STRUCTURAL CHARACTERIZATION OF TYPE IV PILUS BIOGENESIS PROTEINS

A thesis submitted to The University of Manchester for the degree of PhD in the Faculty of Life Sciences

2012

Jamie-lee Berry
List of contents

Abstract………………………………………………………………………………………………10
Declaration and copyright statement…………………………………………………………11
Acknowledgements……………………………………………………………………………12
Abbreviations……………………………………………………………………………………13

Chapter 1: Introduction
1.1 Bacterial secretion systems…………………………………………………………………………………16
1.2 The chaperone/ usher pathway – P pili and type I pili……………………………………………………17
   - Two-usher model for pilus biogenesis………………………………………………………18
   - Single-usher model for pilus biogenesis………………………………………………20
1.3 Type IV secretion – the conjugative pilus…………………………………………………………………22
   - Type IVB secretion system…………………………………………………………………25
1.4 The secretin family of outer membrane proteins………………………………………………………26
   - Secretin accessory factors…………………………………………………………………29
1.5 Type III secretion – The injectisome………………………………………………………………………32
   - The needle complex………………………………………………………………………..33
   - The translocon……………………………………………………………………………..38
1.6 Type II secretion and type IV pilus biogenesis……………………………………………………………40
1.7 The pilus and pseudopilus appendage……………………………………………………………………44
   - Prepilin processing………………………………………………………………………..44
   - The type IV pilus fibre……………………………………………………………………44
   - The T2SS pseudopilus……………………………………………………………………48
1.8 Inner membrane components of the T2SS and T4PB system…………………………………………49
   - A multi-protein complex is formed at the inner membrane of the T2SS………………49
   - The inner membrane complex of the T4PB system…………………………………..51
   - The polytopic inner membrane platform protein…………………………………….53
1.9 T2SS and T4PB system ATPases…………………………………………………………………………56
1.10 GspC and PilP proteins…………………………………………………………………………………59
1.11 Aims and objectives of the project……………………………………………………………………63

Chapter 2: Materials and Methods
2.1 Cloning and expression materials………………………………………………………………………64
2.2 Composition of buffers and reagents…………………………………………………………………67
   2.2.1 Chemicals………………………………………………………………………………67
   2.2.2 Cloning enzymes……………………………………………………………………..67
   2.2.3 Antibiotics and inducers……………………………………………………………..67
   2.2.4 Growth media………………………………………………………………………..67
   2.2.5 Composition of M9 minimal media…………………………………………………68
   2.2.6 Buffers and reagents for agarose gel electrophoresis………………………………68
## 2.3 Cloning protein production and purification protocols

<table>
<thead>
<tr>
<th>2.3.1 Transformation procedure</th>
<th>2.3.2 PCR amplification of genomic sequences</th>
<th>2.3.3 Agarose gel electrophoresis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.4 Preparation of vector and target sequence</td>
<td>2.3.5 Ligation of constructs</td>
<td>2.3.6 Protein production trials</td>
</tr>
<tr>
<td>2.3.7 Large-scale protein production</td>
<td>2.3.8 Cell lysis</td>
<td>2.3.9 Immobilized Metal-ion Affinity Chromatography (IMAC)</td>
</tr>
<tr>
<td>2.3.10 Dialysis buffer-exchange</td>
<td>2.3.11 Desalting column buffer-exchange</td>
<td>2.3.12 Thrombin treatment</td>
</tr>
<tr>
<td>2.3.13 Concentration of purified protein</td>
<td>2.3.14 Size-exclusion chromatography (SEC)</td>
<td>2.3.15 SEC with Multi-angle Laser Light Scattering (SEC-MALLS)</td>
</tr>
</tbody>
</table>

## 2.4 High-throughput cloning and protein production at OPPF

<table>
<thead>
<tr>
<th>2.4.1 High-throughput PCR</th>
<th>2.4.2 AMPure magnetic bead PCR purification</th>
<th>2.4.3 InFusion™ ligation reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.4 High-throughput transformation</td>
<td>2.4.5 High-throughput culture and plasmid recovery</td>
<td>2.4.6 High-throughput protein production trials</td>
</tr>
<tr>
<td>2.4.7 Ni²⁺-NTA robotic protein production screening</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## 2.5 Analysis of proteins

<table>
<thead>
<tr>
<th>2.5.1 SDS-PAGE</th>
<th>2.5.2 Western blot analysis</th>
<th>2.5.3 Measuring protein concentration</th>
</tr>
</thead>
</table>

## 2.6 Protein structural studies

<table>
<thead>
<tr>
<th>2.6.1 Preparation of crystal trials and cryoprotection of crystals</th>
<th>2.6.2 Data acquisition</th>
<th>2.6.3 Preparation of ¹⁵N- and ¹³C-labelled protein samples for NMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.4 Data collection</td>
<td>2.6.5 NMR solution structure calculation &amp; validation</td>
<td>2.6.6 NmPiiQ&lt;sub&gt;345-545&lt;/sub&gt; model</td>
</tr>
</tbody>
</table>

## 2.7 Protein interaction studies
2.7.1 Isothermal titration calorimetry (ITC).................................................................92
2.7.2 Surface plasmon resonance....................................................................................92
2.7.3 NMR chemical shift perturbation (NMR-CSP).....................................................92

Chapter 3: High-throughput production, purification and characterization of individual subdomains from the type IV pilus dependent secretin PilQ
3.1 Introduction..............................................................................................................94
3.2 Domain organization of secretins...........................................................................95
3.3 Selection of targets..................................................................................................98
3.4 High-throughput cloning and protein production .................................................103
  - Soluble protein production with IPTG induction......................................................104
  - Soluble protein production with auto-induction medium........................................108
3.5 Other PilQ constructs from Neisseria and Pseudomonas........................................112
3.6 Neisseria meningitidis KH-type domain production and purification.........................114
3.7 Analysis of the multimeric state of NmPilQ^{343-545}............................................116
3.8 Characterization of NmPilQ^{343-545} by NMR......................................................118
3.9 PilQ beta-domains..................................................................................................120
  - Characterization of beta-domains by NMR..............................................................122
  - Characterization of a B1/2-domain (NmPilQ^{24-329})............................................124
3.10 Discussion.............................................................................................................126

Chapter 4: Structural studies of domains from the type IV pilus-dependent secretin PilQ
4.1 Introduction.............................................................................................................131
4.2 Assignment of NmPilQ^{343-545}...........................................................................134
4.3 Production and purification of NmPilQ^{343-442}....................................................141
4.4 Structure determination of NmPilQ^{343-442}.........................................................143
4.5 Model for NmPilQ^{343-545} by CS-ROSETTA and homology modelling..................149
4.6 Structure determination of a B2 domain from PilP by NMR (NmPilQ^{225-329})........152
4.7 Assignment of a B1 domain from Aeromonas (AhPilQ^{26-130}).............................158
4.8 Discussion..............................................................................................................160

Chapter 5: Structural analysis of the secretin-interacting lipoprotein PilP by X-ray crystallography
5.1 Introduction .............................................................................................................162
5.2 Similarity of PilP to GspC ......................................................................................164
5.3 Domain organization of PilP proteins......................................................................166
5.4 Production and Purification of PilP proteins...........................................................168
  - Protein production trials..........................................................................................168
  - Purification of NmPilP^{C}.........................................................................................170
## Contents

- Purification of PaPilP\textsuperscript{C} ............................................................................................................. 171

5.5 Analysis of oligomeric state of NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C} .................................................. 173

5.6 Structural studies of NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C} by X-ray Crystallography .................................. 176
  - Crystallization and data collection .................................................................................................................. 176
  - Structure determination of PaPilP\textsuperscript{C} .............................................................................................. 181

5.7 Discussion .................................................................................................................................................. 185
  - Comparison to NmPilP NMR structure (2IVW) .................................................................................. 185
  - Comparison to PaPilP NMR structure (2LC4) .................................................................................. 185
  - Comparison to GspC of the T2SS ............................................................................................................. 188
  - Conclusions and future direction ............................................................................................................... 190

## Chapter 6: Establishing protein-protein interactions within the type IV pilus biogenesis system: PilP, PilQ and PilE

6.1 Introduction ................................................................................................................................................ 192

**Part A: Analysis of interaction between NmPilQ\textsuperscript{343-545} and NmPilP\textsuperscript{C}**

6.2 Nuclear magnetic resonance chemical shift perturbation (NMR-CSP) ............................................. 194

6.3 Structural model for the complex formed by NmPilP\textsuperscript{C} / NmPilQ\textsuperscript{343-545} ................................. 199

6.4 Isothermal titration calorimetry .................................................................................................................. 203

6.5 Surface plasmon resonance ........................................................................................................................ 205

6.6 Probing an interaction between NmPilP\textsuperscript{C} and the PilQ B2-domain .............................................. 208

**Part B: Investigation of an interaction between PilQ and the pilin subunit, PilE by NMR**

6.7 Production and purification of the pilus subunit, PilE ........................................................................... 209

6.8 NMR titration experiment: PilE and \textsuperscript{15}N-NmPilQ\textsuperscript{343-545} .............................................................. 212

6.9 Discussion .................................................................................................................................................. 213

## Chapter 7: General Discussion .................................................................................................................. 216

**Appendix** .................................................................................................................................................. 230

**References** .................................................................................................................................................. 245
List of figures

Figure 1.2.1: PapC usher and model for P pilus assembly .............................................................. 19
Figure 1.2.2: Structure of the FimD usher during substrate translocation ............................................. 21
Figure 1.3.1: Substrate transition by type IV secretion ........................................................................... 24
Figure 1.4.1: Crystal structures of T2SS and T3SS secretin N-terminal domains .................................. 28
Figure 1.4.2: Low-resolution electron microscopy secretin structures ................................................... 28
Figure 1.4.3: Crystal structures of secretin accessory factors ................................................................. 31
Figure 1.5.1: Architecture of the T3SS needle assembly ......................................................................... 36
Figure 1.5.2: Schematic representation of the T3SS basal body ............................................................ 37
Figure 1.6.1: The type II secretion system and type IV pilus biogenesis system ..................................... 41
Figure 1.7.1: Structure of the type IV pilus subunit and model for the assembled pilus ......................... 47
Figure 1.9.1: Crystal structure of PilT from Aquifex aeolicus .................................................................. 58

Figure 3.2.1: Schematic representation of the domain organization of secretins ..................................... 96
Figure 3.3.1: Alignment and secondary structure prediction of PilQ protein sequences ....................... 100
Figure 3.4.1: Soluble protein production using IPTG ............................................................................... 106
Figure 3.4.2: Soluble protein production using auto-induction medium ............................................... 109
Figure 3.5.1: Protein production trials of PilQ domains from Neisseria and Pseudomonas .................. 113
Figure 3.6.1: IMAC purification and thrombin protease treatment of NmPilQ 343-545 ......................... 115
Figure 3.7.1: Analysis of the multimeric state of NmPilQ 343-545 ......................................................... 117
Figure 3.8.1: Characterization of NmPilQ 343-545 by NMR ................................................................. 119
Figure 3.9.1: Sequence alignment of PilQ beta domains produced in this study ................................. 121
Figure 3.9.2: Characterization of PilQ beta-domains by NMR .............................................................. 123
Figure 3.9.3: IMAC purification and SEC of NmPilQ 24-329 ................................................................. 125

