spectrometer, the $Spa33_{208-293}$ dimer was observed to oligomerise to form both a dimer of dimers and a trimer of dimers (**Figure 7.16**). Peaks likely corresponding to higher molecular weight $(Spa33_{208-293})_2$ complexes were also present, but were unable to be assigned due to low levels of alternative $Spa33_{220-293}$ species from differential carboxypeptidase A cleavage (**Figure 7.16, asterisk**). Therefore, the $Spa33_{220-293}$ dimer had the propensity to oligomerise both in solution, as had previously been observed for FliN (Brown *et al.*, 2005) and $HrcQ_B$-C (Fadouloglou *et al.*, 2009), and in the gas phase, suggesting self-association between $Spa33_{192-293}$ dimers could be important in mediating C-ring formation. However, it was notable that $(Spa33_{220-293})_2$ oligomers were only observable by native mass spectrometry under very soft activation conditions, in contrast to the 50 V collision energy that could be used to observe intact His-$Spa33_{10}$-$(Spa33_{192-293})_{20}$ complexes (**Figure 7.9**). Therefore, either His-Spa33 or the N-terminus of $Spa33_{192-293}$ could also be contributing to oligomerisation of the His-Spa33-$(Spa33_{192-293})_2$ complex. In support of this, a trimer of $Spa33_{220-293}$ dimers was observed, whilst the equivalent His-$Spa33_3$-$(Spa33_{192-293})_6$ complex was absent from the native mass spectrum obtained for His-Spa33/$Spa33_{192-293}$ complexes (**Figure 7.9**). Therefore, complexes containing an odd number of

His-Spa33-(Spa33$_{192-293}$)$_2$ subunits could be destabilised by His-Spa33 or the N-terminus of Spa33$_{192-293}$.

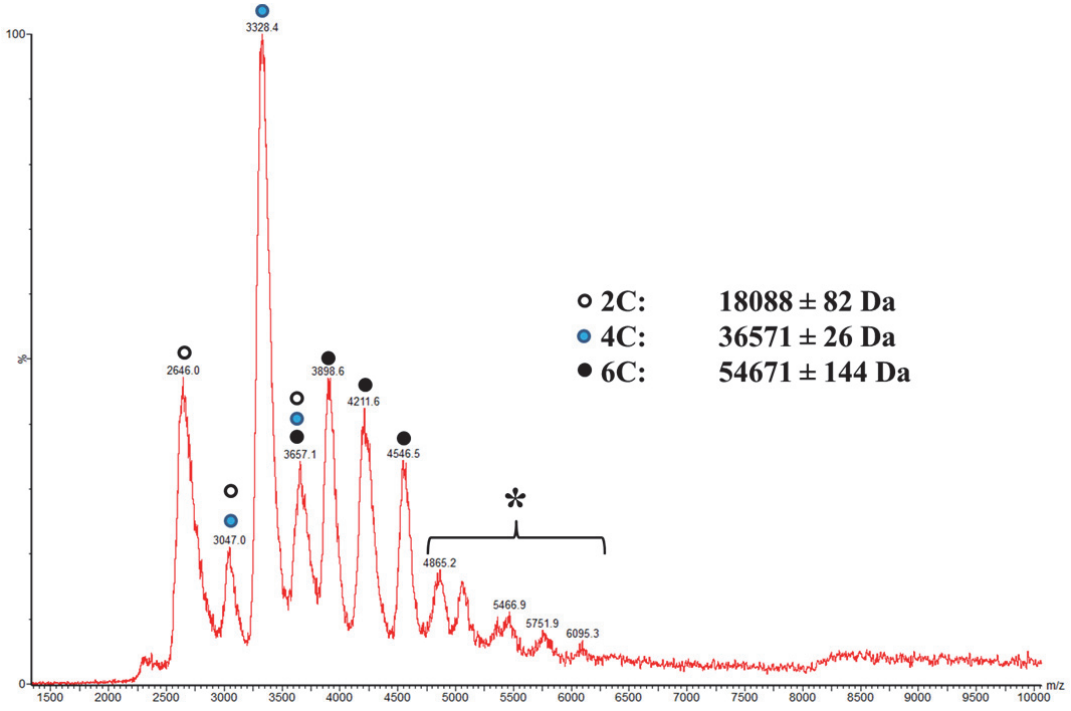


Figure 7.16 *Native Mass Spectrometry of Spa33$_{220-293}$* .

Spectrum (provided by Dr Julien Marcoux) obtained for 0.87 mg/ml (100 µM) Spa33$_{220-293}$ in 200 mM ammonium acetate (pH 7.0) on a Q2B spectrometer at 10 V collision energy. The stoichiometry of complexes comprising Spa33$_{220-293}$ (C) are indicated with respect to their corresponding peaks in the spectrum and their representative molecular weight range in the inset. The peaks indicated by an asterisk could not be assigned due to low levels of several different Spa33$_{220-293}$ species from incomplete carboxypeptidase A digestion.

The oligomerisation of Spa33$_{220-293}$ beyond a tetramer was in contrast to the stable tetramers formed by isolated FliN (Brown *et al.*, 2005) and HrcQ$_B$-C (Fadouloglou *et al.*, 2009) in solution. Furthermore, given that formation of a ring tetramer of Spa33$_{220-293}$ would involve both equivalent, 2-fold symmetry-related binding sites from each dimer (**Figure 7.14A**), it would be difficult to picture how such a tetramer could associate with a further Spa33$_{220-293}$ dimer to form the stable hexamer observed in the native mass spectrum (**Figure 7.16, 6C**). Indeed, the current model for flagellar C-ring formation suggests a bridging interaction by FliM-C domains is required to bring FliN tetramers together (Sarkar *et al.*, 2010b). The N-terminal 57 residues of *E. coli* FliN were found to be required for formation of the ring tetramer (Paul & Blair, 2006),

yet the $Spa33_{220-293}$ dimer, which is equivalent to residues 56-127 of *E. coli* FliN from sequence alignment (**Figure 7.12A**), does not require an equivalent N-terminal extension to oligomerise. Furthermore, the helices at the extreme N- and C-termini of FliN that predominantly mediate formation of the ring tetramer both within the crystal lattice (Brown *et al.*, 2005) and in solution (Paul & Blair, 2006) were not present in the $Spa33_{208-293}$ crystal structure (**Figure 6.15**), suggesting $Spa33_{220-293}$ would not be able to form interactions equivalent to those present in the FliN ring tetramer. However, $(Spa33_{220-293})_2$ could be self-associating via the equivalent interactions found within its crystal lattice, found to be the same as that formed by $Spa33_{208-293}$ in the previous chapter, suggesting it polymerises in a ring tetramer-like arrangement offset by ~14 Å per dimer subunit (**Figure 7.14A&B**). As only higher order His-Spa33/$Spa33_{192-293}$ complexes containing a $Spa33_{192-293}$ tetramer were observed in the native mass spectrum (**Figure 7.9**), the interaction with His-Spa33 could be forcing $(Spa33_{192-293})_2$ to form flat ring tetramers analogous to those observed for FliN.

In contrast, the formation of an elongated tetramer of $Spa33_{208-293}$ would involve only one of two equivalent binding sites from each dimer (**Figure 7.14C**), suggesting another dimer could associate with a tetramer at the second binding site to form the observed hexamer. In this manner, it could be perceived that $Spa33_{192-293}$ dimers further associate to mediate assembly of the C-ring, although the interaction interface between each dimer would have to be adjusted from the current model (**Figure 7.4C**) to increase the angle of association from 120^o (Fadouloglou *et al.*, 2004) to the ~170^o required to form a ~34 member ring. Interestingly, the FliN mutants that prevent association with FliM map to the same molecular surface as the $Spa33_{208-293}$ dimer-dimer interface in the elongated tetramer (**Figure 7.12C, magenta**). Therefore, this binding site could be required for formation of a FliN tetramer, which subsequently interacts more productively with FliM, suggesting an alternative explanation for the results from the pull-down assay (Sarkar *et al.*, 2010b). Similarly, this $Spa33_{220-293}$ tetramer could be stabilised in the presence of full-length Spa33, as evidenced by the presence of (His-Spa33)$_2$-($Spa33_{192-293}$)$_4$ oligomers in the native mass spectrum (**Figure 7.9**).

However, this observation of $Spa33_{220-293}$ hexamers in the absence of full-length Spa33 could also imply that complexes of $Spa33_{220-293}$ were not biologically relevant, indicating predictions of the tetramer structure based on the behaviour of isolated $Spa33_{220-293}$ should be treated with caution. Furthermore, the higher order oligomerisation of a putative $Spa33_{220-293}$ tetramer could be mediated by a distinct interface from that involved in the dimer-dimer interaction or an alternative model for $Spa33_{220-293}$ dimer association could still apply. Indeed, a side by side arrangement of $Spa33_{220-293}$ dimers would perhaps better account for the distinct legs of density seen in the EM reconstruction of the *S. typhimurium* C-ring (Thomas *et al.*, 2006), leaving the molecular surface implicated in the FliM-FliN interaction vacant for Spa33 binding (**Figure 7.12C**). Clearly, further work is required to ascertain the molecular nature of C-ring assembly in both flagellar and translocation-associated T3SSs.

7.3 Towards Structural Characterisation of the *Shigella* T3SS C-ring

Optimisation of His-Spa33-$(Spa33_{192-293})_2$ Complex Purification for Crystallisation Trials

As discussed above, the nature of both the Spa33-C monomer and the interface between Spa33 and $(Spa33_{192-293})_2$ can currently only be hypothesised, whilst full-length Spa33 remains completely structurally uncharacterised. Therefore, a high resolution structure of the Spa33-$(Spa33_{192-293})_2$ complex would provide tremendous insight into C-ring arrangement. However, the yield of the His-Spa33-$(Spa33_{192-293})_2$ complex from the protocol outlined previously was insufficient for extensive crystallisation trials, with only ~0.35 mg complex being obtained from 1 l B834 DE3 (pLysS) pET28b-*spa33*. SDS-PAGE analysis of samples taken from the culture before and after induction of expression showed that much more His-Spa33 had been produced than $Spa33_{192-293}$ (**Figure 7.2**), indicating that translation initiated from the internal RBS and GTG codon was unsurprisingly less efficient. It was hypothesised that this excess His-Spa33 was unstable in the absence of $Spa33_{192-293}$ and was driven into the observed inclusion bodies (**Figure 7.7A, lane 4**). Therefore a construct was designed that would allow co-expression of

His-Spa33 and $Spa33_{192-293}$ from separate promoters, with high translation efficiency for both proteins being facilitated by an optimal RBS and ATG initiation codon. The DNA sequence encoding His-Spa33 (*his-spa33*) was cloned between the *NcoI* and *BamHI* sites of multiple cloning site (MCS) I of the pETDuet-1 vector (Novagen), whilst the sequence encoding $Spa33_{192-293}$ (*spa33*$_{192-293}$) was cloned between the *XhoI* and *NdeI* sites of MCSII to form pETDuet-*his-spa33/spa33*$_{192-293.}$

Amplification of *his-spa33* and *spa33*$_{192-293}$ was achieved using the pET28b-*spa33* construct as a template and the primers outlined in **Table 7.2**. The standard PCR protocol (**Appendix 8.2.4**) specifically produced ample amounts of *his-spa33* (T_a 51-55^{o}C; **Figure 7.17A**) and *spa33*$_{192-293}$ (T_a 60-64^{o}C; **Figure 7.17B**). To remove primers and dNTPs, each reaction was then supplemented with 6 µl *ExoI* 10 x buffer (Fermentas), 5 µl *ExoI* (Fermentas) and 4 µl H_2O and incubated at 37^{o}C for 15 min, then 80^{o}C for 15 min. Reaction mixtures were then pooled pairwise and the inserts purified using the QIAquick PCR Purification Kit (Qiagen) to give ~200 ng/µl DNA.

Table 7.2 Primers used to amplify *his-spa33* and *spa33*$_{192-293}$.

Insert	Forward	Reverse
his-spa33	5'– TA AGA AGG AGA TAT ACC ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AG -3'	5' – GCC GAG CTC GAA TTC GGA TCC TTA CTC CTT TAC CAT CCA AGA AC -3'
spa33$_{192-293}$	5' – TAA GAA GGA GAT ATA CAT ATG AAT GAT AAT AAT GAG GCA AAA ATT AAT CTG TCA G -3'	5'- GGT TTC TTT ACC AGA CTC GAG TTA CTC CTT TAC CAT CCA AGA ACT AAT CTC G -3'

The empty pETDuet vector was sequentially digested with *NcoI* then *BamHI*, as described previously for restriction digests during pBAD-*mxiGN*$_{2-125}$ and pGEX-*mxiGN*$_{2-125}$ cloning. Subsequent separation of a sample of the vector double digest reaction on a 1% agarose gel

demonstrated that digestion of the vector had proceeded to completion (**Figure 7.17C**). The linearised vector was subsequently purified using the QIAquick PCR Purification Kit (Qiagen), giving ~40 µg/ul DNA. The purified *his-spa33* insert was then cloned into the linearised pETDuet vector by homologous recombination using the In-Fusion Dry-Down PCR Cloning Kit (Clontech); the pellet was resuspended in 15 µl of a 5:1 molar ratio of insert:vector and incubated at 42°C for 30 min, before quenching with 40 µl 10 mM Tris-HCl, 1 mM EDTA (pH 8.0). 7 µl of each In-Fusion reaction was subsequently transformed into 200 µl *E. coli* DH5α cells using the gentle heat-shock protocol (**Appendix 8.2.1**). Transformants selected on 100 µg/ml ampicillin agar plates were propagated in a 5 ml overnight culture and the constructs purified using the QIAprep miniprep kit (Qiagen). The presence of the *his-spa33* insert in MCSI was confirmed by DNA sequencing with the DuetDown1 primer (Source BioScience).

pETDuet-*his-spa33* was then sequentially digested with *XhoI* then *NdeI*, with subsequent sample analysis showing the digest had proceeded to completion (**Figure 7.17D**). The linearised

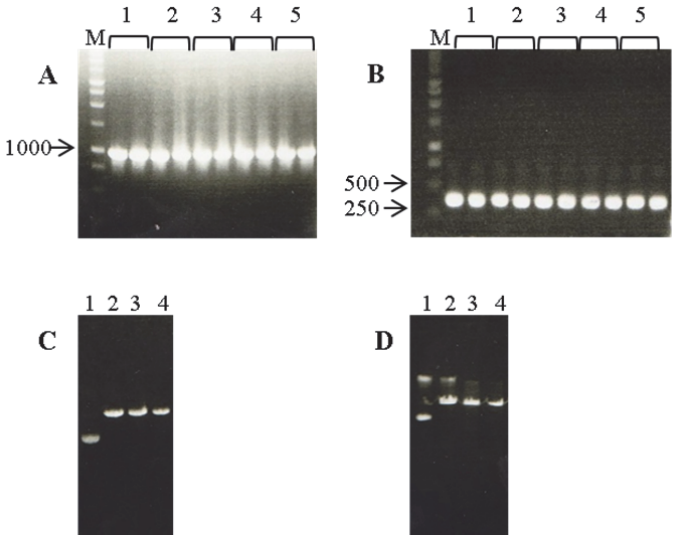


Figure 7.17 *1% agarose gels of DNA fragments from pETDuet-his-spa33/spa33$_{192-293}$ cloning.*

A *his-spa33* insert (977 bp) amplified using the standard PCR protocol. M 1 kb DNA ladder (Promega). T_a = 1: 60°C, 2: 61°c, 3: 62°C, 4: 63°C and 5: 64°C. The relevant marker is indicated in bp.

B *spa33$_{192-293}$* insert (345 bp) amplified using the standard PCR protocol. M 1 kb DNA ladder (Promega). T_a = 1: 51°C, 2: 52°C, 3: 53°C, 4: 54°C and 5: 55°C. The relevant markers are indicated in bp.