Figure 4.1.1: Schematic illustration of secretin domain organization ................................................... 133
Figure 4.2.1: 1H-15N-HSQC spectra for NmPilQ 343-545 ....................................................................... 136
Figure 4.2.2: Secondary structure content of NmPilQ 343-545 ............................................................... 139
Figure 4.3.1: Purification of NmPilQ 343-442 by size-exclusion chromatography ............................... 142
Figure 4.4.1: Assigned $^1$H$^{15}$N-HSQC spectra for NmPilQ$^{343-442}$ ......................................................... 144
Figure 4.4.2: Solution structure and structural homologues of NmPilQ$^{323-442}$ ..................................................... 148
Figure 4.5.1: Models for the complete NmPilQ$^{343-545}$ N0/N1 structure .......................................................... 151
Figure 4.6.1: Assigned $^1$H$^{15}$N-HSQC for NmPilQ$^{224-329}$ .......................................................... 153
Figure 4.6.2: NMR structure of NmPilQ$^{224-329}$ and structural homologues ..................................................... 156
Figure 4.6.3: ConSurf analysis of NmPilQ$^{224-329}$ .......................................................... 157
Figure 4.7.1: Chemical shift and secondary structure assignment of AhPilQ$^{26-130}$ ................................................ 159

Figure 5.2.1: Domain structure of PilP and GspC Proteins .......................................................... 165
Figure 5.3.1: Alignment of PilP protein sequences .......................................................... 167
Figure 5.4.1: Small-scale protein production trials of NmPilP$^C$ and PaPilP$^C$ ..................................................... 169
Figure 5.4.2: IMAC purification of NmPilP$^C$ and PaPilP$^C$ .......................................................... 172
Figure 5.5.1: Size Exclusion Chromatography of NmPilP$^C$ .......................................................... 174
Figure 5.5.2: SEC of NmPilP$^C$ and PaPilP$^C$ using a Superdex 75 column ..................................................... 175
Figure 5.5.3: SEC-MALLS analysis of NmPilP$^C$ using a Superdex 75 column ..................................................... 175
Figure 5.5.1: PaPilP$^C$ crystals produced in this study .......................................................... 178
Figure 5.6.1: Electron density for the PaPilP$^C$ Pro155- Trp161 loop .......................................................... 180
Figure 5.6.3: Zinc atoms bound within the PaPilP$^C$ C2 crystal .......................................................... 180
Figure 5.6.4: Ramachandran plots for PaPilP$^C$ .......................................................... 180
Figure 5.6.5: Structural features of PaPilP$^C$ .......................................................... 183
Figure 5.6.6: Tryptophan residue exposed on the surface of PaPilP$^C$ and NmPilP .......................................................... 183
Figure 5.6.7: Electrostatic surface of PaPilP$^C$ and NmPilP .......................................................... 184
Figure 5.6.8: ConSurf analysis of PaPilP$^C$ .......................................................... 184
Figure 5.7.1: Comparison of NmPilP and PaPilP$^C$ .......................................................... 187
Figure 5.7.2: Comparison of Pseudomonas PilP NMR structure and PaPilP$^C$ .......................................................... 187
Figure 5.7.3: Comparison of PaPilP$^C$ and GspC-HR .......................................................... 189
Figure 5.7.4: Alignment of PilP$^C$ and GspC-HR sequences .......................................................... 189

Figure 6.2.1: Titration of unlabelled NmPilQ$^{343-545}$ into $^{15}$N-NmPilP$^C$ .......................................................... 195
Figure 6.2.2: Identification of the NmPilQ$^{343-545}$ binding site on NmPilP$^C$ .......................................................... 196
Figure 6.2.3: Identification of the NmPilP\textsuperscript{C} binding site on NmPilQ\textsuperscript{343-545} and NmPilQ\textsuperscript{343-442} ..........198
Figure 6.3.1: Structural model for the complex formed by NmPilP\textsuperscript{C}/NmPilQ\textsuperscript{343-545} ......................202
Figure 6.4.1: ITC experiments with NmPilP\textsuperscript{C} and NmPilQ\textsuperscript{343-545} ..................................................204
Figure 6.5.1: Analysis of the interaction between NmPilP\textsuperscript{C} and NmPilQ\textsuperscript{343-545} by SPR ..................207
Figure 6.6.1: Analysis of NmPilP\textsuperscript{C} and NmPilQ\textsuperscript{224-329} by NMR-CSP ........................................208
Figure 6.7.1: Production and purification of NmPilE\textsuperscript{36-170} ..........................................................211
Figure 6.7.2: SEC Analysis of NmPilE\textsuperscript{36-170} .................................................................................211
Figure 6.8.1: Analysis of NmPilQ\textsuperscript{343-545} and NmPilE\textsuperscript{36-170} by NMR-CSP .................................212

Figure 7.1: Cryo-EM reconstruction of NmPilQ ..............................................................220
Figure 7.2: Relative domain orientations in secretin N-terminal domains ....................223
Figure 7.3: Model for the periplasmic domains of NmPilQ ..............................................223
Figure 7.4: Model for the trans-periplasmic assembly formed by PilQ and PilP ............229
List of tables

Table 1.6: T2SS and T4PB system nomenclature.................................................................43
Table 2.1.1: Primers for constructs produced at OPPF..........................................................65
Table 2.1.2: Primers for constructs produced at the home laboratory.................................66
Table 2.1.3: E. coli cloning and protein production strains....................................................66
Table 2.1.4: Protein production vectors..................................................................................66
Table 2.1.5: Genomic DNA templates....................................................................................66
Table 3.4: High-throughput small-scale production of PilQ constructs..................................111
Table 3.5: PilQ constructs from Neisseria and Pseudomonas...................................................113
Table 4.6: NMR Structure calculation statistics for NmPilQ$^{224-329}$ ....................................145
Table 4.4: NMR Structure calculation statistics for NmPilQ$^{343-442}$ ....................................154
Table 5.6: Data collection statistics for PaPilP$^C$ crystals.......................................................179
Abstract

Type IV pili, or fimbriae, are long, thin proteinaceous appendages found on the surface of many well-known pathogens. They mediate a variety of important virulence functions for the organism, such as twitching motility, biofilm formation, uptake of genetic material and host cell recognition and adhesion. Pili are formed by the rapid polymerization and de-polymerization of the pilin subunit, and this is orchestrated by a complex macromolecular machine which spans the bacterial cell envelope, requiring a variety of gene products. The type IV pilus biogenesis system is closely related to the bacterial type II secretion system, one of six designated multi-protein cell envelope complexes which are dedicated to the specific secretion of exotoxins and virulence factors. Many of these secretion systems also produce fimbrial structures to facilitate the extrusion of their substrates or to communicate with the host. As they form crucial virulence factors, the secretion systems and the type IV pilus biogenesis system have become attractive potential antimicrobial targets and obtaining structural and functional information for the components of these systems is an important first step towards achieving this.

Type IV pili appear on the surface of bacteria through an outer membrane pore, PilQ, which is a member of the secretin family. Secretins are also found in the type II and III secretion systems, but the way in which they are regulated remains unclear. PilQ forms a dodecameric chamber in the outer membrane with a large vestibule which reaches into the periplasm, composed of its N-terminal domains. In this project, N-terminal domains from PilQ were produced in recombinant form and their structures determined by NMR. One of these domains revealed an eight-stranded beta-sandwich structure which appears to be unique to type IV pilus secretins and has not been structurally characterized before. Another revealed an alpha/beta type fold which is common to secretins of other systems. In the second part of this project, the interaction formed between the N-terminal alpha/beta domains of PilQ and an essential inner membrane-anchored lipoprotein, PilP, was probed by NMR chemical shift perturbation. Based on changes to the $^{15}$N-HSQC spectra the binding site was mapped onto each protein to produce a computational model for the complex formed between the two. Using a recent cryo-EM structure for the Neisseria PilQ dodecamer determined by colleagues, it was possible to model the PilQ N-terminal domains in complex with PilP into the electron density map. This produced a model for the trans-periplasmic assembly formed by PilQ and PilP in the type IV pilus biogenesis system, and led to the conclusion that the PilQ dodecamer needs to disassemble considerably at the base to accommodate a pilus fibre. The novel beta-domains might therefore function to gate or open the secretin, and PilP may play a role in stabilizing the secretin during this and serve to connect the outer and inner membrane system components.
Declaration

No portion of the work in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (http://www.campus.manchester.ac.uk/medialibrary/policies/intellectualproperty.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations: (http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of theses.
I would like to thank my supervisor Professor Jeremy Derrick for giving me the opportunity to pursue a PhD in his group, for being an insightful and enthusiastic supervisor throughout my project, and for sharing his expertise on all aspects.

I would like to thank Dr. Marie Phelan for her time and patience whilst teaching me the principles of NMR, how to record spectra and how to analyse data as well as many other protein biochemistry tips and techniques. Also importantly I thank Marie for her input and expertise with regards to the advanced data analysis and structure determination for the NMR aspects of the project, without which it would not have been possible. Also, thanks to Professor Lu-Yun Lian, for allowing me to use the facilities at the NMR centre and for her help and advice.

I would like to thank the members of my lab group and surrounding lab groups for their help and advice throughout the project: Dr. Muhammad Saleem, Dr. Vijaykumar Karuppiah, Darin Hassan, Angela Thistlethwaite, Zoe Caldwell, Dr. Rana Dajani, Dr. Muhammad Imran, Dr. James Birtley, Dr. Mark Young and Dr. Robert Cleverly.

I would like to thank everybody at OPPF for allowing me to use their facilities for high-throughput protein production and for their assistance.

I would like to thank the contributors to the publication resulting from this work: Dr. Marie Phelan, Dr. Richard F. Collins, Tomas Adomavicius, Professor Tone Tønjum, Stefan A. Frye, Dr. Louise Bird, Dr. Ray Owens, Professor Robert Ford, Professor Lu-Yun Lian and Professor Jeremy Derrick.