C & D Samples from vector double digest reactions.
C Digestion of pETDuet. 1 undigested vector, 2 +ve control for *NcoI* digestion, 3 +ve control for *BamHI* digestion and 4 double digest with *NcoI* and *BamHI*.
D Digestion of pETDuet-*his-spa33*. 1 undigested vector, 2 +ve control for *XhoI* digestion, 3 +ve control for *NdeI* digestion and 4 double digestion with *XhoI* and *NdeI*.

vector was purified and *spa33*$_{192\text{-}293}$ inserted as for *his-spa33*. The presence of *spa33*$_{192\text{-}293}$ in MCSII was confirmed by DNA sequencing with the T7R primer.

Expression from pETDuet-*his-spa33/spa33*$_{192\text{-}293}$ was carried out in B834 DE3 (pLysS) cells as described previously for the pET28b-*spa33* construct. SDS-PAGE analysis showed that both His-Spa33 (35.6 kDa) and Spa33$_{192\text{-}293}$ (11.6 kDa) had been overexpressed at their expected molecular weights (**Figure 7.18**). Subsequent purification of His-Spa33-(Spa33$_{192\text{-}293}$)$_2$ resulted in a yield of ~2.5 mg complex from 1 l B834 DE3 (pLysS) pETDuet-*his-spa33/spa33*$_{192\text{-}293}$, a ~7-fold increase in the yield obtained after expression from pET28b-*spa33* and sufficient for high-throughput screening of crystallisation conditions.

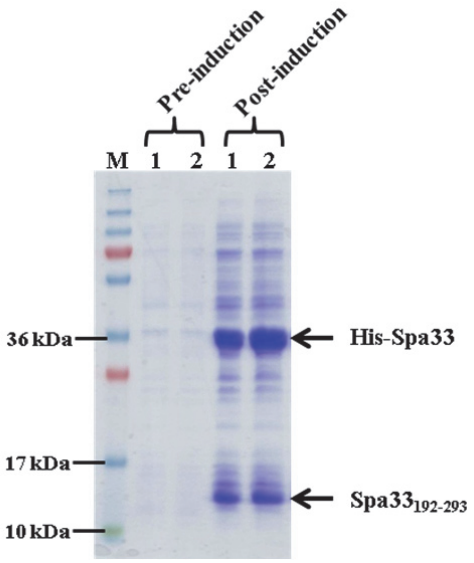


Figure 7.18 *15% SDS-PA gel showing over-expression of His-Spa33 and Spa33*$_{192\text{-}293}$ *in B834 (DE3) pLysS.*

Samples taken from repeat cultures 1 and 2 pre- and post-induction with 1 mM IPTG are labelled. M: PageRuler Plus Prestained Protein Ladder (Fermentas).

MALS (**Figure 7.8**) and native mass spectrometry (**Figure 7.9**) data obtained from the purified His-Spa33-(Spa33$_{192\text{-}293}$)$_2$ complex indicated that further oligomerisation was highly dynamic and concentration-dependent, providing a non-ideal, heterogeneous sample for crystallisation. However, the observation that a C-terminal tag impedes oligomerisation of the Spa33$_{220\text{-}293}$ dimer (**Figure 7.15B**) suggests that addition of a small tag to the C-terminus of Spa33$_{192\text{-}293}$ via site-directed mutagenesis of the pETDuet-*his-spa33/spa33*$_{192\text{-}293}$ construct could potentially impede oligomerisation of His-Spa33-(Spa33$_{192\text{-}293}$)$_2$, to provide a more homogeneous sample.

Alternatively, methylation of the 37 lysine residues within His-Spa33-$(Spa33_{192-293})_2$ could reduce productive interactions between complexes.

The formation of crystal contacts could also be encouraged by the removal of flexible regions from the His-Spa33-$(Spa33_{192-293})_2$ complex. For example, treatment with thrombin would specifically cleave the N-terminal His-tag, whilst limited proteolysis with enzymes such as trypsin and subtilisin could reduce the His-Spa33-$(Spa33_{192-293})_2$ complex to the core interacting domains. As removal of the presumably flexible N-terminus of $Spa33_{208-293}$ increased the crystallisation efficiency of the dimer (previous chapter), it would be interesting to investigate whether the additional 28 residues at the N-terminus of $Spa33_{192-293}$ are redundant for complex formation and could be removed for crystallisation trials.

Does Spa33-$(Spa33_{192-293})_2$ Form a C-ring *in vitro*?

As there were peaks in the native mass spectrum representative of a His-$Spa33_{10}$-$(Spa33_{192-293})_{20}$ complex (**Figure 7.9**), His-Spa33-$(Spa33_{192-293})_2$ clearly had the propensity to form high molecular weight complexes *in vitro*. If FliM and Spa33 are analogously arranged within their respective C-rings, such a 10:20 complex of His-Spa33 and $Spa33_{192-293}$ would represent ~30% of a complete flagellar-like C-ring, assuming the presence of ~34 FliM (Zhao *et al.*, 1996a, Thomas *et al.*, 1999, Thomas *et al.*, 2006). Furthermore, given the concentration-dependent nature of His-Spa33-$(Spa33_{192-293})_2$ oligomerisation (**Figure 7.8**) and the solubility of His-Spa33-$(Spa33_{192-293})_2$ up to 20 mg/ml (340 μM), observable amounts of complexes larger than His-$Spa33_{10}$-$(Spa33_{192-293})_{20}$ could be formed at higher concentrations. It is therefore intriguing whether such significant portions of the C-ring display the expected order and curvature.

Unfortunately, visualisation of such complexes by TEM is precluded by the requirement for a 0.1 mg/ml (17 μM) sample, as isolated His-Spa33-$(Spa33_{192-293})_2$ complexes would be expected to predominate at such a low concentration. Therefore, cross-linking and subsequent dilution of highly concentrated His-Spa33-$(Spa33_{192-293})_2$ would be required for visualisation of high molecular weight complexes. Furthermore, as it is unfeasible to perform native mass

spectrometry at His-Spa33-(Spa33$_{192-293}$)$_2$ concentrations >100 µM, cross-linking would also be required to ascertain whether complexes larger than His-Spa33$_{10}$-(Spa33$_{192-293}$)$_{20}$ are able to form at an observable level *in vitro*. A preliminary cross-linking experiment was carried out by incubating 10 mg/ml (170 µM) His-Spa33-(Spa33$_{192-293}$)$_2$ with a 10-molar excess of the primary amine cross-linkers dimethyl adipimidate (DMA; 8.6 Å spacer arm) or dithiobispropionimidate (DTBP; 11.9 Å spacer arm) in 100 mM HEPES (pH 7.5), 150 mM NaCl at room temperature for 1 h, followed by quenching of the reaction with 130 mM Tris-HCl (pH 6.8). Subsequent SDS-PAGE analysis showed the formation of cross-linked species (**Figure 7.19, red**), although the accumulation of disulphide-bonded dimers of His-Spa33 in the absence of reducing agent precluded further analysis by native mass spectrometry or TEM (**Figure 7.19, asterisks**). Interestingly, a species below the 55 kDa marker was more likely to correspond to Spa33-(Spa33$_{192-293}$) (47.1 kDa) than the Spa33-(Spa33$_{192-293}$)$_2$ complex (58.8 kDa; **Figure 7.19, FL-C**), perhaps implying preferential binding to just one molecule of Spa33$_{192-293}$ within the homodimer. Whilst the thiol within the DTBP spacer arm prevents cross-linking in the presence of reducing agent, further trials with a reduced sample of His-Spa33-(Spa33$_{192-293}$)$_2$ and DMA could be carried out. In addition a GraFix protocol (Stark, 2010) could be implemented to allow gentle intermolecular cross-linking of His-Spa33-(Spa33$_{192-293}$)$_2$ complexes, whilst separating the resulting high molecular weight complexes by density gradient ultracentrifugation.

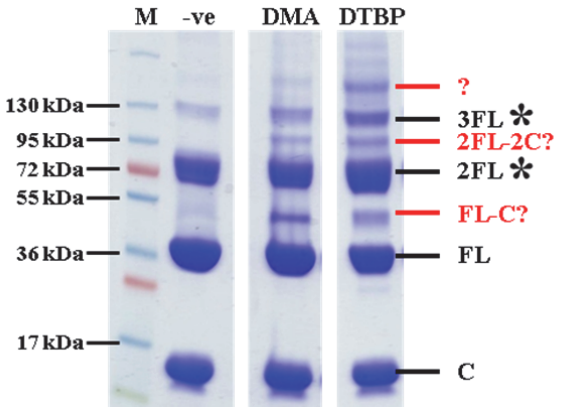


Figure 7.19 *Preliminary cross-linking of His-Spa33-(Spa33$_{192-293}$)$_2$ complexes.*

4-20% mini-Protean TGX gel (Bio-Rad) of cross-linked His-Spa33/Spa33$_{192-293}$. 170 µM His-Spa33-(Spa33$_{192-293}$)$_2$ was cross-linked with a 10 molar excess of DMA or DTBP as indicated. The negative control (-ve) was an equivalent His-Spa33-(Spa33$_{192-293}$)$_2$ sample incubated in the absence of cross-linker. Disulphide-bonded oligomers of His-Spa33 (FL) are indicated with asterisks, whilst putative cross-linked complexes of His-Spa33 (FL) and Spa33$_{192-293}$ (C) are shown in red.

The C-ring Within the Context of the *Shigella* T3SS

TEM visualisation of high molecular weight complexes of His-Spa33-$(Spa33_{192-293})_2$ *in vitro* could show that they do not form partial ring structures, implying that they must be localised to the base of the T3SS *in vivo* in order for a structured C-ring to form. Indeed, deletion of either of the IMR proteins resulted in a diffuse arrangement of fluorescently-labelled YscQ in the cytoplasm of *Y. enterocolitica*, indicating the C-ring localises to the membrane and assembles after formation of the T3SS basal body (Diepold *et al.*, 2010). Furthermore, results presented in Chapter 5 suggested that needle-appendages were absent from basal body complexes lacking MxiG-N, reminiscent of the phenotype of the *spa33*$^-$ strain (Morita-Ishihara *et al.*, 2006). Therefore, localisation of the His-Spa33-$(Spa33_{192-293})_2$ to the membrane could be mediated by a direct interaction with MxiG-N. Indeed, residues 80-217 of Spa33 have previously been shown to interact with GST-MxiG (Morita-Ishihara *et al.*, 2006). However, the nature of this putative interaction is unknown, as evidence discussed in Chapter 3 indicated this was unlikely to be mediated by a proposed interaction with phosphorylated T167 of Spa33 (Barison *et al.*, 2012). Even when localised to the base of the T3SS, there is evidence to suggest both FliM and FliN turnover in the flagellar C-ring (Delalez *et al.*, 2010, Fukuoka *et al.*, 2010). Therefore, the C-ring of translocation T3SSs could also be a dynamic complex rather than a static structure.

Within the flagellar T3SS, the interaction between the C-ring and IMR is not direct and instead mediated by the FliG protein (Thomas *et al.*, 2006). A high resolution structure of full-length FliG from *Aquifex aeolicus* shows the 328 residues to form three α-helical domains (Lee *et al.*, 2010a). The N-terminal domain of FliG (FliG-N) interacts strongly with the C-terminus of the IMR component FliF (Levenson *et al.*, 2012), whilst both the middle- (FliG-M) and C-terminal (FliG-C) domains interact with FliM-M (Paul *et al.*, 2011). Interestingly, a search of the PDB using the fold and function assignment server (FFAS)(Jaroszewski *et al.*) with the 209 residue protein YscK from *Y. pestis* identified residues 30-185 of FliG (FFAS score -6.23), indicating homology to FliG-N and FliG-M. As MxiK within the *S. flexneri* T3SS is a likely orthologue of YscK (Jouihri *et al.*, 2003), it could be proposed that MxiK bridges the middle domain of Spa33

and the IMR. Indeed, interactions between Spa33 and MxiK have previously been observed (Jouihri *et al.*, 2003, Morita-Ishihara *et al.*, 2006), whilst YscK was found to be required for C-ring formation in *Y. pestis* (Diepold *et al.*, 2010). Although MxiJ homologues show sequence identity to the N-terminus of FliF (Crepin *et al.*, 2005), the MxiJ C-terminus is truncated in comparison, with only 10 residues in the cytoplasm (Allaoui *et al.*, 1992), and EPEC EscJ lacks a cytosolic domain. Therefore, a putative interaction between the IMR and MxiK is perhaps more likely to occur via MxiG-N. In addition, the important role of FliG in torque generation and motor reversal within the flagellar T3SS (Stock *et al.*, 2012) would presumably not apply within the translocation-associated T3SS of *Shigella*. Therefore, the very different functional requirements of FliG and MxiK could account for their weak homology and different sizes.

Assembly of the C-ring was also found to be impaired in the absence of the MxiN homologue and ATPase in *Y. pestis*, although activity of the latter was not required (Diepold *et al.*, 2010). Indeed, several studies reported the isolation of soluble complexes of Spa33, MxiN, Spa47 and MxiK (Jouihri *et al.*, 2003, Morita-Ishihara *et al.*, 2006, Johnson & Blocker, 2008), indicating they likely form a structural unit *in vivo*. Evidence discussed in Chapter 6 suggested there was an interaction between Spa33-C and the extreme N-terminus of MxiN, suggesting MxiN could bridge the C-ring and ATPase in the *Shigella* T3SS in an analogous manner to FliH in the flagellar T3SS (McMurry *et al.*, 2006). Therefore, data contributed by previous literature and this thesis could be used to construct a speculative model for the *in vivo* arrangement and interactions of the Spa33-$(Spa33_{192\text{-}293})_2$ complex at the base of the *Shigella* T3SS (**Figure 7.20**).

7.4 Future Perspectives

$Spa33_{192\text{-}293}$ was identified as an alternative translation product of *spa33* that associated as a homodimer with Spa33. This Spa33-$(Spa33_{192\text{-}293})_2$ complex oligomerised *in vitro*, demonstrating for the first time that C-ring proteins from translocation-associated T3SSs have the propensity to form ordered, high molecular weight complexes. This oligomerisation was

found in part to be mediated by self-association of the $Spa33_{192-293}$ dimer. Whilst highlighting that translocation-associated T3SSs likely have two proteins structurally and functionally equivalent to FliM and FliN, results from this work revealed discrepancies with accepted stoichiometries for the flagellar C-ring. Therefore, structural characterisation of both the isolated Spa33-$(Spa33_{192-293})_2$ complex and its oligomers has the potential to provide huge insight into the arrangement of both flagellar and translocation-associated C-rings.