I would like to thank Professor Jeremy Derrick, The Faculty of Life Sciences and the BBSRC for obtaining the funding for the project.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine triphosphate</td>
<td>ATP</td>
</tr>
<tr>
<td>Amino-</td>
<td>N-</td>
</tr>
<tr>
<td>Ammonium persulphate</td>
<td>APS</td>
</tr>
<tr>
<td>Angstroms</td>
<td>Å</td>
</tr>
<tr>
<td>Base pairs</td>
<td>bp</td>
</tr>
<tr>
<td>Bicinchoninic acid</td>
<td>BCA</td>
</tr>
<tr>
<td>Bovine serum albumin</td>
<td>BSA</td>
</tr>
<tr>
<td>5-bromo-4-chloro-indolyl-β-D-galactopyranoside</td>
<td>X-gal</td>
</tr>
<tr>
<td>5-Bromo-4-chloro-3-indolyl phosphate</td>
<td>BCIP</td>
</tr>
<tr>
<td>Calf intestine alkaline phosphatase</td>
<td>CIAP</td>
</tr>
<tr>
<td>Carboxy-</td>
<td>C-</td>
</tr>
<tr>
<td>Chemical shift perturbation</td>
<td>CSP</td>
</tr>
<tr>
<td>Cryo-electronmicroscopy</td>
<td>cryo-EM</td>
</tr>
<tr>
<td>Daltons</td>
<td>Da</td>
</tr>
<tr>
<td>Degrees centigrade</td>
<td>°C</td>
</tr>
<tr>
<td>Deoxyribonuclease</td>
<td>DNase</td>
</tr>
<tr>
<td>Deoxyribonucleic acid</td>
<td>DNA</td>
</tr>
<tr>
<td>Electron microscopy</td>
<td>EM</td>
</tr>
<tr>
<td>Ethylenediaminetetraacetic acid</td>
<td>EDTA</td>
</tr>
<tr>
<td>Fast protein liquid chromatography</td>
<td>FPLC</td>
</tr>
<tr>
<td>Grammes</td>
<td>µg</td>
</tr>
<tr>
<td>Gravitational acceleration</td>
<td>g</td>
</tr>
<tr>
<td>Heteronuclear single quantum coherence</td>
<td>HSQC</td>
</tr>
<tr>
<td>Histidine (6x) tag</td>
<td>His-tag</td>
</tr>
<tr>
<td>Hydrogen ion concentration (negative log)</td>
<td>pH</td>
</tr>
<tr>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
<td>HEPES</td>
</tr>
<tr>
<td>Immobilized metal affinity chromatography</td>
<td>IMAC</td>
</tr>
<tr>
<td>Isoelectric point</td>
<td>IEP</td>
</tr>
<tr>
<td>Isopropyl- β-D-1-thiogalactopyranoside</td>
<td>IPTG</td>
</tr>
<tr>
<td>Isothermal titration calorimetry</td>
<td>ITC</td>
</tr>
<tr>
<td>Kilocalorie</td>
<td>kcal</td>
</tr>
<tr>
<td>Kilodaltons</td>
<td>kDa</td>
</tr>
<tr>
<td>Kelvin</td>
<td>K</td>
</tr>
<tr>
<td>Luria Bertani</td>
<td>LB</td>
</tr>
<tr>
<td>Microcalorie</td>
<td>µcal</td>
</tr>
<tr>
<td>Microgrammes</td>
<td>µg</td>
</tr>
<tr>
<td>Microlitres</td>
<td>µl</td>
</tr>
<tr>
<td>Micrometres</td>
<td>µm</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Mill-Absorption unit</td>
<td>mAu</td>
</tr>
<tr>
<td>Milliamperes</td>
<td>mA</td>
</tr>
<tr>
<td>Milligrammes</td>
<td>mg</td>
</tr>
<tr>
<td>Millilitres</td>
<td>ml</td>
</tr>
<tr>
<td>Millimetres</td>
<td>mm</td>
</tr>
<tr>
<td>Millimoles per litre</td>
<td>mM</td>
</tr>
<tr>
<td>Milliseconds</td>
<td>ms</td>
</tr>
<tr>
<td>Millivolts</td>
<td>mV</td>
</tr>
<tr>
<td>Molecular weight</td>
<td>Mw</td>
</tr>
<tr>
<td>Moles per litre</td>
<td>M</td>
</tr>
<tr>
<td>2-(N-morpholino)ethanesulphonic acid</td>
<td>MES</td>
</tr>
<tr>
<td>Nanogrammes</td>
<td>ng</td>
</tr>
<tr>
<td>Nanolitres</td>
<td>nl</td>
</tr>
<tr>
<td>Nanometres</td>
<td>nm</td>
</tr>
<tr>
<td>Neisseria meningitides</td>
<td>Nm</td>
</tr>
<tr>
<td>Nitro blue tetrazolium</td>
<td>NBT</td>
</tr>
<tr>
<td>Nuclear magnetic resonance</td>
<td>NMR</td>
</tr>
<tr>
<td>Nuclear overhauser effect spectroscopy</td>
<td>NOESY</td>
</tr>
<tr>
<td>Nuclear overhauser effect</td>
<td>NOE</td>
</tr>
<tr>
<td>One-dimensional</td>
<td>1D</td>
</tr>
<tr>
<td>Optical density at 600nm</td>
<td>OD$_{600}$</td>
</tr>
<tr>
<td>Oxford Protein Production Facility</td>
<td>OPPF</td>
</tr>
<tr>
<td>Parts per million</td>
<td>ppm</td>
</tr>
<tr>
<td>Per cent</td>
<td>%</td>
</tr>
<tr>
<td>Phosphate buffered saline</td>
<td>PBS</td>
</tr>
<tr>
<td>Phi</td>
<td>φ</td>
</tr>
<tr>
<td>Picomoles</td>
<td>pmol</td>
</tr>
<tr>
<td>Piconewtons</td>
<td>pN</td>
</tr>
<tr>
<td>Protein data bank</td>
<td>PDB</td>
</tr>
<tr>
<td>Psi</td>
<td>ψ</td>
</tr>
<tr>
<td>Polyacrylamide gel electrophoresis</td>
<td>PAGE</td>
</tr>
<tr>
<td>Polyethylene glycol</td>
<td>PEG</td>
</tr>
<tr>
<td>Polymerase chain reaction</td>
<td>PCR</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>Pa</td>
</tr>
<tr>
<td>Revolutions per minute</td>
<td>rpm</td>
</tr>
<tr>
<td>Response units</td>
<td>RU</td>
</tr>
<tr>
<td>Root mean square deviation</td>
<td>RMSD</td>
</tr>
<tr>
<td>Secretion systems type I-VI</td>
<td>T1SS-T6SS</td>
</tr>
<tr>
<td>Scanning transmission electron microscopy</td>
<td>STEM</td>
</tr>
<tr>
<td>Sigma</td>
<td>σ</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Size exclusion chromatography</td>
<td>SEC</td>
</tr>
<tr>
<td>Sodium dodecylsulphate</td>
<td>SDS</td>
</tr>
<tr>
<td>Super optimal broth with catabolite repression</td>
<td>SOC</td>
</tr>
<tr>
<td>Surface plasmon resonance</td>
<td>SPR</td>
</tr>
<tr>
<td>Three-dimensional</td>
<td>3D</td>
</tr>
<tr>
<td>Transmission electron microscopy</td>
<td>TEM</td>
</tr>
<tr>
<td>Tris (hydroxymethyl) aminomethane</td>
<td>Tris</td>
</tr>
<tr>
<td>Type IV pilus biogenesis</td>
<td>T4PB</td>
</tr>
<tr>
<td>Ultraviolet</td>
<td>UV</td>
</tr>
<tr>
<td>Units</td>
<td>U</td>
</tr>
<tr>
<td>Volts</td>
<td>V</td>
</tr>
<tr>
<td>Volume to volume</td>
<td>v/v</td>
</tr>
<tr>
<td>Weight to volume</td>
<td>w/v</td>
</tr>
<tr>
<td>Yeast extract tryptone media</td>
<td>DYT</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Bacterial secretion systems

The secretion of effector proteins, nucleoproteins and the transfer of genetic material are all fundamental to prokaryotic life. Bacteria accomplish the export of these factors from within the cell and across the cytoplasmic membrane usually via the Sec machinery, or general secretory pathway (GSP) as it sometimes referred to, if they are exported as unfolded precursors. Alternatively, the twin-arginine translocation (TAT) pathway serves to transport fully folded proteins, usually cofactor-bound enzymes, out of the cytoplasm. Gram-positive bacteria produce a single plasma membrane encapsulated by thick cell wall layer; in contrast gram-negative bacteria face an extra challenge due to the presence of not only the cytoplasmic, or inner membrane, but also the lipopolysaccharide-containing outer membrane.

At least six major mechanisms of effector protein and nucleic acid translocation have evolved to deal with this second translocation step, type I-VI. Some systems, namely the type I, III and IV systems, have developed so that effectors are extruded via one step directly from the cytoplasm, without utilising the Sec machinery for inner membrane translocation. Other systems are terminal branches of the GSP or TAT routes, and involve a periplasmic intermediate substrate. Because all of these systems secrete virulence factors, they provide attractive therapeutic targets in the growing age of antibiotic resistance, and therefore remain an active area of research. Several secretion systems assemble a pilus or fimbrial structure on the cell surface to facilitate the secretion of substrates or to communicate with the host. The CU pathway assembles adhesive pili, the type IV secretion system is an adapted conjugation system which uses pili for the transport of toxins and nucleoproteins, and the type III secretion system uses a needle-structure to inject effectors directly into the host. The type II secretion system and the type IV pilus biogenesis system form a closely related platform for the secretion of exotoxins and the presentation of cell surface pili respectively. Type IV pili are not used to secrete effectors, although they are used in genetic transfer, but instead to communicate with the host and govern other pathogenic functions such as twitching motility and biofilm formation. The following sections
discuss those secretion systems involved in fimbrial assembly in bacteria, concentrating on gram-negative paradigms. The systems which do not utilize a secretin outer membrane component, the chaperone/ usher pathway and the type IV secretion system (T4SS), are discussed first, followed by the type III secretion system (T3SS) and a detailed account of the closely related type II secretion system (T2SS) and the type IV pilus biogenesis (T4PB) system. The type II, type III and type IV pilus biogenesis systems all rely on an outer membrane pore formed by a member of the secretin family of proteins, which form a large part of the focus of this PhD project.

1.2 The Chaperone/ Usher pathway – P pili and type I pili

The chaperone/ usher (CU) pathway of pilus biogenesis is perhaps the most ubiquitous means of fimbrial secretion in bacteria, and responsible for assembly of adhesive fibres on the surface of gram-negative pathogens, which form an important class of virulence factors for host cell recognition and attachment. \(^5,10\) CU systems generally consist of an outer membrane β-barrel ‘usher,’ a ‘chaperone’ and pilus components or adhesins, which are often found encoded within a genetic operon. CU systems have been categorized into at least six different classes based on the usher component, as the number of chaperones, fimbrial subunits and adhesin-encoding genes can vary. Two such systems are well characterized, namely the FimD and PapC CU pathways of *Escherichia coli* which mediate the biogenesis of ‘type I’ and ‘P’ pili respectively. \(^5,11\) Many bacteria possess more than one CU system, for instance there are up to 12 CU pili encoded within a single *Escherichia coli* genome. The sequential expression of different pili and adhesins determines the final host tissue destination, for example the bladder (type I) or the kidney (P). \(^5,11\) The type I pilus appendage consists of a tip adhesin, FimH, and two tip complex proteins FimG and FimF, followed by a several copies of the FimA subunit which form a broader, rigid helical rod and form the main pilus structure. \(^10,12\) The P pilus consists of a tip complex of the adhesin PapG, a minor pilin subunit PapF, followed by several PapE pilus subunits to form a narrow fibrillum of about 2 nm diameter. This is followed by the main pilus helical rod structure of 6.8 nm diameter, consisting of thousands of copies of the PapA protein. A ‘termination subunit’ PapH lies at the base of the pilus. \(^5,13,14\)
The CU pathway forms a terminal branch of the general secretory pathway, and therefore begins with the sec-dependent translocation of nascent pilus subunits across the inner membrane. These then travel to the outer membrane in a binary complex with their cognate chaperones (FimC for type I pili, PapD for P pili).\(^{11,13}\) CU pilus subunits adopt an incomplete immunoglobulin-like fold lacking the C-terminal beta-strand, with an N-terminal extension. The X-ray structure of FimC chaperone in complex with the FimH adhesin pilus subunit revealed that a donor-strand complementation mechanism takes place, whereby the chaperone donates the missing strand to the pilus subunit completing the immunoglobulin-like fold and protecting the exposed hydrophobic core.\(^{15}\) When the complex meets the outer membrane usher (FimD for type I and PapC for P pili), pilus subunits are polymerized by a mechanism termed donor-strand exchange. The N-terminal extension of an incoming chaperone/ subunit complex releases the chaperone from the previous subunit by displacing the beta-strand of the chaperone, subsequently joining two pilus subunits together. Repetition of this allows more subunits to be incorporated into the growing pilus fibre.\(^{5,11,16}\)