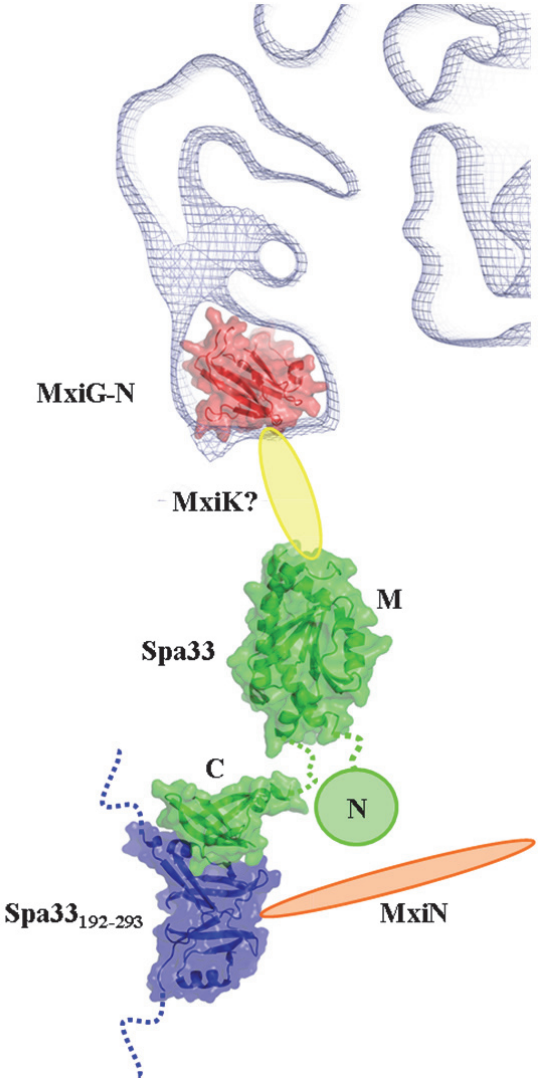


Figure 7.20 *A model for the Spa33-$(Spa33_{192-293})_2$ complex in the context of the Shigella T3SS.*

$Spa33_{192-293}$ (blue) is represented by the crystal structure of $Spa33_{208-293}$ (Chapter 6). Spa33 (green) is represented by the Spa33-C monomer model (this chapter) for the C-terminal domain (C) and the crystal structure of FliM-M from *T. maritima* (pdb id 2hp7)(Park *et al.*, 2006) for the middle domain (M). The position of $Spa33_{208-293}$ relative to the Spa33-C monomer model is derived from cluster 3 from HADDOCK docking (this chapter). MxiG-N (red) is represented by the NMR structure of MxiG-N_{6-112} (Chapter 2) docked into density from the C24 *S. flexneri* NC EM map contoured to 1σ (Hodgkinson *et al.*, 2009). The structurally uncharacterised Spa33 N-terminal domain, MxiK (yellow) and MxiN (orange) are represented by ovals in their putative positions within the *Shigella* T3SS.

References

Abrusci, P., M. Vergara-Irigaray, S. Johnson, M. Beeby, D. Hendrixson, P. Roversi, M. Friede, J. E. Deane, G. Jenson, C. M. Tang & S. M. Lea, (In Press) Architecture of the major component of the type III secretion system export apparatus. *Nature Structural and Molecular Biology*.

Alderwick, L. J., V. Molle, L. Kremer, A. J. Cozzone, T. R. Dafforn, G. S. Besra & K. Fuetterer, (2006) Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 2558-2563.

Allaoui, A., P. Sansonetti, R. Ménard, S. Barzu, J. Mounier, A. Phalipon & C. Parsot, (1995) MxiG, a membrane protein required for secretion of *Shigella* spp. Ipa invasins: involvement in entry into epithelial cells and in intercellular dissemination. *Molecular Microbiology* **17**: 461-470.

Allaoui, A., P. J. Sansonetti & C. Parsot, (1992) MxiJ, a lipoprotein involved in secretion of *Shigella* Ipa invasins, is homologous to YscJ, a secretion factor of the *Yersinia* Yop proteins. *Journal of Bacteriology* **174**: 7661-7669.

Allaoui, A., P. J. Sansonetti & C. Parsot, (1993) MxiD, an outer membrane protein necessary for the secretion of the *Shigella flexneri* Ipa invasins. *Molecular Microbiology* **7**: 59-68.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers & D. J. Lipman, (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.

Andrade, M. A., P. Chacón, J. J. Merelo & F. Morán, (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Engineering* **6**: 383-390.

Aramini, J., (2010) http://www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time.

Arnold, R., S. Brandmaier, F. Kleine, P. Tischler, E. Heinz, S. Behrens, A. Niinikoski, H. W. Mewes, M. Horn & T. Rattei, (2009) Sequence-based prediction of type III secreted proteins. *PLoS Pathogens* **5**: e1000376.

Arnold, T. & D. Linke, (2001) The use of detergents to purify membrane proteins. In: Current Protocols in Protein Science. pp. 4.8.1-4.8.30.

Badea, L., S. Beatson, M. Kaparakis, R. Ferrero & E. Hartland, (2009) Secretion of flagellin by the LEE-encoded type III secretion system of enteropathogenic *Escherichia coli*. *BMC Microbiology* **9**: 30.

Bahrani, F., P. Sansonetti & C. Parsot, (1997) Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infection and Immunity* **65**: 4005-4010.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst & J. A. McCammon, (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10037-10041.

Banci, L., I. C. Felli & R. Kummerle, (2002) Direct detection of hydrogen bonds in monomeric superoxide dismutase: biological implications *Biochemistry* **41**: 2913-2920.

Barison, N., J. Lambers, R. Hurwitz & M. Kolbe, (2012) Interaction of MxiG with the cytosolic complex of the type III secretion system controls *Shigella* virulence. *The FASEB Journal* **26**: 1717-1726.

Barnhart, M. M. & M. R. Chapman, (2006) Curli biogenesis and function. *Annual Review of Microbiology* **60**: 131-147.

Barta, M. L., N. E. Dickenson, M. Patil, A. Keightley, G. J. Wyckoff, W. D. Picking, W. L. Picking & B. V. Geisbrecht, (2012) The structures of coiled-coil domains from type III secretion system translocators reveal homology to pore-forming toxins. *Journal of Molecular Biology* **417**: 395-405.

Bax, A. & M. Ikura, (1991) An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *Journal of Biomolecular NMR* **1**: 99-104.

Betts, H. J., L. E. Twiggs, M. S. Sal, P. B. Wyrick & K. A. Fields, (2008) Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of *Chlamydia trachomatis*. *Journal of Bacteriology* **190**: 1680-1690.

Biemans-Oldehinkel, E., N. Sal-Man, W. Deng, L. J. Foster & B. B. Finlay, (2011) Quantitative proteomic analysis reveals formation of an EscL-EscQ-EscN type III complex in enteropathogenic *Escherichia coli*. *Journal of Bacteriology* **193**: 5514-5519.

Bingle, L. E. H., C. M. Bailey & M. J. Pallen, (2008) Type VI secretion: a beginner's guide. *Current Opinion in Microbiology* **11**: 3-8.

Bischoff, D. S. & G. W. Ordal, (1992) Identification and characterization of FliY, a novel component of the *Bacillus subtilis* flagellar switch complex. *Molecular Microbiology* **6**: 2715-2723.

Blanc, E., P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea & G. Bricogne, (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallographica Section D* **60**: 2210-2221.

Blocker, A., P. Gounon, E. Larquet, K. Niebuhr, V. Cabiaux, C. Parsot & P. Sansonetti, (1999) The tripartite type III secreton of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *The Journal of Cell Biology* **147**: 683-693.

Blocker, A., N. Jouihri, E. Larquet, P. Gounon, F. Ebel, C. Parsot, P. Sansonetti & A. Allaoui, (2001) Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Molecular Microbiology* **39**: 652-663.

Blocker, A. J., J. E. Deane, A. K. Veenendaal, P. Roversi, J. L. Hodgkinson, S. Johnson & S. M. Lea, (2008) What's the point of the type III secretion system needle? *Proceedings of the National Academy of Sciences of the United States of America* **105**: 6507-6513.

Bolanos-Garcia, V. M. & O. R. Davies, (2006) Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1760**: 1304-1313.

Botteaux, A., M. Sani, C. A. Kayath, E. J. Boekema & A. Allaoui, (2008) Spa32 interaction with the inner-membrane Spa40 component of the type III secretion system of *Shigella flexneri* is required for the control of the needle length by a molecular tape measure mechanism. *Molecular Microbiology* **70**: 1515-1528.

Botteaux, A., M. P. Sory, L. Biskri, C. Parsot & A. Allaoui, (2009) MxiC is secreted by and controls the substrate specificity of the *Shigella flexneri* type III secretion apparatus. *Molecular Microbiology* **71**: 449-460.

Braun, P., C. Ockhuijsen, E. Eppens, M. Koster, W. Bitter & J. Tommassen, (2001) Maturation of *Pseudomonas aeruginosa* elastase. Formation of the disulphide bonds. *Journal of Biological Chemistry* **276**: 26030-26035.

Brown, P. N., M. A. A. Mathews, L. A. Joss, C. P. Hill & D. F. Blair, (2005) Crystal structure of the flagellar rotor protein FliN from *Thermotoga maritima*. *Journal of Bacteriology* **187**: 2890-2902.

Brown, P. N., M. Terrazas, K. Paul & D. F. Blair, (2007) Mutational analysis of the flagellar protein FliG: sites of interaction with FliM and implications for organization of the switch complex. *Journal of Bacteriology* **189**: 305-312.

Broz, P., C. A. Mueller, S. A. Müller, A. Philippsen, I. Sorg, A. Engel & G. R. Cornelis, (2007) Function and molecular architecture of the *Yersinia* injectisome tip complex. *Molecular Microbiology* **65**: 1311-1320.

Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti & C. Parsot, (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Molecular Microbiology* **38**: 760-771.

Büttner, D., (2012) Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiology and Molecular Biology Reviews* **76**: 262-310.

Byeon, I. J., H. Li, H. Song, A. M. Gronenborn & M. D. Tsai, (2005) Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. *Nature Structural and Molecular Biology* **12**: 987-993.

Bzymek, K. P., B. Y. Hamaoka & P. Ghosh, (2012) Two translation products of *Yersinia* yscQ assemble to form a complex essential to type III secretion. *Biochemistry* **51**: 1669–1677.

Canutescu, A. A., A. A. Shelenkov & R. L. Dunbrack, (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Science* **12**: 2001-2014.

Carr, H. Y. & E. M. Purcell, (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review* **94**: 630-638.

Cascales, E. & C. Cambillau, (2012) Structural biology of type VI secretion systems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1102-1111.

Cavanagh, J., (2007) *Protein NMR spectroscopy : principles and practice,* p. 885. Academic Press, Burlington, Mass, London.

Cegelski, L., G. R. Marshall, G. R. Eldridge & S. J. Hultgren, (2008) The biology and future prospects of antivirulence therapies. *Nature Reviews Microbiology* **6**: 17-27.

Chatterjee, S., D. Zhong, B. A. Nordhues, K. P. Battaile, S. Lovell & R. N. De Guzman, (2011) The crystal structures of the *Salmonella* type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate. *Protein Science* **20**: 75-86.

Chen, R., L. Li & Z. Weng, (2003) ZDOCK: An initial-stage protein-docking algorithm. *Proteins: Structure, Function, and Bioinformatics* **52**: 80-87.

Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins & J. D. Thompson, (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**: 3497-3500.

Claret, L., S. R. Calder, M. Higgins & C. Hughes, (2003) Oligomerization and activation of the FliI ATPase central to bacterial flagellum assembly. *Molecular Microbiology* **48**: 1349-1355.

Cordes, F. S., K. Komoriya, E. Larquet, S. Yang, E. H. Egelman, A. Blocker & S. M. Lea, (2003) Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *Journal of Biological Chemistry* **278**: 17103-17107.

Cordier, F. & S. Grzesiek, (1999) Direct observation of hydrogen bonds in proteins by interresidue 3hJNC' scalar couplings. *Journal of the American Chemical Society* **121**: 1601-1602.

Cordier, F., L. Nisius, A. J. Dingley & S. Grzesiek, (2008) Direct detection of N-H...O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nature Protocols* **3**: 235-241.

Cornette, J. L., K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky & C. DeLisi, (1987) Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *Journal of Molecular Biology* **195**: 659-685.

Cornilescu, G., B. E. Ramirez, M. K. Frank, G. M. Clore, A. M. Gronenborn & A. Bax, (1999) Correlation between 3hJNC' and hydrogen bond length in proteins. *Journal of the American Chemical Society* **121**: 6275-6279.

Cowtan, K., (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D* **62**: 1002-1011.

Cowtan, K., (2010) Recent developments in classical density modification. *Acta Crystallographica Section D* **66**: 470-478.

Crepin, V. F., S. Prasannan, R. K. Shaw, R. K. Wilson, E. Creasey, C. M. Abe, S. Knutton, G. Frankel & S. Matthews, (2005) Structural and functional studies of the enteropathogenic *Escherichia coli* type III needle complex protein EscJ. *Molecular Microbiology* **55**: 1658-1670.

Das, C., T. S. Ghosh & S. S. Mande, (2011) Computational analysis of the ESX-1 region of *Mycobacterium tuberculosis*: insights into the mechanism of type VII secretion system. *PLoS ONE* **6**: e27980.

Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. Bryan Arendall, III, J. Snoeyink, J. S. Richardson & D. C. Richardson, (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl. Acids Res.* **35**: W375-383.

de Vries, S. J., M. van Dijk & A. M. J. J. Bonvin, (2010) The HADDOCK web server for data-driven biomolecular docking. *Nature Protocols* **5**: 883-897.

Deane, J. E., P. Abrusci, S. Johnson & S. M. Lea, (2010) Timing is everything: the regulation of type III secretion. *Cellular and Molecular Life Sciences* **67**: 1065-1075.

Deane, J. E., S. C. Graham, E. P. Mitchell, D. Flot, S. Johnson & S. M. Lea, (2008a) Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system. *Molecular Microbiology* **69**: 267-276.

Deane, J. E., P. Roversi, F. S. Cordes, S. Johnson, R. Kenjale, S. Daniell, F. Booy, W. D. Picking, W. L. Picking, A. J. Blocker & S. M. Lea, (2006) Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 12529-12533.

Deane, J. E., P. Roversi, C. King, S. Johnson & S. M. Lea, (2008b) Structures of the *Shigella flexneri* type 3 secretion system protein MxiC reveal conformational variability amongst homologues. *Journal of Molecular Biology* **377**: 985-992.

Deeds, E. J., J. A. Bachman & W. Fontana, (2012) Optimizing ring assembly reveals the strength of weak interactions. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 2348-2353.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer & A. Bax, (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**: 277-293.

Delalez, N. J., G. H. Wadhams, G. Rosser, Q. Xue, M. T. Brown, I. M. Dobbie, R. M. Berry, M. C. Leake & J. P. Armitage, (2010) Signal-dependent turnover of the bacterial flagellar switch protein FliM. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 11347-11351.

Delepelaire, P., (2004) Type I secretion in gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1694**: 149-161.

Derewenda, U., A. Mateja, Y. Devedjiev, K. M. Routzahn, A. G. Evdokimov, Z. S. Derewenda & D. S. Waugh, (2004) The structure of *Yersinia pestis* V-Antigen, an essential virulence factor and mediator of immunity against plague. *Structure* **12**: 301-306.

Dickenson, N. E., L. Zhang, C. R. Epler, P. R. Adam, W. L. Picking & W. D. Picking, (2010) Conformational changes in IpaD from *Shigella flexneri* upon binding bile salts provide insight into the second step of type III secretion. *Biochemistry* **50**: 172-180.

Diepold, A., M. Amstutz, S. Abel, I. Sorg, U. Jenal & G. R. Cornelis, (2010) Deciphering the assembly of the *Yersinia* type III secretion injectisome. *The EMBO Journal* **29**: 1928-1940.

Diepold, A., U. Wiesand, M. Amstutz & G. R. Cornelis, (2012) Assembly of the *Yersinia* injectisome: the missing pieces. *Molecular Microbiology* **85**: 878-892.

Diepold, A., U. Wiesand & G. R. Cornelis, (2011) The assembly of the export apparatus (YscR,-S,-T,-U,-V) of the *Yersinia* type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer. *Molecular Microbiology* **82**: 502-514.

Durocher, D., I. A. Taylor, D. Sarbassova, L. F. Haire, S. L. Westcott, S. P. Jackson, S. J. Smerdon & M. B. Yaffe, (2000) The molecular basis of FHA domain: phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Molecular Cell* **6**: 1169-1182.

Dyer, C. M., A. S. Vartanian, H. Zhou & F. W. Dahlquist, (2009) A molecular mechanism of bacterial flagellar motor switching. *Journal of Molecular Biology* **388**: 71-84.

Emsley, P., B. Lohkamp, W. G. Scott & K. Cowtan, (2010) Features and development of Coot. *Acta Crystallographica Section D* **66**: 486-501.

Epler, C. R., N. E. Dickenson, E. Bullitt & W. L. Picking, (2012) Ultrastructural analysis of IpaD at the tip of the nascent MxiH type III secretion apparatus of *Shigella flexneri*. *Journal of Molecular Biology* **420**: 29-39.