**- Two-usher model for pilus biogenesis**

The usher components serve as pores in the outer membrane and assembly sites for the pilus structures. Ushers are ~800 residue proteins with four distinct domains. The N- and C-termini both form soluble periplasmic domains, whilst the central portion of the protein forms a 24-stranded kidney-shaped \(\beta\)-barrel. A ‘plug’ domain forms within the barrel between strands \(\beta6\) and \(\beta7\), occluding the lumen and is held in place by the \(\beta5/6\) hairpin. This would need to be displaced to open the conductance channel and allow passage of the fibre (figure 1.2.1a).\(^{10,13}\) Remaut *et al* determined a low-resolution cryo-EM reconstruction of the entire CU system, including the outer membrane FimD component, FimC, and the FimF/G/H pilus tip intermediate. This indicated that the FimD pores are functional dimers, with both monomers forming conductance channels, although only one channel extrudes the growing pilus whilst the other remains vacant.\(^{10}\) These observations were incorporated into a model for pilus assembly, where the N-terminal domain of one usher recruits the chaperone/ pilus subunit complex and the second usher acts as the translocation pore for the growing fibre (figure 1.2.1b).\(^{10,12}\)
Figure 1.2.1: PapC usher and model for P pilus assembly

a. Crystal structure of the PapC monomer (PDB 2VQI) viewed from the side (left) and from the top (right). The plug domain is coloured red and the β5/6 hairpin is coloured green. *Figure produced using CCP4MG.*

b. Schematic model for P pilus assembly by the chaperone/ usher pathway. The figure on the left shows the growing pilus tip, comprising the FimH distal tip adhesin and the FimG and FimF tip complex proteins. The periplasmic chaperone FimC (C2) forms a complex with the incoming FimA pilin subunit by donor strand complementation. This is recruited by the second usher N-terminal domain (usher 2 NTD). On the right, the C2-FimA complex is presented to FimF and donor-strand exchange takes place. This releases FimF from its chaperone (C1), and incorporates FimA into the pilus where it will begin to form the main helical rod. The N-terminal domain of usher 1 (usher 1 NTD) is then released and free to recruit the next FimC: A complex.
**Single-usher model for pilus biogenesis**

A recent study by Phan *et al* captured the FimD usher in complex with its FimC:FimH substrate. In this elaborate crystal structure, the ‘plug’ domain has been displaced from the barrel lumen and resides in the periplasm, and the β-barrel has undergone considerable rearrangement to lose its characteristic kidney shape and form a much more circular open pore, 32Å in diameter. The FimH adhesin contains a lectin domain, for host cell receptor binding, and a pilin-like domain for binding to the chaperone. The FimH lectin domain has replaced the plug domain; this structure therefore provides a snapshot of the translocator during substrate transport (*figure 1.2.2*).  

Perhaps most interesting is that the results from this study show that a single usher protomer forms a functional pilus assembly machine. The C-terminal periplasmic region of the usher formed two immunoglobulin-like domains, which were shown bound to the FimC:FimH complex. Therefore, the C-terminal domain contributes the second binding site required to hold the complex whilst the next complex is recruited; this role was performed by the N-terminal domain of a second usher in the previous model.  

This shows that a monomeric usher can act as a pilus assembly machine, where the chaperone/ subunit complex (subunit A) at the base of the pilus is bound to the C-terminal domain, whilst the N-terminal domain recruits a new complex (subunit B). Subunit B is placed to undergo donor-strand exchange with subunit A. The chaperone is then displaced from subunit A and also releases the C-terminal domain. Then, subunit B and its chaperone can dissociate from the N-terminal domain and transfer to the C-terminal domain, effectively pushing subunit A into the translocation channel.
Figure 1.2.2: Structure of the FimD usher during substrate translocation

The crystal structure of the FimD:FimC:FimH complex is shown on the left (PDB ID 3RFZ), and a simplified, schematic representation of the single usher model is shown on the right. In the crystal structure FimH subunit B is held in complex with its FimC chaperone at the FimD C-terminal domains (CTD). The FimD N-terminal domain (NTD) is free to recruit the next chaperone/ subunit complex. This will undergo donor strand exchange with the first, polymerizing the FimH subunits. Figure produced using CCP4MG.
1.3 Type IV secretion – the conjugative pilus

Type IV secretion systems are versatile, envelope spanning machines used by both gram-negative and gram-positive pathogens and symbionts for the intercellular secretion of protein effectors and nucleoproteins for the purpose of genetic exchange. Effectors are usually secreted into the cytoplasm of the target cell by direct host cell contact via a fimbrial structure. However, these systems may also be used to secrete into the extracellular environment. These diverse secretion machines all require ATPase activity for secretion, and they can be grouped on the basis of their contribution to pathogenesis. The first group comprises the conjugation systems, which mediate horizontal DNA transfer between bacteria; for example, to confer antibiotic resistance or also to transfer DNA into host cells for pathogenesis. A second group includes systems involved in DNA uptake from the environment for bacterial transformation, and a third group involves those systems responsible for the secretion of protein effectors. This is usually mediated by direct contact with the target cell, but effectors may also be released into the extracellular milieu, as in the case of the *Bordetella pertussis* toxin, and nucleoprotein complexes from *Neisseria gonorrhoeae*.6,18

Secretion by conjugative T4SS is Sec-independent so occurs via one step across both membranes. Systems can be generalized by three main substructures: a coupling protein, which receives substrates at the cytoplasmic entrance of the channel, a transport channel spanning the cell envelope, and a cell surface pilus to mediate host cell attachment.6 The archetypical T4SS is the VirB DNA delivery system of the plant pathogen *Agrobacterium tumefaciens*, which consists of the VirB1-VirB11 and VirD4 proteins. This system is responsible for the export of T-DNA, a transfer-prepared single stranded plasmid molecule, which has been ‘nicked’ by a relaxase enzyme. T-DNA is exported as a nucleoprotein complex with the VirD2 protein.19

There is a reasonably comprehensive model available of the molecular architecture of the VirB transporter and the transit of T-DNA (figure 1.3.1). One of the first studies to define the pathway of T-DNA secretion through the T4SS used crosslinking and T-DNA precipitation to identify substrate contact partners. The coupling protein, VirD4, recognises the nucleoprotein substrate first at the inner
membrane. It is then transferred to the VirB11 ATPase, which belongs to the family of hexameric ‘traffic’ ATPases which are also found in type II secretion and type IV pilus biogenesis systems. VirB11 is believed to then co-ordinate the ATPase activity of VirD4 and a third ATPase in the complex, VirB4. Using energy from ATP hydrolysis by VirD4/VirB11/VirB4 the substrate is then passed to the inner membrane proteins VirB6 and VirB8, which form an early part of the transenvelope channel, and subsequently to the pilus subunit VirB2.\textsuperscript{6,20}

Fronzes \textit{et al} reported a 15Å resolution cryo-EM structure of the core complex of the T4SS encoded on the pKM101 plasmid, which included homologues of three VirB proteins: VirB7, VirB9 and VirB10. The complex forms a large chamber, 80Å at its widest point, composed of fourteen copies of the three protein constituents, which form a double-walled ring. This has two distinct layers, termed the inner (I) and outer (O) layers. The I-layer lies in the periplasm, resembles a cup and is formed by the N-terminal regions of VirB9 and VirB10, the latter of which inserts into the inner membrane. The O-layer is composed of the outer membrane lipoprotein VirB7 and the C-terminal parts of VirB9 and VirB10, and this closes into a cap, narrowing the channel to only 10Å, which would need to be opened to allow passage of substrates.\textsuperscript{21} The crystal structure of the entire O-layer revealed that the cap in the outer membrane was composed of a hydrophobic ring of two-helix bundles contributed by VirB10; therefore this protein spans both membranes. This time the cap aperture is 32Å in diameter, implying that VirB10 opens at the cap to allow substrate passage. The authors propose a mechanism where VirB10 acts as an energy sensor, responding to the substrate bound, and catalytically active ATPases at the inner membrane induce a conformational change which is passed to the helical cap region to open the channel\textsuperscript{20,22}. 


Figure 1.3.1: Substrate transition by type IV secretion

Schematic cut-away representation of the type IV secretion core complex. The coupling protein, VirD4, recruits the substrate, which may be a relaxase-treated plasmid. This is passed to the traffic ATPase VirB11 which co-ordinates the ATPase activity of VirD2 and VirB4 to transfer the substrate across the inner membrane to the VirB6 and VirB8 channel proteins. The substrate is then passed to VirB9 and VirB2, the pilus component. The O-layer is formed by VirB7 and the C-terminal regions of VirB9 and VirV10. The I-layer is formed by the N-terminal regions of VirB9 and VirB10, and lined by VirB6 and VirB8. The pilus tip adhesin, VirB5, is also shown. 6,20-21
- Type IVB secretion system

A second major subclass of T4SS first became apparent in *Legionella pneumophila*,\(^{23}\) termed the type IVb secretion system (T4SSb), and a growing wealth of genetic information now suggests these systems are found in over 20 species of bacteria.\(^{24}\) The T4SSb contains some VirB-like proteins, and additional components that do not have any counterpart in VirB-like systems (T4SSa).\(^{13,25}\) The T4SSa and T4SSb also differ in the morphology of the conjugative pilus, with T4SSa pili forming long, flexible fibres at around 2-20\(\mu\)m in length, whereas T4SSb pili appear as rigid rods and are generally shorter at around 1\(\mu\)m.\(^{13}\) *Legionella pneumophila* possesses examples of both T4SSa and T4SSb. The LvhB proteins share extensive similarity with the VirB proteins and mediate plasmid transfer. It is the Dot/Icm system, however, that is absolutely required for virulence and falls into the T4SSb subcategory.\(^{26}\)

The VirB complex described above is thus far the only T4SS core complex that has been isolated and structurally characterized. However, genetic inspection and analysis by biochemical fractionation techniques of the *Legionella* Dot gene products has identified a core subcomplex consisting of DotC/D/F/G/H bridging the inner and outer membranes, akin to the T4SSa. Furthermore, there were a total of sixteen proteins associated with the inner membrane, suggesting a much more complex inner membrane platform in comparison to T4SSa. DotF and DotG are predicted inner membrane proteins, but they co-fractionate with outer membrane components DotC/D/H therefore form a complex with the outer membrane proteins.\(^{25}\) DotH is a periplasmic, outer membrane-associated protein that requires the outer membrane lipoproteins DotC and DotD for its correct targeting, reminiscent to the pilotin-mediated targeting of outer membrane secretins, which is discussed later (1.4). Interestingly, the crystal structure of the DotD lipoprotein revealed a fold similar to that found in periplasmic domains of secretins from the T2SS and T4PB system, the structure-function implications of which are unclear whilst the T4SSb remains poorly characterized. It is possible, however, that DotD may form a higher order ring structure, similar to that of secretins, presumably incorporating DotC and DotH, which also reflects the VirB7/9/10 ring observed in the T4SSa.\(^{24}\)
1.4 The secretin family of outer membrane proteins

Several secretion systems that produce pili or fimbrial-like structures are more complex and utilize outer membrane secretins for translocation across the outer membrane. These include the T3SS, the T2SS and the T4PB system. Secretins play major roles in these systems, enabling several vital pathogenic functions, including protein export, host-cell invasion, DNA uptake in Haemophilus influenzae and filamentous phage biogenesis.\textsuperscript{27,28} They form stable, SDS and heat resistant, homo-multimers of between twelve and fifteen subunits. An apparent large, central cavity of 50-80\textdegree would allow transport of the various substrates through the multimer,\textsuperscript{28} although whether this cavity forms a continuous channel remains debatable.

Over ninety members of the secretin family have been sequenced so far, and several organisms express multiple paralogues: for example, two are found in Neisseria meningitidis, five in Vibrio cholerae and eight in Pseudomonas aeruginosa.\textsuperscript{29} The family encompasses a diverse range of proteins, varying in subunit number, amino acid length and sequence with no single amino acid conserved throughout the entire family. Major size and sequence variation occurs at the N-terminus, whereas a region of approximately 300 aligned residues at the C-terminus are present in the majority of family members.\textsuperscript{29} Considering this, it would be logical that the N-terminal regions provide the diversity to function in a number of secretion systems. Also, as most transport systems involving secretins are complex and require many additional proteins, the N-terminal region may be responsible for interaction with other system components within the periplasm.\textsuperscript{28}

Currently there is no high-resolution information available for the structure of a complete secretin multimer to indicate how they form oligomers within the outer membrane. The well-conserved C-terminal region is predicted to be of beta-strand content and it is accepted that this forms the transmembrane region. It is also accepted that the C-terminal region is both necessary and sufficient for oligomerization,\textsuperscript{30-33} yet its topology in both the monomeric and multimeric form remains to be determined. There is, however, more information available regarding the N-terminal domains of secretins. In recent years crystal structures of N-terminal domains from secretins of the T2SS and T3SS have been solved.
The domains are termed N0-N3 based on their sequential positions from the N-terminus. These share a common modular domain structure of two $\alpha/\beta$ folds, one of which has been likened to a DNA-binding domain known as a KH-domain, and the other bears a similar fold to a domain from a TonB-dependent signalling receptor and also to bacteriophage tail proteins although the significance of this remains unclear. In addition, the mutual orientation of these domains with respect to one another differs in each structure, suggesting that the N-terminal regions may have a considerable degree of flexibility.

Several low-resolution EM structures have shed some light on secretin assembly, and highlighted some potentially important differences within the family. Studies of T2SS secretins have invariably revealed a constricted channel. Cryo-EM reconstructions of PulD from the *Klebsiella* T2SS appeared as a three-ring stack forming an overall ‘cup and saucer shape’ (figure 1.4.2a). The base of the cup-like structure occludes the channel formed by the rest of the protein. The structure of the PulD homologue, GspD, also forms a constricted channel, which opens to 75Å in the periplasm and narrows to 55Å in the central regions, before a periplasmic gate completely closes the channel (figure 1.4.2b). A very similar arrangement was seen in the region of T3SS needle complex that corresponds to the secretin, which also shows an area of density blocking the channel. The T4PB secretin PilQ (figure 1.4.2c), however, differs somewhat in that it is closed at both ends with a large vestibule inside the structure, yet a common theme still emerges which indicates that secretins have to be gated in some manner, and that they may adopt both open and closed states.
Figure 1.4.1 Crystal structures of T2SS and T3SS secretin N-terminal domains
High-resolution structures have been solved for the N-terminal domains of GspC (PDB ID 3EZJ) and EscC (PDB ID 3GR5) secretins. N1/N2-domains (magenta) have a KH-domain fold, and the N0-domain (blue) adopts a slightly different fold often found in secretin domains and other secretion system proteins, such as DotD. *Figure produced using CCP4MG.*

Figure 1.4.2 Low-resolution electron microscopy secretin structures
a. PulD from the T2SS. The trypsin-resistant core is coloured blue. *Figure taken from reference 33.*
b. GspD from the T2SS, shown with the atomic resolution structures of the N-terminal domains N0-N3 modelled into the map volume. *Figure taken from reference 37.*
c. Side, top and bottom projections of the PIlQ reconstruction from the T4PB system. The image shows that the chamber is closed at both ends. *Figure taken from reference 39.*
- Secretin accessory factors

Integral outer membrane proteins use Omp85 for membrane insertion, and this includes secretins. In the absence of Omp85 in *Neisseria meningitidis*, PilQ fails to form multimers and a decreased amount of PilQ is detectable on the cell surface by immunofluorescence.\(^{40,41}\) In the case of secretins, however, a number of partner lipoproteins have been implicated as additional requirements for their correct outer membrane insertion and multimerization. Such lipoproteins are sometimes referred to as pilotins, inferring that they mediate a chaperone-like targeting role. However, they sometimes form a stable, rather than transient, association with secretins which is different from the behaviour of chaperone proteins.\(^8\) In addition, the structure and apparent function of these proteins varies from system to system; in some systems, they are not required for the secretin to reach the outer membrane, and growing evidence suggests that the term ‘pilotin’ should be used loosely, if at all. Therefore they are discussed here as secretin accessory factors.

One such lipoprotein for which there is evidence of a chaperone-like function is MxiM from the T3SS of *Shigella flexneri*. MxiM binds monomers of the secretin MxiD in the periplasm, presumably preventing premature oligomerization. The structure of MxiM is that of a ‘cracked’ \(\beta\)-barrel, disrupted by an \(\alpha\)-helix (*figure 1.4.3a*).\(^{42}\) Within the cleft of the barrel, MxiM binds an 18-mer of its cognate secretin which, in turn, adopts a more defined structure within the groove. The hydrophobic binding cleft can also bind lipid and may prompt the pilotin to release the secretin to the membrane by an allosteric switch mechanism.\(^{42,43}\)

PilW, an outer membrane lipoprotein of the T4PB system, is important for the assembly of the PilQ secretin. PilW forms a superhelical arrangement of six tetratricopeptide repeats, and is structurally distinct from MxiM (*figure 1.4.3b*).\(^{44}\) In *Neisseria meningitidis*, pilW mutants fail to form heat-stable SDS-resistant multimers of PilQ, therefore it was suggested that PilW may act to stabilize the secretin.\(^{45}\) Similarly, in *Myxococcus xanthus*, PilQ monomers are found in the membrane fraction when the PilW equivalent protein, Tgl, is absent.\(^{46}\) The secretin is still found associated with the outer membrane but, given its low degree of beta-strand topology, it is unlikely to insert correctly as a monomer. In *Pseudomonas*
aeruginosa, however, the PilW homologue PilF has been demonstrated to be required for both outer membrane targeting and assembly of PilQ. A pilF strain showed an accumulation PilQ monomers like the previous studies in other organisms; however when the Cys18 lipidation site of PilF was substituted for alanine, so that PilF is not sorted to the outer membrane, assembled PilQ was detected in both the inner and outer membrane fraction, which is more consistent with a pilotin role.  

PilF and PilW are both sequence and structural orthologues, but the differences in these findings may indicate that there are species-specific differences in the precise protein function.

The T2SS secretin from Klebsiella oxytoca, PulD, requires the PulS lipoprotein for outer membrane insertion. In the absence of PulS, PulD localizes to the inner membrane and is degraded.  

Recently, the crystal structures of PulS from Klebsiella oxytoca and its Dickeya dadantii (Erwinia) orthologue OutS, have both been solved. The protein forms a novel four-helix bundle, distinct from other secretin auxiliary lipoproteins. The first and second helices form an antiparallel hairpin, which is wrapped by the hairpin formed between the second and third helices (figure 1.4.3c). A disulphide bridge between helices four and two is believed to be important for the structural integrity of the protein, and its ability to bind to the secretin. NMR cross-titration studies showed the pilotin binds the secretin at an unstructured, 18-residue stretch at the C-terminus, which becomes structured upon binding, somewhat reflecting the operation of MxiM.

The exact function of auxiliary lipoproteins remains unclear, and their structure and function varies between secretion systems. However, there is a strong common underlying theme, in that they all play a crucial role in maintaining the secretion pore. It seems that, in addition to secretin-associated lipoproteins, the outer membrane lipoproteins of other systems such as DotC/DotD and VirB7 of T4SS, share overlapping functions within their respective systems, although they are dissimilar at the sequence level.
Figure 1.4.3 Crystal structures of secretin accessory factors

a. The crystal structure of MxiM$^{42}$ from the T3SS reveals a β-barrel disrupted by an α-helix. In the structure a detergent acyl chain sits in the hydrophobic core of the barrel. (PDB ID 19YL).

b. The crystal structure of PilW from the T4PB system$^{44}$ reveals a superhelical arrangement of six tetratricopeptide repeats (PDB ID 2VQ2).

c. The crystal structure of OutS from the T2SS$^{50}$ reveals a novel four-helix bundle held by a disulphide, which is indicated by an arrow (PDB ID 3UTK).

All figures produced using CCP4MG.
1.5 Type III secretion – The injectisome

Type III secretion system (T3SS) machineries are utilized by gram-negative organisms for delivery of effector proteins into a host cell, the functions of which are then used to the bacterium's advantage for symbiosis or pathogenicity. They are distinct from other secretion systems due to the lack of a signal peptide in the secreted proteins, and the requirement for host cell contact for secretion. A variety of plant and animal pathogens, including *Escherichia*, *Pseudomonas*, *Shigella*, *Salmonella*, *Xanthomonas* and *Yersinia* species, use one or more of these systems which are often referred to as injectisomes, owing to the needle-like complex which spans the periplasm and projects several nanometres from the outer membrane. These systems also use three sets of proteins known as 'translocators' to form a pore in the target cell membrane and inject the effectors into its cytoplasm.

The type III machine shares sequence similarity of some components with those of the bacterial flagellum. This is indicative of a common ancestry and evolutionary divergence between the two systems. Indeed, the requirement for the flagellum arose prior to the first eukaryotes, so it is generally accepted that the type T3SS is derived from the flagellar system. It is likely that type III systems have arisen due to lateral gene transfer, as they are often encoded in extra-chromosomal elements and pathogenicity islands, and because bacteria often have more than one system. The systems have been grouped into seven main families of conserved genes, which have remained clustered at their loci, while the effectors are often encoded outside of this block and may even have been acquired from eukaryotes.