Erhardt, M., K. Namba & K. T. Hughes, (2010) Bacterial nanomachines: the flagellum and type III injectisome. *Cold Spring Harbor Perspectives in Biology* **2**: a000299.

Espina, M., A. J. Olive, R. Kenjale, D. S. Moore, S. F. Ausar, R. W. Kaminski, E. V. Oaks, C. R. Middaugh, W. D. Picking & W. L. Picking, (2006) IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infection and Immunity* **74**: 4391-4400.

Evans, L. D. B. & C. Hughes, (2009) Selective binding of virulence type III export chaperones by FliJ escort orthologues InvI and YscO. *FEMS microbiology letters* **293**: 292-297.

Evans, P., (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D* **62**: 72-82.

Fadouloglou, V. E., M. N. Bastaki, A. E. Ashcroft, S. E. V. Phillips, N. J. Panopoulos, N. M. Glykos & M. Kokkinidis, (2009) On the quaternary association of the type III secretion system HrcQB-C protein: Experimental evidence differentiates among the various oligomerization models. *Journal of Structural Biology* **166**: 214-225.

Fadouloglou, V. E., A. P. Tampakaki, N. M. Glykos, M. N. Bastaki, J. M. Hadden, S. E. Phillips, N. J. Panopoulos & M. Kokkinidis, (2004) Structure of HrcQB-C, a conserved component of the bacterial type III secretion systems. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 70-75.

Farmer, B. T. & R. A. Venters, (1995) Assignment of side-chain 13C resonances in perdeuterated proteins. *Journal of the American Chemical Society* **117**: 4187-4188.

Fattori, J., A. Prando, A. Martini Martins, F. Henrique dos Santos Rodrigues & L. Tasic, (2011) Bacterial secretion chaperones. *Protein and Peptide Letters* **18**: 158-166.

Francis, N. R., G. E. Sosinsky, D. Thomas & D. J. DeRosier, (1994) Isolation, characterization and structure of bacterial flagellar motors containing the switch complex. *Journal of Molecular Biology* **235**: 1261-1270.

Fröbel, J., P. Rose & M. Müller, (2012) Twin-arginine-dependent translocation of folded proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1029-1046.

Fujii, T., M. Cheung, A. Blanco, T. Kato, A. J. Blocker & K. Namba, (2012) Structure of a type III secretion needle at 7-Å resolution provides insights into its assembly and signaling mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 4461-4466.

Fukuoka, H., Y. Inoue, S. Terasawa, H. Takahashi & A. Ishijima, (2010) Exchange of rotor components in functioning bacterial flagellar motor. *Biochemical and Biophysical Research Communications* **394**: 130-135.

Fushman, D., S. Cahill & D. Cowburn, (1997) The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration. *Journal of Molecular Biology* **266**: 173-194.

Galkin, V. E., W. H. Schmied, O. Schraidt, T. C. Marlovits & E. H. Egelman, (2010) The structure of the *Salmonella typhimurium* type III secretion system needle shows divergence from the flagellar system. *Journal of Molecular Biology* **396**: 1392-1397.

Goddard, T. D. & D. G. Kneller, SPARKY 3. In. University of California, San Francisco.

González-Pedrajo, B., G. M. Fraser, T. Minamino & R. M. Macnab, (2002) Molecular dissection of *Salmonella* FliH, a regulator of the ATPase FliI and the type III flagellar protein export pathway. *Molecular Microbiology* **45**: 967-982.

González-Pedrajo, B., T. Minamino, M. Kihara & K. Namba, (2006) Interactions between C ring proteins and export apparatus components: a possible mechanism for facilitating type III protein export. *Molecular Microbiology* **60**: 984-998.

Gophna, U., E. Z. Ron & D. Graur, (2003) Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* **312**: 151-163.

Grzesiek, S., J. Anglister & A. Bax, (1993) Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *Journal of Magnetic Resonance, Series B* **101**: 114-119.

Grzesiek, S. & A. Bax, (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **114**: 6291-6293.

Grzesiek, S. & A. Bax, (1992b) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **99**: 201-207.

Guichon, A., D. Hersh, M. R. Smith & A. Zychlinsky, (2001) Structure-function analysis of the *Shigella* virulence factor IpaB. *Journal of Bacteriology* **183**: 1269-1276.

Güntert, P., W. Braun & K. Wüthrich, (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *Journal of Molecular Biology* **217**: 517-530.

Güntert, P., C. Mumenthaler & K. Wüthrich, (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *Journal of Molecular Biology* **273**: 283-298.

Hammet, A., B. Pike, C. McNees, L. Conlan, N. Tenis & J. Heierhorst, (2003) FHA domains as phospho-threonine binding modules in cell signaling. *IUBMB Life* **55**: 23-27.

Hill, C. M., I. R. Bates, G. F. White, F. Ross Hallett & G. Harauz, (2002) Effects of the osmolyte trimethylamine-N-oxide on conformation, self-association, and two-dimensional crystallization of myelin basic protein. *Journal of Structural Biology* **139**: 13-26.

Hodges, K. & G. Hecht, (2012) Interspecies communication in the gut, from bacterial delivery to host-cell response. *The Journal of Physiology* **590**: 433-440.

Hodgkinson, J. L., A. Horsley, D. Stabat, M. Simon, S. Johnson, P. C. A. da Fonseca, E. P. Morris, J. S. Wall, S. M. Lea & A. J. Blocker, (2009) Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nature Structural and Molecular Biology* **16**: 477-485.

Holm, L. & P. Rosenström, (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research* **38**: W545-W549.

Hueck, C. J., (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews* **62**: 379-433.

Hwang, T.-L., S. Mori, A. J. Shaka & P. C. M. van Zijl, (1997) Application of phase-modulated CLEAN chemical exchange spectroscopy (CLEANEX-PM) to detect water-protein proton exchange and intermolecular NOEs. *Journal of the American Chemical Society* **119**: 6203-6204.

Hwang, T.-L., P. C. M. van Zijl & S. Mori, (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical exchange (CLEANEX-PM) approach with a fast-HSQC (FHSQC) detection scheme. *Journal of Biomolecular NMR* **11**: 221-226.

Ibarra, J. A. & O. Steele-Mortimer, (2009) *Salmonella* – the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. *Cellular Microbiology* **11**: 1579-1586.

Ibuki, T., K. Imada, T. Minamino, T. Kato, T. Miyata & K. Namba, (2011) Common architecture of the flagellar type III protein export apparatus and F- and V-type ATPases. *Nature Structural and Molecular Biology* **18**: 277-282.

Ikura, M., L. E. Kay, R. Tschudin & A. Bax, (1990) Three-dimensional NOESY-HMQC spectroscopy of a 13C-labeled protein. *Journal of Magnetic Resonance (1969)* **86**: 204-209.

Irikura, V. M., M. Kihara, S. Yamaguchi, H. Sockett & R. M. Macnab, (1993) *Salmonella typhimurium* fliG and fliN mutations causing defects in assembly, rotation, and switching of the flagellar motor. *Journal of Bacteriology* **175**: 802-810.

Jackson, M. W. & G. V. Plano, (2000) Interactions between type III secretion apparatus components from *Yersinia pestis* detected using the yeast two-hybrid system. *FEMS Microbiology Letters* **186**: 85-90.

Jaroszewski, L., L. Rychlewski, Z. Li, W. Li & A. Godzik, FFAS03: a server for profile–profile sequence alignments. *Nucleic Acids Research* **33**: W284-W288.

Johnson, D. L. & J. B. Mahony, (2007) *Chlamydophila pneumoniae* PknD exhibits dual amino acid specificity and phosphorylates Cpn0712, a putative type III secretion YscD homolog. *J. Bacteriol.* **189**: 7549-7555.

Johnson, D. L., C. B. Stone & J. B. Mahony, (2008) Interactions between CdsD, CdsQ, and CdsL, three putative *Chlamydophila pneumoniae* type III secretion proteins. *Journal of Bacteriology* **190**: 2972-2980.

Johnson, S. & A. Blocker, (2008) Characterization of soluble complexes of the *Shigella flexneri* type III secretion system ATPase. *FEMS Microbiology Letters* **286**: 274-278.

Johnson, S., P. Roversi, M. Espina, A. Olive, J. E. Deane, S. Birket, T. Field, W. D. Picking, A. J. Blocker, E. E. Galyov, W. L. Picking & S. M. Lea, (2007) Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *The Journal of Biological Chemistry* **282**: 4035-4044.

Jouihri, N., M.-P. Sory, A.-L. Page, P. Gounon, C. Parsot & A. Allaoui, (2003) MxiK and MxiN interact with the Spa47 ATPase and are required for transit of the needle components MxiH and MxiI, but not of Ipa proteins, through the type III secretion apparatus of *Shigella flexneri*. *Molecular Microbiology* **49**: 755-767.

Journet, L., C. Agrain, P. Broz & G. R. Cornelis, (2003) The needle length of bacterial injectisomes is determined by a molecular ruler. *Science* **302**: 1757-1760.

Kabsch, W., (2010) XDS. *Acta Crystallographica Section D* **66**: 125-132.

Karplus, M., (1959) Contact electron-spin coupling of nuclear magnetic moments. *The Journal of Chemical Physics* **30**: 11-15.

Kay, L. E., M. Ikura, R. Tschudin & A. Bax, (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **89**: 496-514.

Kay, L. E., G. Y. Xu, A. U. Singer, D. R. Muhandiram & J. D. Formankay, (1993) A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. *Journal of Magnetic Resonance, Series B* **101**: 333-337.

Kim, D. W., G. Lenzen, A.-L. Page, P. Legrain, P. J. Sansonetti & C. Parsot, (2005) The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 14046-14051.

Kim, J. F., (2001) Revisiting the chlamydial type III protein secretion system: clues to the origin of type III protein secretion. *Trends in Genetics* **17**: 65-69.

Kimbrough, T. & S. Miller, (2000) Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 11008-11013.

Koradi, R., M. Billeter & K. Wüthrich, (1996) MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics* **14**: 51-55.

Korotkov, K. V., E. Pardon, J. Steyaert & W. G. J. Hol, (2009) Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure* **17**: 255-265.

Kostakioti, M., C. L. Newman, D. G. Thanassi & C. Stathopoulos, (2005) Mechanisms of protein export across the bacterial outer membrane. *Journal of Bacteriology* **187**: 4306-4314.

Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak & M. M. Levine, (1999) Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization* **77**: 651-666.

Krissinel, E. & K. Henrick, (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D* **60**: 2256-2268.

Krissinel, E. & K. Henrick, (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**: 774-797.

Kubori, T., Y. Matsushima, D. Nakamura, J. Uralil, M. Lara-Tejero, A. Sukhan, J. Galán & S. Aizawa, (1998) Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**: 602-605.

Lara-Tejero, M., J. Kato, S. Wagner, X. Liu & J. E. Galan, (2011) A sorting platform determines the order of protein secretion in bacterial type III systems. *Science* **331**: 1188-1191.

Lario, P. I., R. A. Pfuetzner, E. A. Frey, L. Creagh, C. Haynes, A. T. Maurelli & N. C. Strynadka, (2005) Structure and biochemical analysis of a secretin pilot protein. *The EMBO Journal* **24**: 1111-1121.

Lee, H., C. Yuan, A. Hammet, A. Mahajan, E. S. W. Chen, M.-R. Wu, M.-I. Su, J. Heierhorst & M.-D. Tsai, (2008) Diphosphothreonine-specific interaction between an SQ/TQ cluster and an FHA domain in the Rad53-Dun1 kinase cascade. *Molecular Cell* **30**: 767-778.

Lee, L. K., M. A. Ginsburg, C. Crovace, M. Donohoe & D. Stock, (2010a) Structure of the torque ring of the flagellar motor and the molecular basis for rotational switching. *Nature* **466**: 996-1000.

Lee, P.-C., C. M. Stopford, A. G. Svenson & A. Rietsch, (2010b) Control of effector export by the *Pseudomonas aeruginosa* type III secretion proteins PcrG and PcrV. *Molecular Microbiology* **75**: 924-941.

Lee, S.-Y., H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L.-s. Huang, S. Kustu, E. A. Berry & D. E. Wemmer, (2001a) Crystal structure of an activated response regulator bound to its target. *Nature Structural and Molecular Biology* **8**: 52-56.

Lee, V. T., S. K. Mazmanian & O. Schneewind, (2001b) A program of *Yersinia enterocolitica* type III secretion reactions is activated by specific signals. *Journal of Bacteriology* **183**: 4970-4978.

Leslie, A., (2006) The integration of macromolecular diffraction data. *Acta Crystallographica Section D* **62**: 48-57.

Levenson, R., H. Zhou & F. W. Dahlquist, (2012) Structural insights into the interaction between the bacterial flagellar motor proteins FliF and FliG. *Biochemistry* **51**: 5052-5060.

Levine, M. M., K. L. Kotloff, E. M. Barry, M. F. Pasetti & M. B. Sztein, (2007) Clinical trials of *Shigella* vaccines: two steps forward and one step back on a long, hard road. *Nature Reviews Microbiology* **5**: 540-553.

Li, H. & V. Sourjik, (2011) Assembly and stability of flagellar motor in *Escherichia coli*. *Molecular Microbiology* **80**: 886-899.

Login, F. H. & V. E. Shevchik, (2006) The single transmembrane segment drives self-assembly of OutC and the formation of a functional type II secretion system in *Erwinia chrysanthemi*. *Journal of Biological Chemistry* **281**: 33152-33162.

Loquet, A., N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker & A. Lange, (2012) Atomic model of the type III secretion system needle. *Nature* **486**: 276-279.

Lountos, G. T., B. P. Austin, S. Nallamsetty & D. S. Waugh, (2009) Atomic resolution structure of the cytoplasmic domain of *Yersinia pestis* YscU, a regulatory switch involved in type III secretion. *Protein Science* **18**: 467-474.

Lountos, G. T., J. E. Tropea & D. S. Waugh, (2012) Structure of the cytoplasmic domain of *Yersinia pestis* YscD, an essential component of the type III secretion system. *Acta Crystallographica Section D* **68**: 201-209.

Lunelli, M., R. Hurwitz, J. Lambers & M. Kolbe, (2011) Crystal structure of PrgI-SipD: insight into a secretion competent state of the type three secretion system needle tip and its interaction with host ligands. *PLoS Pathogens* **7**: e1002163.

Lycklama a Nijeholt, J. A. & A. J. M. Driessen, (2012) The bacterial Sec-translocase: structure and mechanism. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1016-1028.

Magdalena, J., A. Hachani, M. Chamekh, N. Jouihri, P. Gounon, A. Blocker & A. Allaoui, (2002) Spa32 regulates a switch in substrate specificity of the type III secreton of *Shigella flexneri* from needle components to Ipa proteins. *Journal of Bacteriology* **184**: 3433-3441.

Mahajan, A., C. Yuan, H. Lee, E. S. Chen, P. Y. Wu & M. D. Tsai, (2008) Structure and function of the phosphothreonine-specific FHA domain. *Science Signaling* **1**: re12.

Marion, D., P. C. Driscoll, L. E. Kay, P. T. Wingfield, A. Bax, A. M. Gronenborn & G. M. Clore, (1989) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1.beta. *Biochemistry* **28**: 6150-6156.

Marlovits, T., T. Kubori, A. Sukhan, D. Thomas, J. Galán & V. Unger, (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306**: 1040-1042.

Marlovits, T. C., T. Kubori, M. Lara-Tejero, D. Thomas, V. M. Unger & J. E. Galán, (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441**: 637-640.

Martinez-Argudo, I. & A. J. Blocker, (2010) The *Shigella* T3SS needle transmits a signal for MxiC release, which controls secretion of effectors. *Molecular Microbiology* **78**: 1365-1378.

Matson, J. S. & M. L. Nilles, (2001) LcrG-LcrV interaction is required for control of Yops secretion in *Yersinia pestis*. *Journal of Bacteriology* **183**: 5082-5091.

Matthews, B. W., (1968) Solvent content of protein crystals. *Journal of Molecular Biology* **33**: 491-497.

Matthews, B. W., (1976) X-ray crystallographic studies of proteins. *Annual Review of Physical Chemistry* **27**: 493-493.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni & R. J. Read, (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**: 658-674.

McDowell, M. A., S. Johnson, J. E. Deane, M. Cheung, A. D. Roehrich, A. J. Blocker, J. M. McDonnell & S. M. Lea, (2011) Structural and functional studies on the N-terminal domain of the *Shigella* type III secretion protein MxiG. *Journal of Biological Chemistry* **286**: 30606-30614.

McFerrin, M. B. & E. H. Snell, (2002) The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *Journal of Applied Crystallography* **35**: 538-545.

McMurry, J. L., J. W. Murphy & B. Gonzalez-Pedrajo, (2006) The FliN-FliH interaction mediates localization of flagellar export ATPase FliI to the C ring complex. *Biochemistry* **45**: 11790-11798.

Meiboom, S. & D. Gill, (1958) Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments* **29**: 688-691.

Meitert, T., E. Pencu, L. Ciudin, M. Tonciu, I. Mihai & S. Nicolescu, (1991) Correlation between Congo red binding as virulence marker in *Shigella* species and Sereny test. *Roumanian Archives of Microbiology and Immunology* **50**: 45-52.

Minamino, T., B. González-Pedrajo, M. Kihara, K. Namba & R. M. Macnab, (2003) The ATPase FliI can interact with the type III flagellar protein export apparatus in the absence of its regulator, FliH. *Journal of Bacteriology* **185**: 3983-3988.

Minamino, T. & R. M. Macnab, (2000) FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity. *Molecular Microbiology* **37**: 1494-1503.

Minamino, T. & K. Namba, (2008) Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export. *Nature* **451**: 485-488.

Minamino, T., S. D. J. Yoshimura, Y. V. Morimoto, B. González-Pedrajo, N. Kami-ike & K. Namba, (2009) Roles of the extreme N-terminal region of FliH for efficient localization of the FliH–FliI complex to the bacterial flagellar type III export apparatus. *Molecular Microbiology* **74**: 1471-1483.

Misic, A. M., K. A. Satyshur & K. T. Forest, (2010) *P. aeruginosa* PilT structures with and without nucleotide reveal a dynamic type IV pilus retraction motor. *Journal of Molecular Biology* **400**: 1011-1021.

Molle, V., L. Kremer, C. Girard-Blanc, G. S. Besra, A. J. Cozzone & J.-F. Prost, (2003) An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* **42**: 15300-15309.

Morita-Ishihara, T., M. Ogawa, H. Sagara, M. Yoshida, E. Katayama & C. Sasakawa, (2006) *Shigella* Spa33 Is an essential C-ring component of type III secretion machinery. *Journal of Biological Chemistry* **281**: 599-607.

Mota, L. J., L. Journet, I. Sorg, C. Agrain & G. R. Cornelis, (2005) Bacterial injectisomes: needle length does matter. *Science* **307**: 1278.

Mougous, J. D., C. A. Gifford, T. L. Ramsdell & J. J. Mekalanos, (2007) Threonine phosphorylation post-translationally regulates protein secretion in *Pseudomonas aeruginosa*. *Nature Cell Biology* **9**: 797-803.

Mueller, C. A., P. Broz, S. A. Muller, P. Ringler, F. Erne-Brand, I. Sorg, M. Kuhn, A. Engel & G. R. Cornelis, (2005) The V-Antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* **310**: 674-676.

Muhandiram, D. R. & L. E. Kay, (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *Journal of Magnetic Resonance, Series B* **103**: 203-216.

Natale, P., T. Brüser & A. J. M. Driessen, (2008) Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778**: 1735-1756.

Nouwen, N., N. Ranson, H. Saibil, B. Wolpensinger, A. Engel, A. Ghazi & A. P. Pugsley, (1999) Secretin PulD: Association with pilot PulS, structure, and ion-conducting channel formation. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 8173-8177.

Okon, M., T. F. Moraes, P. I. Lario, A. L. Creagh, C. A. Haynes, N. C. J. Strynadka & L. P. McIntosh, (2008) Structural characterization of the type-III pilot-secretin complex from *Shigella flexneri*. *Structure* **16**: 1544-1554.

Olive, A. J., R. Kenjale, M. Espina, D. S. Moore, W. L. Picking & W. D. Picking, (2007) Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infection and Immunity* **75**: 2626-2629.

Paetzel, M., A. Karla, N. C. J. Strynadka & R. E. Dalbey, (2002) Signal peptidases. *Chemical Reviews* **102**: 4549-4580.

Pallen, M., R. Chaudhuri & A. Khan, (2002) Bacterial FHA domains: neglected players in the phospho-threonine signalling game? *Trends in Microbiology* **10**: 556-563.

Pallen, M. J., S. A. Beatson & C. M. Bailey, (2005) Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiology Reviews* **29**: 201-229.

Pardi, A., M. Billeter & K. Wüthrich, (1984) Calibration of the angular dependence of the amide proton-Cα proton coupling constants, 3JHNα, in a globular protein: Use of 3JHNα for identification of helical secondary structure. *Journal of Molecular Biology* **180**: 741-751.

Park, S.-Y., B. Lowder, A. M. Bilwes, D. F. Blair & B. R. Crane, (2006) Structure of FliM provides insight into assembly of the switch complex in the bacterial flagella motor. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 11886-11891.

Parsot, C., R. Ménard, P. Gounon & P. J. Sansonetti, (1995) Enhanced secretion through the *Shigella flexneri* Mxi-Spa translocon leads to assembly of extracellular proteins into macromolecular structures. *Molecular Microbiology* **16**: 291-300.

Paul, K. & D. F. Blair, (2006) Organization of FliN Subunits in the Flagellar Motor of *Escherichia coli*. *Journal of Bacteriology* **188**: 2502-2511.

Paul, K., G. Gonzalez-Bonet, A. M. Bilwes, B. R. Crane & D. Blair, (2011) Architecture of the flagellar rotor. *The EMBO Journal* **30**: 2962-2971.

Paul, K., J. G. Harmon & D. F. Blair, (2006) Mutational analysis of the flagellar rotorprotein FliN: Identification of surfaces important for flagellar assembly and switching. *Journal of Bacteriology* **188**: 5240-5248.

Pegues, D., M. Hantman, I. Behlau & S. Miller, (1995) PhoP/PhoQ transcriptional repression of *Salmonella typhimurium* invasion genes: evidence for a role in protein secretion. *Molecular Microbiology* **17**: 169-181.

Penno, C., A. Hachani, L. Biskri, P. Sansonetti, A. Allaoui & C. Parsot, (2006) Transcriptional slippage controls production of type III secretion apparatus components in *Shigella flexneri*. *Molecular Microbiology* **62**: 1460-1468.

Penno, C., P. Sansonetti & C. Parsot, (2005) Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in *Shigella flexneri*. *Molecular Microbiology* **56**: 204-214.

Picking, W. L., H. Nishioka, P. D. Hearn, M. A. Baxter, A. T. Harrington, A. Blocker & W. D. Picking, (2005) IpaD of *Shigella flexneri* is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infection and Immunity* **73**: 1432-1440.

Poyraz, O., H. Schmidt, K. Seidel, F. Delissen, C. Ader, H. Tenenboim, C. Goosmann, B. Laube, A. F. Thunemann, A. Zychlinsky, M. Baldus, A. Lange, C. Griesinger & M. Kolbe, (2010) Protein refolding is required for assembly of the type three secretion needle. *Nature Structural and Molecular Biology* **17**: 788-792.

Privé, G. G., (2007) Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**: 388-397.

Rêgo, A. T., V. Chandran & G. Waksman, (2010) Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochemical Journal* **425**: 475-488.

Reis, R. & F. Horn, (2010) Enteropathogenic *Escherichia coli*, *Samonella*, *Shigella* and *Yersinia*: cellular aspects of host-bacteria interactions in enteric diseases. *Gut Pathogens* **2**: 8.

Robien, M. A., B. E. Krumm, M. Sandkvist & W. G. J. Hol, (2003) Crystal structure of the extracellular protein secretion NTPase EpsE of *Vibrio cholerae*. *Journal of Molecular Biology* **333**: 657-674.

Rohde, J. R., A. Breitkreutz, A. Chenal, P. J. Sansonetti & C. Parsot, (2007) Type III secretion effectors of the IpaH family are E3 ubiquitin ligases. *Cell Host and Microbe* **1**: 77-83.

Ross, J. A. & G. V. Plano, (2011) A C-terminal region of *Yersinia pestis* YscD binds the outer membrane secretin YscC. *Journal of Bacteriology* **193**: 2276-2289.

Russell, A. B., R. D. Hood, N. K. Bui, M. LeRoux, W. Vollmer & J. D. Mougous, (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**: 343-347.

Sack, R. B., M. Rahman, M. Yunus & E. H. Khan, (1997) Antimicrobial resistance in organisms causing diarrheal disease. *Clinical Infectious Diseases* **24**: S102-S105.

Sal‐Man, N., W. Deng & B. B. Finlay, (2012) EscI: a crucial component of the type III secretion system forms the inner rod structure in enteropathogenic *Escherichia coli*. *Biochemical Journal* **442**: 119-125.

Sani, M., A. Allaoui, F. Fusetti, G. T. Oostergetel, W. Keegstra & E. J. Boekema, (2007) Structural organization of the needle complex of the type III secretion apparatus of *Shigella flexneri*. *Micron* **38**: 291-301.

Sanowar, S., P. Singh, R. A. Pfuetzner, I. Andre, H. Zheng, T. Spreter, N. C. J. Strynadka, T. Gonen, D. Baker, D. R. Goodlett & S. I. Miller, (2010) Interactions of the transmembrane polymeric rings of the *Salmonella enterica* serovar typhimurium type III secretion system. *mBio* **1**: e00158-00110.

Sansonetti, P. J., (2001) Rupture, invasion and inflammatory destruction of the intestinal barrier by *Shigella*, making sense of prokaryote-eukaryote cross-talks. *FEMS Microbiology Reviews* **25**: 3-14.

Sansonetti, P. J., (2004) War and peace at mucosal surfaces. *Nature Reviews Immunology* **4**: 953-964.

Sansonetti, P. J., D. J. Kopecko & S. B. Formal, (1982) Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infection and Immunity* **35**: 852-860.

Sarkar, M. K., K. Paul & D. Blair, (2010a) Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 9370-9375.

Sarkar, M. K., K. Paul & D. F. Blair, (2010b) Subunit organization and reversal-associated movements in the flagellar switch of *Escherichia coli*. *Journal of Biological Chemistry* **285**: 675-684.

Schraidt, O., M. D. Lefebre, M. J. Brunner, W. H. Schmied, A. Schmidt, J. Radics, K. Mechtler, J. E. Galan & T. C. Marlovits, (2010) Topology and organization of the *Salmonella typhimurium* type III secretion needle complex components. *PLoS Pathogens* **6**: e1000824.

Schraidt, O. & T. C. Marlovits, (2011) Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. *Science* **331**: 1192-1195.

Schubot, F. D., M. W. Jackson, K. J. Penrose, S. Cherry, J. E. Tropea, G. V. Plano & D. S. Waugh, (2005) Three-dimensional structure of a macromolecular assembly that regulates type III secretion in *Yersinia pestis*. *Journal of Molecular Biology* **346**: 1147-1161.

Schuch, R. & A. Maurelli, (2001) MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of *Shigella*, interact with and stabilize the MxiD secretin in the cell envelope. *Journal of Bacteriology* **183**: 6991-6998.

Scopes, R. K., (1994) *Protein purification: principles and practice*. Springer.

Sedgwick, H., J. E. Cameron, W. C. K. Poon & S. U. Egelhaaf, (2007) Protein phase behavior and crystallization: Effect of glycerol. *The Journal of Chemical Physics* **127**: 125102-125106.

Sekiya, K., M. Ohishi, T. Ogino, K. Tamano, C. Sasakawa & A. Abe, (2001) Supermolecular structure of the enteropathogenic *Escherichia coli* type III secretion system and its direct interaction with the EspA-sheath-like structure. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 11638-11643.

Shen, Y., F. Delaglio, G. Cornilescu & A. Bax, (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* **44**: 213-223.

Silva-Herzog, E., F. Ferracci, M. W. Jackson, S. S. Joseph & G. V. Plano, (2008) Membrane localization and topology of the *Yersinia pestis* YscJ lipoprotein. *Microbiology* **154**: 593-607.

Simorellis, A. K. & P. F. Flynn, (2006) Fast local backbone dynamics of encapsulated ubiquitin. *Journal of the American Chemical Society* **128**: 9580-9581.

Sorg, I., S. Wagner, M. Amstutz, S. A. Muller, P. Broz, Y. Lussi, A. Engel & G. R. Cornelis, (2007) YscU recognizes translocators as export substrates of the *Yersinia* injectisome. *The EMBO Journal* **26**: 3015-3024.

Spaeth, K. E., Y. S. Chen & R. H. Valdivia, (2009) The *Chlamydia* type III secretion system C-ring engages a chaperone-effector protein complex. *PLoS Pathogens* **5**: e1000579.

Spreter, T., C. K. Yip, S. Sanowar, I. André, T. G. Kimbrough, M. Vuckovic, R. A. Pfuetzner, W. Deng, A. C. Yu, B. B. Finlay, D. Baker, S. I. Miller & N. C. J. Strynadka, (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nature Structural and Molecular Biology* **16**: 468-476.

Stark, H., (2010) GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. In: Methods in Enzymology. J. J. Grant (ed). Academic Press, pp. 109-126.

Stein, N., (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **41**: 641-643.

Stock, D., K. Namba & L. K. Lee, (2012) Nanorotors and self-assembling macromolecular machines: the torque ring of the bacterial flagellar motor. *Current Opinion in Biotechnology* **23**: 1-10.

Stone, C. B., D. L. Johnson, D. C. Bulir, J. D. Gilchrist & J. B. Mahony, (2008) Characterization of the putative type III secretion ATPase CdsN (Cpn0707) of *Chlamydophila pneumoniae*. *Journal of Bacteriology* **190**: 6580-6588.

Sukhan, A., T. Kubori, J. Wilson & J. E. Galán, (2001) Genetic analysis of assembly of the *Salmonella enterica* serovar typhimurium type III secretion-associated needle complex. *Journal of Bacteriology* **183**: 1159-1167.

Suzuki, T. & C. Sasakawa, (2001) Molecular basis of the intracellular spreading of *Shigella*. *Infection and Immunity* **69**: 5959-5966.

Tamano, K., S. Aizawa, E. Katayama, T. Nonaka, S. Imajoh-Ohmi, A. Kuwae, S. Nagai & C. Sasakawa, (2000) Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. *The EMBO Journal* **19**: 3876-3887.

Tamano, K., E. Katayama, T. Toyotome & C. Sasakawa, (2002) *Shigella* Spa32 is an essential secretory protein for functional type III secretion machinery and uniformity of its needle length. *Journal of Bacteriology* **184**: 1244-1252.

Tang, H., S. Billings, X. Wang, L. Sharp & D. F. Blair, (1995) Regulated underexpression and overexpression of the FliN protein of *Escherichia coli* and evidence for an interaction between FliN and FliM in the flagellar motor. *Journal of Bacteriology* **177**: 3496-3503.

Tang, H., T. F. Braun & D. F. Blair, (1996) Motility protein complexes in the bacterial flagellar motor. *Journal of Molecular Biology* **261**: 209-221.

Thomas, D. R., N. R. Francis, C. Xu & D. J. DeRosier, (2006) The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of *Salmonella enterica* serovar typhimurium. *Journal of Bacteriology* **188**: 7039-7048.

Thomas, D. R., D. G. Morgan & D. J. DeRosier, (1999) Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 10134-10139.

Tsai, M.-D., (2002) FHA: A signal transduction domain with diverse specificity and function. *Structure* **10**: 887-888.