The Ysc family includes systems of many pathogens including *Yersinia*, *Pseudomonas*, and *Vibrio* amongst others. There may also be a system of the Ysc family encoded in the non-pathogen *Myxococcus xanthus*. Other families include the Inv-Mxi-Spa systems of *Salmonella* and *Shigella* species, the Ssa-Esc systems also of *Salmonella* as well as *Escherichia coli*, the Hrc-Hrp 1 and 2 systems which include all type III systems of plant pathogens, and the Rhizobiales type III systems which allow symbiosis of *Rhizobium* species with leguminous plants. The seventh family comprises the type III systems of *Chlamydiae*. These
are non-proteobacteria with type III systems whose components are not clustered within a genetic locus like those of the other five families. The distinct families have been reviewed extensively by Troisfontaines and Cornelis.52

- The needle complex

The T3SS are extremely complex and involve over 20 proteins although they do all seem to share two core structural components. The first is known as the needle complex, or the ‘injectisome’, and this bears architectural resemblance to the flagellar apparatus.7,38,53 The second is commonly known as the ‘translocon’, or tip-complex. The needle-complex is a supermolecular structure, composed of an external needle anchored into the bacterial envelope by a ‘multi-ring’ base that spans both membranes.54 Considerable evidence has helped to elucidate the structure and assembly of this complex in *Salmonella typhimurium*. A low molecular weight protein subunit, PrgI, is secreted by the base upon the recruitment of various other export accessory proteins to form the ~50 nm projection known as the needle, the length of which is believed to be controlled by an accessory protein, InvJ, which is also secreted in the same manner.55 In the *Yersinia* Ysc system, the prototypical type T3SS, extrusion of the needle subunit, YscF, continues until the needle reaches the length of the InvJ homologue YscP, which is thought to act as a molecular ruler to control needle length. In this model, the ends of the YscP protein anchor the central ‘ruler’, by attaching to the basal body and the growing needle tip.55,56 The needle is a tubular appendage, traversed by a channel of about 25Å diameter.57 The 3D structure of the MxiH needle appendage of *Shigella* was solved to 16Å by x-ray fibre diffraction and EM reconstructions. This revealed a helical architecture with a central channel.58 If protein substrates are extruded through the channel in the needle, they are likely to pass through unfolded due to these size constraints.

A very recent comprehensive study from Loquet et al69 used a combination of solid-state NMR, immunoelectron microscopy and Rosetta modelling to elucidate the atomic structure of the *Salmonella typhimurium* needle. Secondary structure analysis of 13C recombinant wild-type PrgI subunits determined from conformation-dependent chemical shifts revealed two long helices connected by a loop, as well as a 5-residue N-terminal domain of rigid structure, characteristic of neither α-helix
or β-sheet conformation. The collection of intrasubunit distance restraints revealed a helical hairpin motif, as shown in figure 1.5.1 by red lines. Intersubunit distance restraints meanwhile (blue lines), enabled the group to elucidate the subunit interfaces in the helical needle structure. This highlighted the importance of the last five C-terminal residues, which are essential for needle formation, as they form multiple intersubunit interactions. In addition, the rigid N-terminal domain forms strong interactions with the hairpin loop of a neighbouring subunit along the filament axis. Immunoelectron microscopy showed this domain to be surface exposed. The inner surface of the channel formed by PrgI consists mostly of polar, more conserved residues, whereas the needle surface is less conserved, which may represent a bacterial strategy for evading the immune response. 59

The basal body consists of one pair of protein rings in the inner membrane, connected to a secretin ring in the outer membrane by a periplasmic protein rod, PrgJ in Salmonella, which locks the needle substructure into the base. 55 These homo-oligomeric rings are composed of two membrane-anchored lipoproteins, namely PrgH and PrgK. 53 Sucrose density centrifugation has shown that the acylated N-terminus of the E. coli PrgK homologue, EscJ, localises to the bacterial inner membrane. As EscJ lacks a transmembrane domain, this ring is probably peripheral, and not integral to, the inner membrane. 60 Mutagenesis studies have shown that, although the secretin InvG can form multimeric rings without PrgH-PrgK, its oligomerization is enhanced in the presence of the inner membrane ring proteins. 61 PrgH and PrgK, however, require one another’s presence to form stable complexes. All three proteins are required for PrgI needle substructures to traverse the outer membrane. In the absence of PrgH-PrgK function no needle protein was detected and this was also the case in loss-of-function mutants of other components of the export machinery. Conversely, in the invG mutant, PrgI was detected in small amounts, although needle substructures were not extruded. It is likely that PrgI in this scenario remained associated with PrgH-PrgK in the periplasm, supporting the role of InvG as the secretion pore and consistent with an association of PrgH-PrgK with other components of the export machinery. 61

Marlovits and colleagues noted some interesting conformational differences between the structure of the base alone and the fully assembled needle-
complex. The base appeared to form a 300Å tall cylinder with a cup-like projection on the cytoplasmic side of the first inner membrane ring. This cup-like protrusion moved downwards in the fully assembled needle-complex, whilst the second inner membrane ring was clamped inwards, remodelling the cavity beneath the basal plate. The authors hypothesised that these changes may form the basis of specificity switching - from the secretion of the needle subunit PrgI, the inner-rod protein PrgJ and the regulatory protein InvJ, to the secretion of effector proteins. The same group later investigated this further by comparing cryo-electron micrographs obtained from wild-type and ∆InvJ strains of Salmonella typhimurium. Complexes isolated from the mutant lacked the PrgJ inner-rod component, and the conformation of the base resembled that observed in the wild-type before needle assembly in the previous study. In addition, the ∆InvJ mutant was unable to switch substrate specificity to effector proteins. Needle substructures were able to form, but these were unusually long and improperly anchored to the base. These studies supported the role of the regulator InvJ in needle length control, the requirement of the inner-rod structure for needle anchoring and the significance of the observed conformational changes in specificity switching.

The secretin component is generally accepted to form an oligomeric ring of twelve to fifteen subunits, much like its homologues in other secretion systems, whereas needle-complex particles are reported to be most abundant in 20- and 21-fold rotational symmetries in Salmonella typhimurium. However, there is confliction as to the exact stoichiometry of the three base ring components. In Shigella, the peripheral component of the inner membrane ring appeared to display 24-fold rotational symmetry. The 1.8Å crystal structure of the inner membrane protein EscJ (YscJ/PrgK) from enteropathogenic Eschericia coli revealed a flat triangular shaped molecule comprising two globular α/β domains connected by a linker. The interaction surface between monomers was extensive, involving the burial of several charged hydrogen bonding partners. The resulting ring-model from this structure maintained rotational symmetry at 24-fold for EscJ, although this cannot be extrapolated to conclude that this is the case for all orthologues in vivo. It is possible that needle complexes may not always be identical, with varying symmetries across different species. Following this, the crystal structures for the
N-terminal domains of the secretin EscC from EPEC, and the inner membrane component PrgH from *Salmonella*, showed that these also share a common α/β fold with EscJ (PrgK) of two α-helices flanked against a three-strand β-sheet, despite no sequence conservation. In the same study, the PrgH-EscJ/PrgK ring was then modelled as a 24-mer ring into the cryo-EM volume for the entire *Salmonella typhimurium* basal body, including the InvG secretin. The EscC secretin structure was modelled into the density as a 12-mer, as discussed earlier. Another study, also based on cryo-EM of the *Salmonella* needle complex, revealed that the basal body organizes as a 15-mer ring of InvG secretin subunits in the outer ring, and a 24-fold arrangement of PrgH/K subunits to form the inner rings. This results in a symmetry mismatch, and therefore an overall low C3 symmetry to the complex (*figure 1.5.2*). Together these models provide the most detailed representation to date of the T3SS envelope-spanning complex.

**Figure 1.5.1: Architecture of the T3SS needle assembly**

Loquet *et al* collected a set of intermolecular (blue lines) and intramolecular (red lines) distance restraints by solid-state NMR of recombinant PrgI needle subunits. These revealed a helical hairpin structure with a rigid N-terminal domain of five residues. Neighbouring subunits are connected by an α1-α1 and α2-α2 helix-helix packing arrangement. At this stage subunits i+5 and i+6 were indistinguishable, until Rosetta structure calculations produced a model for complete helical channel. *Figure taken from reference 59.*
Figure 1.5.2: Schematic representation of the T3SS basal body

The protein rings are sketched according to the cryo-EM volume of the needle complex from *Salmonella typhimurium*. PrgH and PrgK form the inner rings, whilst the secretin InvG forms the outer rings. The positions of PrgH, PrgK and InvG subunits are shown according to the docking of atomic structures into the cryo-EM map by Schraidt and Marlovits. The Prgl needle substructure, which traverses the channel, is also shown.
- The translocon

The second major structural component of the T3SS is the translocation pore, which is formed in the host cell membrane to enable the injection of effector proteins. This is mediated by a set of three proteins, two hydrophobic and one hydrophilic. These proteins are generally sequence conserved within type III system families whilst between families there are equivalents that are functionally analogous but not necessarily similar in sequence.\(^{\text{51}}\) The general consensus is that a tip complex is formed at the distal end of the needle substructure, probably consisting of the hydrophilic translocator, whilst a pore is formed by insertion of the hydrophobic translocators into the host membrane. There is debate, however, as to the composition and structure of the tip complex, and very little detail has been elucidated regarding the structure of the translocation pore. Despite this, several studies have helped to define the requirements for individual components of the translocon.

One such study on \textit{Pseudomonas aeruginosa} used haemolysis of erythrocytes by osmotic shock as an indicator of efficient pore formation. This group identified that all three proteins - PcrV, PopD and PopB in \textit{Pseudomonas} - are simultaneously required for translocation of effectors. They found that PcrV, the hydrophilic component, did not co-fractionate with the membrane whereas the two hydrophobic components PopB and PopD, were detected in membrane fractions. However, while \textit{ΔpcrV} mutants still efficiently secreted the hydrophobic components, they did not display any pore forming activity, so the hydrophilic component appears to be essential for translocon pore function by perhaps acting as an assembly platform for pore formation.\(^{\text{63}}\)

As the hydrophilic components such as LcrV of \textit{Yersinia} and PcrV of \textit{Pseudomonas} are surface-exposed antigens, the requirement of these components for haemolysis was tested again, but this time by binding antibodies to the antigens. The antibodies protected erythrocytes from haemolysis even in low doses. The group reinforced their results with competition experiments, by keeping antibody levels constant and increasing levels of recombinant antigen in the infection mixture, to ensure inhibition of haemolysis was a consequence of direct antibody-antigen interaction.\(^{\text{64}}\) It is noteworthy that an earlier study...
highlighted deletions in YopD, a hydrophobic component, that did not inhibit haemolysis yet disrupted effector microinjection\textsuperscript{65} therefore this result questions the veracity of haemolysis as a reflection of functional pore formation.