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

Veenendaal, A. K., J. L. Hodgkinson, L. Schwarzer, D. Stabat, S. F. Zenk & A. J. Blocker, (2007) The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Molecular Microbiology* **63**: 1719-1730.

Vogler, A. P., M. Homma, V. M. Irikura & R. M. Macnab, (1991) *Salmonella typhimurium* mutants defective in flagellar filament regrowth and sequence similarity of FliI to F0F1, vacuolar, and archaebacterial ATPase subunits. *Journal of Bacteriology* **173**: 3564-3572.

Vold, R. L., J. S. Waugh, M. P. Klein & D. E. Phelps, (1968) Measurement of spin relaxation in complex systems. *The Journal of Chemical Physics* **48**: 3831-3832.

Vuister, G. W. & A. Bax, (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNHα) coupling constants in 15N-enriched proteins. *Journal of the American Chemical Society* **115**: 7772-7777.

Wagner, S., L. Königsmaier, M. Lara-Tejero, M. Lefebre, T. C. Marlovits & J. E. Galán, (2010) Organization and coordinated assembly of the type III secretion export apparatus. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17745-17750.

Waksman, G., (2012) Bacterial secretion comes of age. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1014-1015.

Walker, O., R. Varadan & D. Fushman, (2004) Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from 15N relaxation data using computer program ROTDIF. *Journal of Magnetic Resonance* **168**: 336-345.

Walter, T. S., C. Meier, R. Assenberg, K.-F. Au, J. Ren, A. Verma, Joanne E. Nettleship, R. J. Owens, David I. Stuart & J. M. Grimes, (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**: 1617-1622.

Wang, L., H. R. Eghbalnia, A. Bahrami & J. Markley, (2005) Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications. *Journal of Biomolecular NMR* **32**: 13-22.

Wang, Y.-X., J. Jacob, F. Cordier, P. Wingfield, S. J. Stahl, S. Lee-Huang, D. Torchia, S. Grzesiek & A. Bax, (1999) Measurement of 3hJNC′ connectivities across hydrogen bonds in a 30 kDa protein. *Journal of Biomolecular NMR* **14**: 181-184.

Williamson, M. P., T. F. Havel & K. Wüthrich, (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *Journal of Molecular Biology* **182**: 295-315.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin & K. S. Wilson, (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**: 235-242.

Winter, G., (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**: 186-190.

Wishart, D. S., B. D. Sykes & F. M. Richards, (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**: 1647-1651.

Wood, S. E., J. Jin & S. A. Lloyd, (2008) YscP and YscU switch the substrate specificity of the *Yersinia* type III secretion system by regulating export of the inner rod protein YscI. *Journal of Bacteriology* **190**: 4252-4262.

Worrall, L. J., M. Vuckovic & N. C. J. Strynadka, (2010) Crystal structure of the C-terminal domain of the *Salmonella* type III secretion system export apparatus protein InvA. *Protein Science* **19**: 1091-1096.

Wuthrich, K., (1986) *NMR of Proteins and Nucleic Acids*. John Wiley.

Wuthrich, K., M. Billeter & W. Braun, (1984) Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances. *Journal of Molecular Biology* **180**: 715-740.

Wuthrich, K., G. Wider, G. Wagner & W. Braun, (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *Journal of Molecular Biology* **155**: 311-319.

Yamaguchi, S., H. Fujita, A. Ishihara, S. Aizawa & R. M. Macnab, (1986) Subdivision of flagellar genes of *Salmonella typhimurium* into regions responsible for assembly, rotation, and switching. *Journal of Bacteriology* **166**: 187-193.

Yip, C. K., T. G. Kimbrough, H. B. Felise, M. Vuckovic, N. A. Thomas, R. A. Pfuetzner, E. A. Frey, B. B. Finlay, S. I. Miller & N. C. Strynadka, (2005) Structural characterization of the molecular platform for type III secretion system assembly. *Nature* **435**: 702-707.

Yonekura, K., S. Maki-Yonekura & M. Homma, (2011) Structure of the flagellar motor protein complex PomAB: Implications for the torque-generating conformation. *Journal of Bacteriology* **193**: 3863-3870.

Yu, X.-J., M. Liu, S. Matthews & D. W. Holden, (2011) Tandem translation generates a chaperone for the *Salmonella* type III secretion system protein SsaQ. *Journal of Biological Chemistry* **286**: 36098-36107.

Zarivach, R., W. Deng, M. Vuckovic, H. B. Felise, H. V. Nguyen, S. I. Miller, B. B. Finlay & N. C. J. Strynadka, (2008) Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. *Nature* **453**: 124-127.

Zarivach, R., M. Vuckovic, W. Deng, B. B. Finlay & N. C. J. Strynadka, (2007) Structural analysis of a prototypical ATPase from the type III secretion system. *Nature Structural and Molecular Biology* **14**: 131-137.

Zechner, E. L., S. Lang & J. F. Schildbach, (2012) Assembly and mechanisms of bacterial type IV secretion machines. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1073-1087.

Zenk, S. F., D. Stabat, J. L. Hodgkinson, A. K. Veenendaal, S. Johnson & A. J. Blocker, (2007) Identification of minor inner-membrane components of the *Shigella* type III secretion system 'needle complex'. *Microbiology* **153**: 2405-2415.

Zhang, L., Y. Wang, W. L. Picking, W. D. Picking & R. N. De Guzman, (2006) Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *Journal of Molecular Biology* **359**: 322-330.

Zhao, R., C. D. Amsler, P. Matsumura & S. Khan, (1996a) FliG and FliM distribution in the *Salmonella typhimurium* cell and flagellar basal bodies. *Journal of Bacteriology* **178**: 258-265.

Zhao, R., N. Pathak, H. Jaffe, T. S. Reese & S. Khan, (1996b) FliN is a major structural protein of the C-ring in the *Salmonella typhimurium* flagellar basal body. *Journal of Molecular Biology* **261**: 195-208.

Zhong, D., M. Lefebre, K. Kaur, M. A. McDowell, C. Gdowski, S. Jo, Y. Wang, S. H. Benedict, S. M. Lea, J. E. Galan & R. N. De Guzman, (2012) The *Salmonella* type III secretion system inner rod protein PrgJ is partially folded. *Journal of Biological Chemistry* **287**: 25303-25311.

8. Appendix

8.1 Protein Sequences and Construct Directory

MxiG

```
        10         20         30         40         50         60
MSEAKNSNLA PFRLLVKLTN GVGDEFPLYY GNNLIVLGRT IETLEFGNDN FPENIIPVTD

        70         80         90        100        110        120
SKSDGIIYLT ISKDNICQFS DEKGEQIDIN SQFNSFEYDG ISFHLKNMRE DKSRGHILNG

       130        140        150        160        170        180
MYKNHSVFFF FAVIVVLIII FSLSLKKDEV KEIAEIIDDK RYGIVNTGQC NYILAETQND

       190        200        210        220        230        240
AVWASVALNK TGFTKCRYIL VSNKEINRIQ QYINQRFPFI NLYVLNLVSD KAELLVFLSK

       250        260        270        280        290        300
ERNSSKDTEL DKLKNALIVE FPYIKNIKFN YLSDHNARGD AKGIFTKVNV QYKEICENNK

       310        320        330        340        350        360
VTYSVREELT DEKLELINRL ISEHKNIYGD QYIEFSVLLI DDDFKGKSYL NSKDSYVMLN

       370
DKHWFFLDKN K
```

Cytoplasmic domain (MxiG-N) Periplasmic domain (MxiG-C)

Table 8.1 Outline of MxiG constructs used in this study

Construct	Tag	Cloning Sites	Source	Chapter
pET14b-*mxiGN*$_{1\text{-}126}$	N-terminal His-tag	NdeI / BamHI	S. Johnson	2, 3
pBAD-*mxiGN*$_{2\text{-}125}$	N-terminal His-tag	BglII / KpnI	This study	3
pGEX-2T-*mxiGN*$_{2\text{-}125}$	N-terminal GST-tag	BamHI / EcoRI	This study	3
pUC19-*mxiG* (pAB13)	None	SmaI	Allaoui et al., 1995	3
pAB13-R39A/S61A/S63A	None	*Mutagenesis of pAB13*	This study	3
pAB14-ΔN (deletion of residues 12-107)	None	*Mutagenesis of pAB13*	This study	3

pET14b-*mxiGC*$_{142-371}$	N-terminal His-tag	NdeI / BamHI	S. Johnson	4
pET22b-*mxiGC*$_{142-371}$	N-terminal pelB leader sequence C-terminal His-tag	BamHI / XhoI	This study	4
pBAD-*mxiG*	N-terminal His-tag	BglII / KpnI	Zenk et al., 2007	5
pBAD-*mxiGΔN* (deletion of residues 12-107)	N-terminal His-tag	*Mutagenesis of* pBAD-*mxiG*	This study	5

MxiJ

```
        10         20         30         40         50         60
MIRYKCFILF LLLMLIGCEQ REELISNLSQ RQANEIISVL ERHNITARKV DGGKQGISVQ

        70         80         90        100        110        120
VEKGTFASAV DLMRMYDLPN PERVDISQMF PTDSLVSSPR AEKARLYSAI EQRLEQSLVS

       130        140        150        160        170        180
IGGVISAKIH VSYDLEEKNI SSKPMHISVI AIYDSPKESE LLVSNIKRFL KNTFSDVKYE

       190        200        210        220        230        240
NISVILTPKE EYVYTNVQPV KEIKSEFLTN EVIYVFLGMA VLVVILLVWA FKTGWFKRNKI
```

N terminal signal peptide not present in mature protein

Cytoplasmic domain Periplasmic domain

Table 8.2 Outline of MxiJ constructs used in this study

Construct	Tag	Cloning Sites	Source	Chapter
pET14b-*mxiJ*$_{19-206}$	N-terminal His-tag	NdeI / BamHI	S. Johnson	4
MxiJ$_{232-241}$ (synthesised peptide; Severn Biotech)	None	N/A	This study	3

Spa33

```
        10         20         30         40         50         60
MLRIKHFDAN EKLQILYAKQ LCERFSIQTF KNKFTGSESL VTLTSVCGDW VIRIDTLSFL

        70         80         90        100        110        120
KKKYEVFSGF STQESLLHLS KCVFIESSSV FSIPELSDKI TFRITNEIQY ATTGSHLCCF

       130        140        150        160        170        180
SSSLGIIYFD KMPVLRNQVS LDLLHHLLEF CLGSSNVRLA TLKRIRTGDI IIVQKLYNLL

       190        200        210        220        230        240
LCNQVIIGDY IVNDNNEAKI NLSESNGESE HTEVSLALFN YDDINVKVDF ILLEKNMTIN

       250        260        270        280        290
ELKMYVENEL FKFPDDIVKH VNIKVNGSLV GHGELVSIED GYGIEISSWM VKE
```

Alternative translation product

Table 8.3 Outline of Spa33 constructs used in this study

Construct	Tag	Cloning Sites	Source	Chapter
$Spa33_{163-174}$ (synthesised peptide with and without phosphorylation on T167; Severn Biotech)	None	N/A	This study	3
pET28b-*spa33*$_{208-293}$	C-terminal His-tag	NcoI / XhoI	M. Friede	6
pET28b-*spa33*$_{220-293}$	C-terminal His-tag	*Mutagenesis of pET28b-spa33*$_{208-293}$	This study	6
pET28b-*spa33*	N-terminal His-tag	NdeI / BamHI	J. Deane	7
pETDuet-*his-spa33/spa33*$_{192-293}$	Spa33: N-terminal His-tag $Spa33_{192-293}$: none	Spa33: NcoI / BamHI $Spa33_{192-293}$: XhoI / NdeI	This study	7

MxiN

```
        10         20         30         40         50         60
MKVCNMQKGT LPVSRHHAYD GVVIKRIEKE LCKTIKDRDT ESKKKAICVI KDATKKAESL

        70         80         90        100        110        120
RIDAVCDGYQ IGIQTAFEHI IDYICEWKLK QNENRRNIED YITSLLSENL HDERIISTLL

       130        140        150        160        170        180
EQWLSSLRNT VTELKVVLPK CNLALRKKLE LDLHKYRSDV KIILKYSEGN NYIFCSGNQV

       190        200        210        220        230
VEFSPQDVIS GVKIELAEKL TKNDKKYFKE LAHKKLRQIA EDLLKENPVN D
```

Table 8.4 Outline of MxiN constructs used in this study

Construct	Tag	Cloning Sites	Source	Chapter
$MxiN_{1-18}$ (synthesised peptide; Severn Biotech)	None	N/A	This study	6

8.2 Standard Laboratory Methods

8.2.1 Transformation Protocol

Table 8.5 The heat-shock protocol for transformation

1. Add 1 µl of a ~100 ng/µl plasmid stock to 50 µl competent cells
2. Incubate on ice for 30 min
3. Heat-shock at 42°C for 1 min
4. Incubate on ice for 2 min
5. Add 200 µl 2 x YT medium (16 g Bacto-tryptone, 5 g Bacto-yeast extract, 5 g NaCl in 1 l H_2O)
6. Incubate at 37°C for 1 h
7. Spread 100 µl on LB agar (LB medium, 15 g/l agar) & appropriate antibiotic
8. Grow overnight at 37°C

Where specified, a 'gentle transformation' was performed as above, except cells were transferred to a 15 ml falcon tube after heat-shock to allow greater culture aeration and 450 µl 2 x YT medium added. After 1 h incubation at 37°C, cells were harvested by centrifugation and resuspended in 100 µl 2 x YT medium before spreading on LB agar for overnight growth.

8.2.2 SDS-PAGE

50 µl samples were generally supplemented with 1 x loading dye (**Table 8.6**) from a 3 x stock solution and heated to 95°C for 5 min. For the analysis of protein overexpression in cell culture, 1 ml aliquots taken both pre- and post-induction with IPTG were pelleted and resuspended in 1 x loading dye before heating. Reducing and non-reducing SDS-PAGE differs by the inclusion of 10 mM DTT in the 1 x loading dye of the former. Samples were loaded on to a x % SDS-PA gel (**Table 8.6**), which was subsequently run in 1 x SDS-PAGE running buffer (**Table 8.6**) at 200 V within a Mini-PROTEAN Tetra Electrophoresis System (Bio-Rad). Protein bands were visualised by incubation of the gel in Coomassie Staining Solution (**Table 8.6**) for 1 h, followed by incubation overnight in Destaining Solution (**Table 8.6**), both at room temperature.

Table 8.6 Composition of solutions used in SDS-PAGE

Solution	Composition	Concentration
1 x loading dye	Tris-HCl (pH 6.8) Glycerol SDS Bromophenol Blue DTT (reducing gels only)	0.05 M 10 % 2 % 0.1 % 10 mM
x % SDS-PA gel (x = 12, 13.5 or 15 as specified in text)	*Running Gel:* acrylamide/methylene bisacrylamide solution (37.5:1 ratio) Tris-HCl (pH 8.8) SDS Ammonium Persulphate (APS) Tetramethylethylenediamine (TEMED) *Stacking Gel:* acrylamide/methylene bisacrylamide solution (37.5:1 ratio) Tris-HCl (pH 6.8) SDS APS TEMED	 x % v/v 0.4 M 0.1 % 0.1 % 0.01 % 5.1 % 0.13 M 0.1 % 0.1 % 0.01 %
1 x SDS-PAGE running buffer	Tris Glycine SDS	25 mM 190 mM 1 %
Coomassie Staining Solution	Coomassie Brilliant Blue R-250 Methanol Glacial Acetic Acid	0.12 % 40 % 10 %
Destaining Solution	Methanol Glacial Acetic Acid	30 % 10 %

8.2.3 Site-Directed Mutagenesis

Site-directed mutagenesis was conducted using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene). In general, 50 µl mixtures were prepared with the following amounts of each component: 5 µl 10 x reaction buffer, 3 µl Quiksolution, 1 µl dNTP mix, 1 µl template DNA (25 ng/µl), 1 µl forward primer (125 ng/µl), 1 µl reverse primer (125 ng/µl) and 1 µl *PfuTurbo* (2.5 U/µl). For each modification, a parallel negative control reaction was also set-up to which no *PfuTurbo* was added. Construct modification and amplification was then implemented by a TC-5000 gradient thermal cycler (Techne) using the PCR protocol outlined in **Table 8.7**. T_a for each reaction was taken as the temperature 3°C below the melting temperature (T_m) of the primer sequence complementary to the template DNA (calculated using the Promega BioMath calculator). t_e was taken as 1 min/kb of plasmid length.