Several hypotheses have been proposed regarding the composition and structure of the tip complex. A common proposal, whereby the hydrophilic translocator attaches to the tip of the needle substructure and forms a distinct complex, is vindicated by most studies. STEM analysis of the \textit{Yersinia} needle reveals a distinct tip complex, comprising a head, a neck and a base. The main components isolated from needle fractions were LcrV, YopD and the needle subunit. After cross-linking, LcrV appeared to be the main structural component of this complex explaining its surface-exposed antigen properties\textsuperscript{66} YopD, however, could not be cross-linked into the complex\textsuperscript{67} The LcrV monomer, which has the common structural motif of hydrophilic translocators of a coiled-coil connecting two globular domains\textsuperscript{68} appears to form a pentameric complex at the needle tip\textsuperscript{67} Conversely, in \textit{Shigella}, another model based on crystal packing suggests that the hydrophobic component, IpaB, may exist in complex with four IpaD monomers at the tip. IpaD is the hydrophilic translocator in \textit{Shigella}, which directly attaches to the needle substructure\textsuperscript{69} and also has a structure of two domains around a central coiled-coil in the crystal\textsuperscript{70} A TEM study in \textit{Escherichia coli} revealed yet another possible tip complex structure. Immunogold labelling of anti-EpsA antibodies revealed that EpsA, the corresponding hydrophilic component, also directly associates with the needle but forms a sheath-like structure. At the tip of a comparably shorter needle substructure, this EpsA filament appears to bridge the secreton and the host cell\textsuperscript{71}
1.6 Type II secretion and type IV pilus biogenesis

The remaining two bacterial secretion systems that utilize a secretin OM component for extrusion of their substrates are the T2SS and the T4PB system. Both systems form multi-protein macromolecular machines, which span the cell envelope of gram-negative bacteria. The extensive parallels observed between the two systems, which share several homologous protein components, mean they have long been considered to have descended from a common evolutionary origin. To outline the basic analogy of their architecture, both systems comprise an outer membrane secretin component, which may or may not require a lipoprotein pilotin for correct targeting and assembly, as discussed earlier in the context of the secretin family. The inner membrane components can vary in their arrangement and composition, but usually form a complex consisting of one or more ATPases, a bitopic inner membrane protein or protein complex sometimes referred to as an assembly platform, and a polytopic inner membrane protein. Both systems also include prepilin proteins which are processed by a leader peptidase at the inner membrane, and assemble into filamentous structures. These components often share a degree of structural homology and some sequence conservation across the two systems, and are classed as analogous. Other components may vary according to the different functions of the separate T2SS and T4PB systems, or may even vary within the same system of different organisms. A comparison of the general architecture of the two systems is presented in figure 1.6.1, and the functions of the various components are discussed in the following pages. The gene designation of individual components differs between organisms and between systems. The components of the well-characterized systems are cross-referenced in Table 1.6 for clarity. Both systems will be reviewed, using the generalized ‘Gsp’ nomenclature for the T2SS and the Neisseria ‘Pil’ nomenclature for the T4PB, unless gene products are referred to specifically.
Figure 1.6.1 The type II secretion system and type IV pilus biogenesis system

Schematic representation of the T2SS (left) and T4PB system (right), using the standard ‘Gsp’ and Neisseria ‘Pil’ gene nomenclature. Analogous components from the respective systems are coloured the same. Pilin (PilE) or pseudopilin (GspG) subunits are synthesized as precursors then processed by a prepilin peptidase (GspO/PilD) at the inner membrane before they are incorporated into the growing (pseudo)pilus fibre, by a complex mechanism involving several gene products. Each component is reviewed in the following pages.
The T2SS - often referred to as the terminal branch of the GSP functions to secrete predominantly fully-folded proteins into the extracellular milieu. Up to now studies of type II systems have depicted an envelope-spanning pseudo-pilus assembly complex, composed of at least twelve different gene products.\textsuperscript{73,74} The T2SS uses ATP-binding and hydrolysis to assemble a periplasmic pseudopilus that pumps exoproteins across the outer membrane.\textsuperscript{74} The first type II systems were discovered in \textit{Klebsiella oxytoca} on the ‘Pul’ gene cluster,\textsuperscript{75} and were found to be conserved across gram-negative species. The system is strongly affiliated with pathogenicity; for instance \textit{Pseudomonas aeruginosa} utilizes this system for the secretion of lipase, alkaline phosphatase and exotoxin A,\textsuperscript{73,76} and \textit{Vibrio cholerae} for cholera toxin and neuraminidase secretion amongst other proteases.\textsuperscript{77} Secretion occurs via two membrane translocation steps. The first involves the transport of unfolded proteins bearing a cleavable N-terminal signal peptide across the inner membrane - a process orchestrated universally by the Sec machinery.\textsuperscript{73} Substrates reach their tertiary conformation in the periplasm, and mature exoproteins are then expelled from the cell via the highly specific type II machinery. However another method of inner membrane translocation, the TAT system, has been shown to operate alongside the Sec system and is capable of allowing folded enzymes possessing a twin-arginine motif to cross the inner membrane. Voulhoux and co-workers demonstrated that two Xcp-dependent proteins from \textit{Pseudomonas aeruginosa} utilise the TAT system to enter the periplasm. Taking this, the authors challenge the designation of type II secretion as the general secretory pathway \textit{per se}.\textsuperscript{78} However, ‘Gsp’ is still currently the common term applied to any protein of the T2SS.

The T4PB system facilitates the extrusion and retraction of the type IV pilus appendage from the bacterial surface. Therefore the secreted substrate in this case is the polymeric fibre of the pilin subunit itself. Type IV pili appear as proteinaceous filaments, 6-9 nm in diameter and several microns in length,\textsuperscript{9} which are also able to depolymerize and retract due to the presence of a designated ATPase in the assembly system. Type IV pili govern various important functions relating to the pathogenicity of many bacteria, such as host cell adhesion, twitching motility, DNA uptake and biofilm formation.\textsuperscript{9} The pilin subunit is synthesised as a prepilin precursor, which is processed at the inner membrane by
cleavage of the signal peptide by a leader peptidase. The mature pilin subunit is then incorporated into the growing pilus fibre, which is presented on the cell surface. Type IV pili fall into two categories: IVa and IVb, which can be distinguished by the structure of the pilus itself, the specific function of the pilus and by subtle differences in assembly systems. These are discussed where necessary with respect to individual system components in the following sections.

### Table 1.6: T2SS and T4PB system nomenclature

<table>
<thead>
<tr>
<th>Type II secretion system</th>
<th>Proposed Function</th>
<th>Gsp</th>
<th>Pseudomonas</th>
<th>Klebsiella</th>
<th>Erwini</th>
<th>Vibrio</th>
<th>Aeromonas</th>
<th>Xanthomonas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretin pore</td>
<td>GspD XcpQ PulD OutD EpsD ExeD XpsD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepilin peptidase</td>
<td>GspO XcpA PulO OutO VcpD TapD XpsO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudopilin subunit</td>
<td>GspG XcpT PulG OutG EpsG ExeG XpsG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic ATPase</td>
<td>GspE XcpR PulE OutE EpsE ExeE XpsE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polytropic platform protein</td>
<td>GspF XcpS PulF OutF EpsF ExeF XpsF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner membrane proteins</td>
<td>GspL/M XcpY/Z PulL/M EpsL/M ExeN XpsL/M OutL/M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilotin</td>
<td>GspS PulS OutS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretin-interacting protein</td>
<td>GspC XcpP PulC OutC EpsC XpsN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type IV pilus biogenesis system</th>
<th>Proposed function</th>
<th>Pseudomonas</th>
<th>Neisseria</th>
<th>Myxococcus</th>
<th>Haemophilus</th>
<th>Thermus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretin pore</td>
<td>PiQ PiQ PiQ ComE PiQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepilin peptidase</td>
<td>PiD PiD PiD PiD PiD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilin subunit</td>
<td>PiA PiE PiA PiA PiA4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic ATPase</td>
<td>PiB/T/U PiB/T/U PiB/T PiB PiB/T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polytropic platform protein</td>
<td>PiC PiG PiG PiC PiC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner membrane proteins</td>
<td>PiM/N/O PiM/N/O PiM/N/O ComA/B/C PiM/N/O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilotin</td>
<td>PiF PiW Tgl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secretin-interacting protein</td>
<td>PiP PiP PiP ComD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.7 The pilus and pseudopilus appendage

**- Prepilin processing**

Type IV pili are polymeric fibres of a 15-20 kDa major subunit, which is synthesised as a prepilin precursor. Essential for T4PB\textsuperscript{81,82} is the inner membrane-localized prepilin peptidase - PilD - a homologue of GspO from the T2SS, which cleaves the hydrophobic leader peptide to liberate the mature pilin subunit. In addition, PilD has been shown to provide the subsequent N-methyltransferase activity required for post-translational modification of the now exposed N-terminal residue, and thus represents a bifunctional enzyme.\textsuperscript{80} The equivalent protein required for processing of the pilin-like proteins of the parallel T2SS, GspO, is also found at this location; therefore it is necessary for both pilus biogenesis and protein extrusion. Most PilD homologues contain a two-pair cysteine motif in the N-terminal domain, usually in the form C-X-X-C\textsubscript{21}-C-X-X-C. Alkylation of the sulphydryl groups leads to a loss of enzyme activity, suggesting the cysteine residues play a role in catalysis or substrate binding. However, the XpsO (GspO) peptidase of \textit{Xanthomonas campestris} although lacking this motif is able to complement \textit{pilD} mutants in \textit{Pseudomonas aeruginosa}.\textsuperscript{83}

The prepilin precursor remains anchored in the membrane by α1 before processing by the peptidase, which removes the leader peptide from the cytoplasmic face of the membrane. Meanwhile, oxidative folding of the globular domain proceeds in the periplasm, as an important disulphide is introduced by a thiol-oxidoreductase.\textsuperscript{84}

**- The type IV pilus fibre**

The first published structure of a pilin monomer was that of \textit{Neisseria gonorrhoeae} gonococcal (GC) pilin,\textsuperscript{85} and despite an overall lack of sequence conservation across T4PB systems, pilins share a similar overall architecture.\textsuperscript{9} A conserved, N-terminal stretch of around 53 residues is the most conserved part of the protein and forms an extended helix, which embeds into a globular beta-sheet or sandwich domain at the C-terminus, held by a conserved disulphide.\textsuperscript{9} Based on amino acid sequence and topology, pilins can be subdivided into two groups. Type IVa pilins have an α-amino-methylated N-terminal phenylalanine residue, and a
generally short cleavable leader peptide of 5-6 residues. They are found in many bacteria encompassing a broad host range. Type IVb pilins are generally only found on enteric pathogens of the human intestine, have a longer signal sequence and have a different topology within the beta-sheet to the type IVa group.\textsuperscript{86}

Craig \textit{et al} published a comprehensive study in 2006\textsuperscript{87} in which they determined the structure of the full length type IVa pilin subunit, PilE, from \textit{Neisseria gonorrhoeae} to 2.3Å (\textit{figure 1.7.1a}), and used this to build into a cryo-EM reconstruction of the pilus fibre, to produce a model for an atomic resolution gonococcal (GC) pilus (\textit{figure 1.7.1b}). The N-terminal 53 residues of PilE, which remain conserved throughout pilins, form the α1 helix, the C-terminal half of which is wrapped by a four-stranded antiparallel β-sheet. This forms the common ‘ladle-like’ architecture evident in pilin subunits, of an extended helix with a globular head domain. On either side of the β-sheet region, an αβ-loop connecting the extended helix and the β-sheet regions, and the D-region which is defined by the Cys121-Cys151 disulphide bridge delineating a hypervariable loop, are thought to provide distinct surface chemistry to the common structure.\textsuperscript{87,88} The resulting model of the pilus appendage docked the pilin subunits into a helical fibre, with the hydrophobic N-terminal α1 helix of each subunit packed into the central core and the globular head domains forming the surface of the fibre. The helices provide most of the packing interaction, forming extensive hydrophobic contacts within the pore of the fibre, and grooves on the surface of the fibre are formed between the more loosely associated globular heads. An abundance of positively charged residues line the crevices, which may facilitate uptake of DNA by the GC pilus.\textsuperscript{87}

In a second model, the type IVb TcpA fibre formed by \textit{Vibrio cholerae}, the only interactions between subunits appear to be mediated by the α1 helices and the αβ-loops (\textit{figure 1.7.1b}). The α1 helices are tightly packed with no steric clashes, and each contacts six helices from neighbouring subunits. Due to tight hydrophobic packing no channel traverses the fibre, indicating that assembly occurs from the base. In this second model the D-region is also solvent-exposed, and fewer globular head contacts create deeper cavities and a more variegated pilus surface.\textsuperscript{89} Another model for type IVb pili has been proposed for the ETEC bundle-forming pilus (Bfp). The NMR structure of the bundilin subunit adopts the
common characteristics of other pilins, however the beta-sandwich domain has a contiguous topology that differs from that of type IVa pilins.  