Table 8.7 PCR protocol used for site-directed mutagenesis.

Stage	Temperature / °C	Time	Number of Cycles
Initial denaturation	95	1 min	1
Denaturation	95	50 s	18
Annealing	Reaction-specific T_a	50 s	
Elongation	68	Reaction-specific t_e	
Final elongation	68	7 min	1

Subsequently, each reaction was directly supplemented with 1 µl *DpnI* and incubated at 37°C for 1 h to selectively digest the template DNA. 3 µl β-mercaptoethanol was added to 65 µl aliquots of XL10-Gold ultracompetent cells, before transformation with 3 µl of each reaction by

the 'gentle transformation' protocol outlined in **Appendix 8.2.1**. Constructs were purified from 5 ml overnight cultures using the QIAprep miniprep kit (Qiagen) and the presence of the desired mutation confirmed by DNA sequencing (Source BioScience).

8.2.4 PCR For Insert Preparation

To obtain 50-100 ng/ul of insert, polymerase chain reaction (PCR) was used. In general, 3 x 50 µl mixtures were prepared with the following amounts of each component: 5 µl 10 x buffer I (Novagen), 3 µl $MgCl_2$ (25 mM; Novagen), 5 µl dNTP mix (2 mM; Novagen), 1 µl template DNA (1 ng/µl), 1 µl forward primer (5 µM), 1 µl reverse primer (5 µM) and 0.4 µl KOD polymerase (2.5 U/µl; Novagen). Inserts were then amplified using either the standard PCR protocol (**Table 8.8**) or the touch-down PCR protocol (**Table 8.9**). For the standard PCR protocol and the final 25 cycles of the touch-down PCR protocol, T_a was taken as the temperature 3^oC below T_m of the complete primer sequence. For the first 10 cycles of the touch-down PCR protocol, T_a was taken as the temperature 3^oC below T_m of the primer sequence complementary to the template DNA.

Table 8.8 Standard PCR protocol for amplification of inserts

Stage	Temperature / ^oC	Time	Number of Cycles
Initial denaturation	95	5 min	1
Denaturation	95	15 s	30
Annealing	Reaction-specific T_a	4 s	
Elongation	72	40 s	
Final elongation	72	5 min	1

Table 8.9 Touch-down PCR protocol for amplification of inserts

Stage	Temperature / ^{o}C	Temperature gradient	Time	Number of Cycles
Initial denaturation	95	--	5 min	1
Denaturation	95	--	15 s	10
Annealing	Reaction-specific T_a	0.5^{o}C per cycle	4 s	
Elongation	72	--	40 s	
Denaturation	95	--	15 s	25
Annealing	Reaction-specific T_a	--	4 s	
Elongation	72	--	40 s	
Final elongation	72	--	5 min	1

8.2.5 Agarose Gel Electrophoresis

DNA samples were combined with 6 x Blue/Orange Loading Dye (Promega) and loaded on to a 50 ml gel, comprising 1% agarose, 1 μg/ml ethidium bromide and 1 x TAE buffer (40 mM Tris-acetate pH 8.3, 2 mM disodium EDTA). Gels were run at 70 V in a Mini-Sub Cell GT Cell (Bio-Rad) with 1 x TAE and 0.5 μg/ml ethidium bromide as the running buffer. Bands were subsequently visualised with UV light using a Gel Doc 2000 (Bio-Rad).

8.2.6 ThermoFluor Assay

25 µl reactions (7.5 µl SYPRO Orange dye (Sigma), 12.5 µl buffer condition, 5 µl protein solution) were set-up in a 96 well ABgene PCR plate (Thermo scientific) and sealed with ABsolute QPCR seal (Thermo scientific). Thermal denaturation was then carried out using a Stratagene RTPCR 305 machine, with an initial equilibration at 25^{o}C for 5 min followed by 36 cycles each with a 1.3^{o}C increase in temperature relative to the previous. An initial calibration was carried out to determine the dilutions of both dye and protein required to obtain optimal signal (**Table 8.10**).

Table 8.10 Calibration reactions used to optimise dye and protein dilutions for ThermoFluor

		Dilution of SYPRO Orange dye (7.5 µl per reaction)			
		1:5000	1:1000	1:500	1:100
Concentration of protein in final reaction / mg/ml	1.5				
	0.7				
	0.35				
	0.18				

8.2.7 Western Blotting

SDS-PAGE was carried out as described previously and the resulting SDS-PA gel (within 15 min of completion) was sandwiched next to an Amersham Hybond-P membrane (GE Healthcare) between 4 x sheets of extra thick filter paper (Bio-Rad), all of which were pre-soaked in transfer buffer (20% v/v methanol, 0.0375% w/v SDS, 39 mM glycine, 48 mM Tris). Protein transfer from the gel to the membrane was then carried out in a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad), run at 22 V for 1 h. Successful transfer was confirmed by the presence of the PageRuler Plus Prestained Protein Ladder (Fermentas) on the membrane. The membrane was then treated as outlined in **Table 8.11**.

Table 8.11 Treatment of western blotting membrane after protein transfer

1. Incubate overnight at 4^{o}C in blocking buffer (PBS, 0.1% v/v Tween, 2% w/v milk).
2. Incubate for 1 h at room temperature in blocking buffer containing the specified dilution of primary antibody
3. Wash for 3 x 15 min with blocking buffer.
4. Incubate for 1 h at room temperature in blocking buffer containing the specified dilution of secondary antibody.
5. Wash for 3 x 15 min with blocking buffer.
6. Wash for 3 x 5 min with H_2O.
7. Apply Amersham ECL western Blotting Analysis System (GE Healthcare) for 1 min.
8. Expose blot to Amersham Hyperfilm ECL (GE Healthcare) in the dark for specified amount of time and develop using Xograph developer machine.

8.3 NMR experiments

Table 8.12 Key parameters of NMR spectra collected in this study

Spectrum	Experiment	Sample	Nuclei Measured	Key Experimental Parameters			Chapter
				Number of scans	Total data points (real + imaginary)	Sweepwidth / ppm	
^{1}H,^{15}N-HSQC (2D)	Characterisation of MxiG-N_{1-125}	3 mM MxiG-N_{1-126} in 25 mM Na_2HPO_4 (pH 6.8)	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 1024 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 140	$^{1}H^{N}$ = 16 ^{15}N = 32	2
^{1}H,^{15}N-HSQC (2D)	Assigned amides of MxiG-N_{6-112}	2.5 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 16 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 120	$^{1}H^{N}$ = 16 ^{15}N = 32	2
HN(CO)CA (3D)	Backbone ^{13}C assignment of MxiG-N_{6-112} (triple resonance)	2.5 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i), $^{13}C^{\alpha}$ (i-1)	$^{1}H^{N}$ = 4 ^{15}N = 16 ^{13}C = 16	$^{1}H^{N}$ = 2048 ^{15}N = 32 ^{13}C = 128	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{13}C = 32	2
CBCA(CO)NH (3D)	Backbone & side-chain ^{13}C assignment of MxiG-N_{6-112} (triple resonance)	2.5 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i), $^{13}C^{\alpha}$ (i-1), $^{13}C^{\beta}$ (i-1) N/Q side-chain: ^{1}H, ^{15}N, ^{13}C	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{13}C = 16	$^{1}H^{N}$ = 2048 ^{15}N = 32 ^{13}C = 108	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{13}C = 65	2

CBCANH (3D)	Backbone ^{13}C assignment of MxiG-N_{6-112} (triple resonance)	2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i), $^{13}C^{\alpha}$ (i, i-1), $^{13}C^{\beta}$ (i, i-1)	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{13}C = 16	$^{1}H^{N}$ = 2048 ^{15}N = 32 ^{13}C = 108	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{13}C = 65	2
CC(CO)NH (3D)	Backbone & side-chain ^{13}C assignment of MxiG-N_{6-112} (triple resonance)	2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i) Side-chain: all ^{13}C (i-1)	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{13}C = 16	$^{1}H^{N}$ = 2048 ^{15}N = 32 ^{13}C = 100	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{13}C = 75	2
HNCO (3D)	Backbone carbonyl assignment of MxiG-N_{6-112}	2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i), $^{13}C'$ (i-1)	$^{1}H^{N}$ = 2 ^{15}N = 16 $^{13}C'$ = 16	$^{1}H^{N}$ = 2048 ^{15}N = 32 $^{13}C'$ = 128	$^{1}H^{N}$ = 16 ^{15}N = 32 $^{13}C'$ = 15	2
HNHA (3D)	H^{α} assignment & calculation of $^{3}J_{HNH\alpha}$ values for MxiG-N_{6-112}	2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N, $^{1}H^{\alpha}$	$^{1}H^{N}$ = 4 ^{15}N = 16 $^{1}H^{\alpha}$ = 16	$^{1}H^{N}$ = 2048 ^{15}N = 44 $^{1}H^{\alpha}$ = 120	$^{1}H^{N}$ = 16 ^{15}N = 32 $^{1}H^{\alpha}$ = 16	2
$^{1}H,^{15}N$-NOESY-HSQC (3D)	Backbone ^{1}H assignment (sequential approach) & NOE identification for MxiG-N_{6-112}	2.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N, all ^{1}H within 7 Å	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{1}H = 16	$^{1}H^{N}$ = 2048 ^{15}N = 44 ^{1}H = 160	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{1}H = 16	2
HBHA(CO)NH (3D)	Side-chain ^{1}H assignment of MxiG-N_{6-112}	6.7 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8)	Backbone: $^{1}H^{N}$ (i), ^{15}N (i) Side-chain: $^{1}H^{\alpha}$ (i-1), $^{1}H^{\beta}$ (i-1)	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{1}H = 16	$^{1}H^{N}$ = 2048 ^{15}N = 44 ^{1}H = 128	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{1}H = 16	2

H(CCO)NH (3D)	Side-chain ^{1}H assignment of MxiG-N_{6-112}	6.7 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ (i), ^{15}N (i) Side-chain: all ^{1}H (i-1)	$^{1}H^{N}$ = 8 ^{15}N = 16 ^{1}H = 16	$^{1}H^{N}$ = 2048 ^{15}N = 40 ^{1}H = 128	$^{1}H^{N}$ = 16 ^{15}N = 32 ^{1}H = 16	2
HCCH-TOCSY (3D)	Side-chain ^{1}H assignment of MxiG-N_{6-112}	4 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT in >98% D_2O	^{1}H, ^{13}C to all ^{1}H in spin-coupled system	^{1}H = 8 ^{13}C = 16 ^{1}H = 16	^{1}H = 2048 ^{13}C = 64 ^{1}H = 220	^{1}H = 16 ^{13}C = 75 ^{1}H = 16	2
^{1}H,^{15}N-HSQC (2D)	Backbone amide proton exchange measurements for MxiG-N_{6-112}	0.5 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT in >98% D_2O	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N (slowly-exchanging only)	$^{1}H^{N}$ = 8 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 120	$^{1}H^{N}$ = 16 ^{15}N = 32	2
^{1}H,^{15}N-HSQC (CLEANEX-PM) (2D)	Backbone amide saturation transfer measurements for MxiG-N_{6-112}	5.4 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N (solvent-exposed only)	$^{1}H^{N}$ = 4 ^{15}N = 1	$^{1}H^{N}$ = 2048 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 37.6	2
Long-range H(N)CO-TROSY (2D)	H-bond identification for MxiG-N_{6-112}	5.4 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$ and $^{13}C'$ of H-bond partners	$^{1}H^{N}$ = 128 $^{13}C'$ = 1	$^{1}H^{N}$ = 2048 $^{13}C'$ = 80	$^{1}H^{N}$ = 16 $^{13}C'$ = 18	2
^{1}H,^{13}C-NOESY-HSQC (3D)	NOE identification for MxiG-N_{6-112}	4 mM $^{13}C/^{15}N$-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT in >98% D_2O	^{1}H, ^{13}C to all ^{1}H within 7 Å	^{1}H = 8 ^{13}C = 16 ^{1}H = 16	^{1}H = 2048 ^{13}C = 64 ^{1}H = 220	^{1}H = 16 ^{13}C = 75 ^{1}H = 16	2

^{1}H,^{15}N-HSQC (with & without saturation of ^{1}H magnetization) (2D)	Measurement of steady state ^{1}H-^{15}N NOE for MxiG-N_{6-112}	1.3 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 32 ^{15}N = 16	$^{1}H^{N}$ = 2048 ^{15}N = 120	$^{1}H^{N}$ = 16 ^{15}N = 32	2
^{1}H,^{15}N-HSQC (inversion-recovery) (2D)	Measurement of T1 values for MxiG-N_{6-112}	1.3 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 32 ^{15}N = 16	$^{1}H^{N}$ = 2048 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	2
^{1}H,^{15}N-HSQC (CPMG) (2D)	Measurement of T2 values for MxiG-N_{6-112}	1.3 mM ^{13}C/^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 32 ^{15}N = 16	$^{1}H^{N}$ = 2048 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	2
^{1}H,^{15}N-HSQC (2D)	Titration of MxiG-N_{6-112} with 50 mM P-Thr	138 µM ^{15}N-MxiG-N_{6-112} in 100 mM Tris-HCl (pH7.5), 150 mM NaCl, 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 8 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 100	$^{1}H^{N}$ = 16 ^{15}N = 32	3
1D ^{1}H spectra	Detection of Spa33 P_5 peptide in solution	Saturated solutions of P_5 in 100 mM Tris-HCl (pH 7.5), 150 mM NaCl	All ^{1}H	^{1}H = 64	^{1}H = 16384	^{1}H = 20.7	3
^{1}H,^{15}N-HSQC (2D)	Titration of MxiG-N_{6-112} with Spa33 P_5 peptide	Various - as specified in **Figure 3.11** & **Figure 3.12**.	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 8 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	3
^{1}H,^{15}N-HSQC (2D)	Titration of MxiG-N_{6-112} with 1 mM $MxiJ_{232-241}$	220 µM ^{15}N-MxiG-N_{6-112} in 25 mM Na_2HPO_4 (pH 6.8), 1 mM DTT	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 8 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 100	$^{1}H^{N}$ = 16 ^{15}N = 32	3

^{1}H,^{15}N-HSQC (2D)	Characterisation of MxiG-$C_{161-362/4}$	240 µM ^{15}N-MxiG-$C_{161-362/4}$ in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 24 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	4
^{1}H,^{15}N-HSQC (2D)	Characterisation of $MxiJ_{19-206}$	150 µM ^{15}N-$MxiJ_{19-206}$ in 25 mM Na_2HPO_4 (pH 6.8)	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 24 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 100	$^{1}H^{N}$ = 16 ^{15}N = 32	4
^{1}H,^{15}N-HSQC (2D)	Characterisation of $MxiJ_{19-100}$	200 µM ^{15}N-$MxiJ_{19-100}$ in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 48 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	4
^{1}H,^{15}N-HSQC (2D)	Titration of ^{15}N-$MxiJ_{19-206}$ with MxiG-$C_{161-362/4}$	50 µM ^{15}N-$MxiJ_{19-206}$ in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 48 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	4
^{1}H,^{15}N-HSQC (2D)	Titration of ^{15}N-MxiG-$C_{161-362/4}$ with $MxiJ_{19-206}$	50 µM ^{15}N-MxiG-$C_{161-362/4}$ in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 48 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 128	$^{1}H^{N}$ = 16 ^{15}N = 32	4
^{1}H,^{15}N-HSQC (2D)	Characterisation of $Spa33_{220-293}$	788 µM ^{15}N-$Spa33_{220-293}$ in 100 mM Tris-HCl (pH 7.5), 150 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 96 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 512	$^{1}H^{N}$ = 16 ^{15}N = 32	6
^{1}H,^{15}N-HSQC (2D)	Titration of ^{15}N-$Spa33_{220-293}$-His with $MxiN_{1-18}$	50 µM ^{15}N-$Spa33_{220-293}$-His in 100 mM Tris-HCl (pH 7.5), 150 mM NaCl	Backbone: $^{1}H^{N}$, ^{15}N N/Q/R side-chain: ^{1}H, ^{15}N	$^{1}H^{N}$ = 256 ^{15}N = 16	$^{1}H^{N}$ = 1024 ^{15}N = 192	$^{1}H^{N}$ = 16 ^{15}N = 32	6

8.4 Additional Work

8.4.1 Cloning and Expression of the MxiD Periplasmic Domain

Using the ACEMBL *Multi*coli expression system (ATG:biosynthetics), constructs were designed to enable both MxiD-His expression and MxiD/His-MxiG-$C_{142-371}$ co-expression (**Table 8.13**). Although all inserts were able to be successfully amplified by the touch-down PCR protocol and ligation-independent, ligation and In-Fusion cloning protocols were attempted, only pDC-*his-mxiGC*$_{142-371}$, pACEI-*mxiD*$_{1-177}$ and pACEI-*mxiD*$_{1-177}$*-his* were successfully cloned (**Table 8.13, yellow**). *E. coli* DH5α transformants for the remaining MxiD constructs were unable to be propagated in overnight culture, suggesting leaky expression of the domain was toxic to the host and required the N-terminal signal peptide for periplasmic export.

Table 8.13 Design of constructs for the expression of the MxiD periplasmic domain

Construct	Rationale
pACEI-*mxiD*$_{1-177}$	N-terminal signal peptide; encompasses region equivalent to $EscC_{21-174}$ structure (Spreter et al., 2009); no tag for co-expression with His-MxiG-$C_{142-371}$.
pACEI-*mxiD*$_{1-177}$*-his*	N-terminal signal peptide; encompasses region equivalent to $EscC_{21-174}$ structure (Spreter et al., 2009); C-terminal His-tag for expression alone.
pACEI-*mxiD*$_{22-177}$	Encompasses region equivalent to $EscC_{21-174}$ structure (Spreter et al., 2009); no tag for co-expression with His-MxiG-$C_{142-371}$.
pACEI-*mxiD*$_{22-177}$*-his*	Encompasses region equivalent to $EscC_{21-174}$ structure (Spreter et al., 2009); C-terminal His-tag for expression alone.
pACEI-*mxiD*$_{22-259}$	Encompasses region equivalent to T2SS GspD structure (Korotkov et al., 2009); no tag for co-expression with His-MxiG-$C_{142-371}$.
pACEI-*mxiD*$_{22-259}$*-his*	Encompasses region equivalent to T2SS GspD structure (Korotkov et al., 2009); C-terminal His-tag for expression alone.
pDC-*his-mxiGC*$_{142-371}$	Complete periplasmic domain of MxiG; N-terminal His-tag for co-expression with untagged MxiD constructs; does an interaction with MxiD stabilise MxiG-C in the cytoplasm?

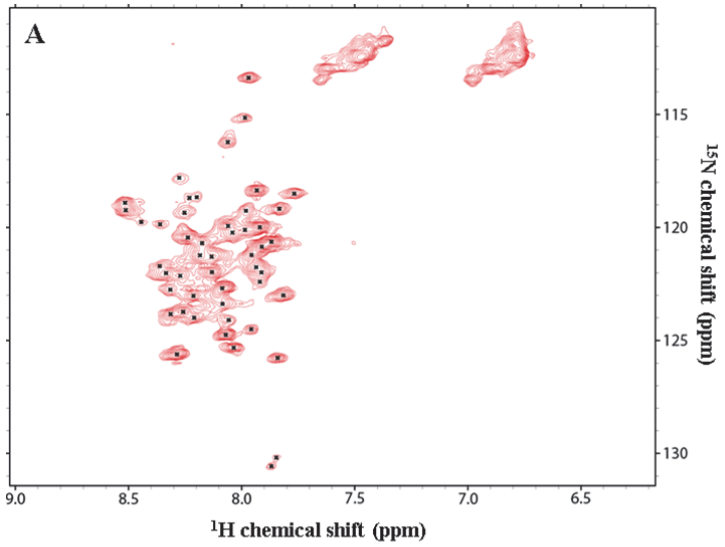
Cre-Lox recombination of pDC-*his-mxiGC*$_{142-371}$ and pACEI-*mxiD*$_{1-177}$ resulted in the formation of a pACEI-*mxiD*$_{1-177}$/pDC-*his-mxiGC*$_{142-371}$ construct. Despite testing various strains of *E. coli*, pACEI-*mxiD*$_{1-177}$-*his* and pACEI-*mxiD*$_{1-177}$/pDC-*his-mxiGC*$_{142-371}$ constructs were only able to be successfully transformed into *E. coli* L56 cells (derived from the Lemo strain; donated by Stockholm University), with growth of colonies after 48 h at 37^{o}C. Expression screening was subsequently carried out with these constructs, in each case testing 5 ml LB & Terrific Broth (Fisher) growth media, 0.1 mM & 1 mM IPTG concentrations and 21^{o}C & 37^{o}C induction temperatures. SDS-PAGE analysis showed no detectable binding of proteins in the soluble fraction to Ni-NTA magnetic agarose beads (Qiagen). A subsequent large-scale expression (2 x 1 l TB, 0.1 mM IPTG, 21^{o}C induction) and Ni-affinity purification of MxiD$_{1-177}$-His and MxiD$_{1-177}$/His-MxiG-C$_{142-371}$ confirmed no detectable overexpression and purification of the expected proteins, suggesting instability and toxicity of the MxiD periplasmic domain in *E. coli* L56 cells. A pCOLA-*mxiD*$_{1-177}$ *-His/mxiM* construct, created by site-directed mutagenesis of pCOLA-*mxiD/mxiM* (donated by M. Cottee), also did not allow overexpression of MxiD$_{1-177}$-His, indicating the toxicity was likely due to the nature of the protein and not the absence of MxiM or the vector used for expression.

8.4.2 Characterisation of MxiIΔ5

The *S. flexneri* inner rod protein MxiI with a truncation of the five C-terminal residues and a C-terminal LEHHHHHH tag (MxiIΔ5) was expressed from a pET21b-*mxiIΔ5* construct (donated by J. Deane). MxiIΔ5 was purified from the soluble fraction by Ni-affinity chromatography and SEC, typically resulting in 500 µl a 100 µM sample of the 11.2 kDa protein. Addition of a 2-fold molar excess of MxiIΔ5 to 50 µM ^{15}N-MxiJ$_{19-206}$ (chapter 4) resulted in no change in peak positions in the ^{1}H,^{15}N-HSQC spectrum (not shown), suggesting the isolated inner rod subunit does not undergo a high affinity interaction with the inner ring subunit from the IMR.

The ^{1}H,^{15}N-HSQC spectrum for ^{15}N-MxiIΔ5 showed a poor peak dispersion in both dimensions, allowing only 49 putative backbone amide peaks to be unambiguously picked for a 100 residue

construct (**Figure 8.1A**). Subsequently, the CD spectrum of 9 µM MxiIΔ5 showed a prominent peak at 200 nm and a slight dip at 220 nm (**Figure 8.1B**). Together these results indicated that isolated MxiIΔ5 had some α-helical content but was predominantly unfolded, suggesting the protein may have to self-polymerise within the inner rod to form a defined tertiary structure. This was in agreement with NMR spectroscopy and CD performed on the *S. typhimurium* inner rod protein PrgJ purified from the insoluble fraction, which ascertained that the only structured portion of the protein was an 18-residue C-terminal α-helix, allowing joint publication of these results (Zhong et al., 2012). Therefore, unlike the needle subunit, studies of the inner rod subunit in isolation cannot be used to provide insight into the structure and function of the intact structure found within the T3SS basal body.



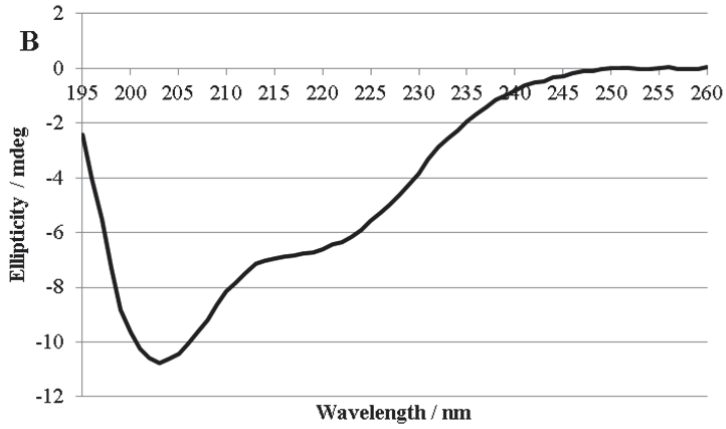


Figure 8.1 *MxiIΔ5 is partially unfolded*.

A $^{1}H,^{15}N$-HSQC spectrum for 96 µM ^{15}N-MxiIΔ5 in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5% D_2O. Putative peaks for backbone amides are indicated by black dots.

B CD spectrum for 9 µM MxiIΔ5 in 10 mM Na_2HPO_4 (pH 7.5), 150 mM NaF acquired at 25°C with four repeats of wavelength scans between 195-260 nm.

8.4.3 SPR Analysis of the DNA Binding Properties of *N. meningitidis* ComP

The type IV pilus of *Neisseria meningitidis* facilitates the selective uptake of DNA containing specific uptake sequences (DUS), required for subsequent incorporation into the bacterial genome and natural transformation[1]. The minor pilin ComP was identified as the elusive DNA-binding receptor and purified for use in *in vitro* interaction studies (Dr A. Cehovin, Imperial College, London). This enabled the relative binding of ComP to 10 bp DUS (ATGCCGTCTGAA) and scrambled SUD (ACGACTTATAAT) and SDU (AAGGCCTGTCAT) sequences to be probed by SPR on a BIAcore 3000 instrument at 25^{o}C. Biotinylated DUS, SUD and SDU ds primers were coupled to streptavidin on the sensor surface (SA chip, GE Healthcare) at equivalent levels (250-300 response unit (RU) increase) and an uncouple channel was left for the collection of a control trace. ComP was then flowed over the chip surface at a flow-rate of 40 µl/min, followed by injection of 10 µl 0.5 M NaCl to remove residual non-specific interactions. Over a 1-30 µM concentration range, ComP showed dose-dependent and saturable binding to the DUS sequence (**Figure 8.2A, black**), with an average K_d of 29 ± 8 µM calculated from six kinetic fits (**Table 8.14**). In contrast, lower equilibrium RU values (R_{eq}) demonstrated lower affinity binding to the SUD and SDU sequences (**Figure 8.2A, purple & red**) that did not reach saturation < 100 µM ComP. Therefore, ComP can specifically recognise the DUS sequence, supporting the proposal that it acts as the receptor for DNA uptake by *N. meningitidis*.

Table 8.14 Kinetic fits of SPR curves of ComP binding to DUS, SUD and SDU sequences.

Dilution series	Repeat	k_{on} / 10^4 1/Ms	k_{off} / 1/s	K_d / uM	Chi
1 (30, 20, 13.3, 8.9, 5.9, 3.9, 1 µM)	1	1.2	0.32	26.0	0.53
	2	9.2	2.7	29.5	1.11
2 (30, 20, 13.3, 8.9, 5.9, 3.9, 1 µM)	1	6.5	1.5	22.5	1.21
	2	2.4	0.50	20.6	0.98
3 (50, 25, 12.5, 6.25, 4, 1 µM)*	1	0.18	0.077	43.0	0.47
	2	0.18	0.061	34.1	1.18

*Denotes dilution series carried out on a different chip with a different batch of purified ComP.

In addition, structural characterisation of ComP and its interaction with DUS sequences by NMR spectroscopy had suggested positively charged residues involved in DNA binding[2]. Indeed, ComP single point mutants showed a lower binding affinity to DUS relative to the wild-type protein, corroborating the involvement of these residues in DUS binding (**Figure 8.2B**).

1 Goodman, S. D. & Scocca, J. J, (1988) Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. USA* **85**, 6982-6986.

2 Cehovin, A., Simpson, P. J., McDowell, M. A., Brown, D. R., Noschese, R., Pallett, M., Brady, J., Baldwin, G. S., Lea, S. M., Matthews, S. J., Pelicic, V., A minor pilin enables competent *Neisseria* to recognise their own DNA during natural transformation. Unpublished.

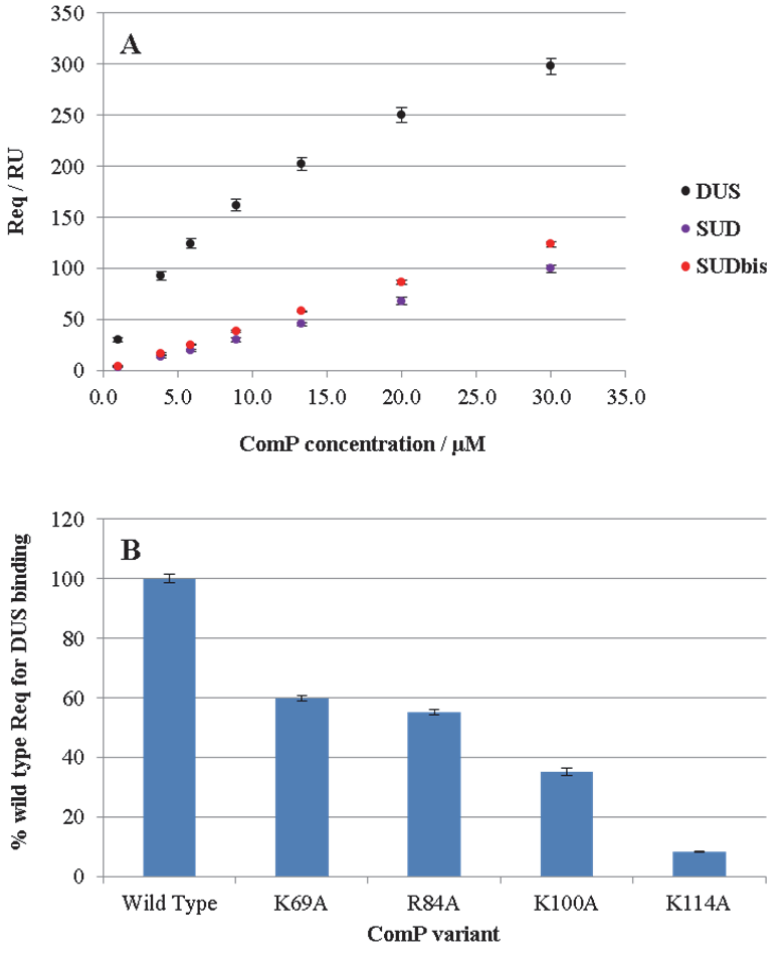


Figure 8.2 *SPR analysis of the DNA binding properties of N. meningitidis ComP.*

A R_{eq} values for 1-30 µM ComP binding to ds DUS, SUD and SDU DNA sequences. Values shown are the average of 6 measurements taken from 2 repeats each of 3 separate dilution series. 267 RU DUS, 262 RU SUD and 264 RU SDU were present on the surface of the SA chip.

B R_{eq} values for 30 µM ComP variants binding to DUS shown as a percentage of R_{eq} for wild-type binding. Values shown are the average of 6 measurements taken from 2 repeats each of 3 separate dilutions. The point mutations are of positively charged residues implicated to be in the ComP DNA binding site from NMR spectroscopy.