The conserved helix and common β-sheet elements of the globular head may provide a molecular scaffold for the structurally variable regions, namely the αβ-loop and D-regions, which mediate interactions between subunits and the environment. If this model is true in vivo, the hypervariable regions of the globular head are predominantly exposed on the surface of pili. This will provide broad antigenic variation to pili and enable them to continually evolve to evade immune attack, making the design of vaccines directed towards these pathogens difficult.

A number of ‘minor’ pilins have also been identified in T4PB systems, which are designated so due to their low abundance with respect to the major pilin subunit. However, these proteins have prepilin-like N-termini, canonical prepilin peptidase cleavage motifs and are required for several type IV pili-mediated functions such as DNA transformation and aggregation, whilst being dispensable for pilus biogenesis. One such protein in Neisseria species, PilX, has a structure that closely resembles the major pilin, PilE and it is incorporated into the fibre. Deletion of PilX results in a loss of bacterial aggregation, therefore it has been proposed that PilX subunits in neighbouring antiparallel fibres of separate organisms may function to brace against each other via their protruding D-regions.
Figure 1.7.1: Structure of the type IV pilus subunit and model for the assembled pilus

a. Crystal structure of full length PilE from *Neisseria gonorrhoea* (PDB ID 2HI2) (left) and truncated TcpA from *Vibrio cholerae* (PDB ID 1OQV) (right). The αβ-loop is shown in blue, the D-region in magenta and the N-terminal extended helix in red. *Figure produced using CCP4MG.*

b. Cryo-EM reconstruction of the *Neisseria* GC pilus (left) and model for the *Vibrio* TcpA fibre. The αβ-loops are coloured green and the D-regions are purple. *Figure taken from reference 9.*
- The T2SS pseudopilus

As mentioned earlier, the T2SS also assembles a pilus-like structure for substrate extrusion, known as the pseudopilus. This might act as a piston, to push directly against exoproteins or prevent backward flow in the channel. Pseudopilins earn their name due to their structural similarity and likely ancestral relationship to pilins of the T4PB system, and the T2SS also possesses a major subunit, GspD, and several minor subunits GspH-K. Overproduction of PulG (GspG) in *Escherichia coli* K-12 showed that the protein was able to assemble into pili that resembled type IV pili; likewise, the T2SS secreton was able to assemble fibres of a T4PB pilin, reiterating the close relationship between these systems.

Mutation and microscopy studies of the Xcp T2SS of *Pseudomonas aeruginosa* revealed that increased levels of the minor pseudopilin XcpX (GspK), an atypical pseudopilin due to its larger size, negatively regulate the length and number of assembled XcpT (GspG) polymers. Also, XcpV (GspI) was the only minor pseudopilin required for pilus assembly, therefore it was suggested to act as an initiator. None of the other minor pseudopilins were necessary to form the appendage, nor could they form an appendage themselves.

A subsequent investigation into the minor pseudopilins by Korotkov and Hol provided a high-resolution structure of a heterotrimeric complex formed by GspK/I/J in *Escherichia coli*. The structure for GspK, the atypical pseudopilin, revealed that it contains a pilin-like domain and an additional unique α-domain. The α-domain, comprising twelve α-helices and four short β-strands, lies between strands β2 and β3 of the pilin-like globular head. The authors propose that the heterotrimer caps the tip of the pseudopilus, preventing other pseudopilins from assembling above it, and may contact either the secretin pore to open the channel or form an interaction with the extruded substrate. Aside from the GspK α-domain, all four minor pseudopilins and the major pseudopilin GspG share the pilin-like topology of type IVa pilins, with a long N-terminal helix and antiparallel β-sheet head, but possess marked distinctions in their variable regions. The major deviation from pilins of the T4PB system, however, is that the pseudopilins do not contain the crucial conserved disulphide bond found within the D-region.
1.8 Inner membrane components of the T2SS and T4PB system

- A multi-protein complex is formed at the inner membrane of the T2SS

Currently in the field, the majority of evidence in the literature describes a multi-protein complex formed at the inner membrane of both the T2SS and the T4PB system. In the T2SS, this complex incorporates GspE/F/L/M and C (see Table 1.6) and it has been characterized in several organisms. For instance, early work on the ‘Eps’ T2SS of *Vibrio cholerae* established a quaternary GspE/L/M complex. EpsE is essential for cholera toxin secretion, and requires an intact Walker A box for this purpose otherwise leading to accumulation of assembled cholera toxin in the periplasm.\(^98-100\) It has therefore been identified as a secretion ATPase. Upon subcellular fractionation of cells, the ATPase EpsE can be found in both the cytoplasm and membrane fractions. However in *epsL* mutants EpsE localizes only to the cytoplasm,\(^98,101\) suggesting that the EpsL gene product is required for recruitment of EpsE to the inner membrane. In addition, the EpsL and EpsM gene products, both integral, bitopic inner membrane proteins, could co-immunoprecipitate from subcellular fractions with an antibody against either component and protect one another from proteolytic degradation, suggesting that they too form a stable complex.\(^99,101,102\) As protein interactions within the T2SS are species specific, Sandkvist *et al* were able to use a domain swapping approach to determine important regions of EpsL for its interaction with EpsE and EpsM. By producing chimeric proteins of EpsL and its *Aeromonas* equivalent ExeL, the group were able to determine that the cytoplasmic N-terminal portion of EpsL interacts with EpsE, and the transmembrane or C-terminal periplasmic part of EpsL forms a complex with EpsM, therefore providing a link across the inner membrane.\(^100\)

An inner membrane T2SS complex was also identified in *Erwinia chrysanthemi*. Using a yeast two-hybrid system, two new binary complexes were identified: GspF/E and GspF/L, incorporating the polytopic inner membrane protein GspF into the complex for the first time.\(^103\) This study also agreed with previous work suggesting that GspM formed part of the complex too. The results suggested that GspE/L form a complex, which subsequently recruits GspF, as the GspF/E interaction is dependent on the presence of GspL.\(^103\) Similarly, an inner membrane
complex was identified in the pullulanase ‘Pul’ T2SS of *Klebsiella oxytoca* which, in agreement with previous findings, included GspE/L/M but a new component, GspC, was also found to be associated.¹⁰⁴ Later studies of Xcp proteins from *Pseudomonas aeruginosa* suggested that the GspC homologue may further strengthen the GspL/M interaction although its involvement in the complex may be transient *in vivo*. This group also used cross-linking and co-purification techniques to show that GspL/M was part of a larger complex with GspE and GspF, which is in agreement with observations from other organisms.¹⁰⁵

More recent structural data have provided a clearer vision of the complex formed by T2SS components at the inner membrane. The periplasmic domain of EpsM forms a dimer, which is believed to be the physiological form.¹⁰⁶ Each monomer is composed of a layer of two alpha helices sandwiched against a four-stranded anti-parallel beta sheet, and is a variant of the ferrodoxin-like fold. A cleft lined with conserved residues between the two monomers at β3 and β3’ forms a possible small ligand binding site. The cytoplasmic domain of EpsL forms a crystallographic dimer, although it is monomeric in solution, and each monomer consists of three distinct domains. It should be noted, however, that the full-length protein does form dimers. Domains I and III are structurally related to the RNaseH-like fold. Two copies of domains with this fold are also found in proteins belonging to the superfamily of actin-like ATPases. Domains II and III appear to mediate the interaction with EpsE in the cytoplasm, which inserts an alpha helix into the cleft between the two domains in the EpsE/L co-crystal.¹⁰⁷,¹⁰⁸ It is likely that a much larger complex forms at the inner membrane due to the proposed hexameric quaternary structure of GspE.¹⁰⁹ This has not yet been built into a model in complex with EpsL; the current EpsE/L heterotramer co-crystal provides evidence for the hypothesis that energy is transferred from EpsE across the inner membrane T2SS components, for the translocation process across the outer membrane.

Interestingly the structure of periplasmic domain of EpsL was later solved, and it also adopts a ferrodoxin-like fold very similar to the periplasmic domain of EpsM, suggesting that the two share a common evolutionary origin. However, the dimer formed by peri-EpsL is distinct to that of EpsM.¹¹⁰ The reason for this remains unclear, and would perhaps become more obvious from a peri-GspL/M complex.
Although the stoichiometries have yet to be defined, there is a well-established inner membrane complex in the T2SS, consisting of a cytoplasmic hexameric ATPase GspE, which interacts with the polytopic GspF protein. These form a complex with the dimeric GspL protein, which crosses the membrane and forms a complex with the integral, dimeric membrane protein GspM in the periplasm.

- The inner membrane complex of the T4PB system

The composition of inner membrane components of the T4PB system was less clear until recently, when structural data emerged highlighting distinct parallels with the T2SS. A systematic genetic analysis in Neisseria meningitidis identified a core set of only six proteins required for the assembly of pilin polymeric fibres. These included four conserved genes clustered within a genetic locus, PilM/N/O/P, as well as the pre pilin peptidase PilD and the ATPase PilF. Sequence and hydropathy analysis of these genes in Pseudomonas show that PilM is likely to be cytoplasmic, where PilN and PilO are type II membrane proteins. PilM also shares sequence similarities with the bacterial cell division protein FtsA, which are concentrated in motifs that are conserved in eukaryotic actin and involved in ATP-binding. PilN and PilO contain similar predicted secondary structure elements and identifies a stable heterodimeric complex between the two. The group also solved the crystal structure of Pseudomonas aeruginosa.

Two concurrent studies from the same laboratory have improved our understanding of the PilM/N/O/P inner membrane complex in the T4PB system of Pseudomonas aeruginosa. The first, a comprehensive study of pilM/N/O/P mutants, showed that although PilM is a cytoplasmic protein, it localizes to the inner membrane and is necessary for the stability of the PilN/O/P complex. Therefore it is likely to be forming an interaction with the cytoplasmic part of PilN or PilO. In addition, loss of any one of the pilM/N/O/P components had a negative impact on the stability of the other proteins. The second study notes that PilN and PilO contain similar predicted secondary structure elements and identifies a stable heterodimeric complex between the two. A high degree of sequence conservation in the cytoplasmic N-terminus of PilN proteins, and the presence of a consensus INLLPW sequence, compared with lack of such conservation in the N-terminus of PilO, suggests that it is PilN mediating the interaction of PilM in the cytoplasm.
the periplasmic domain of PilO, to find it adopts a variant of the ferrodoxin-like fold similar to that observed in the periplasmic domains of GspL and GspM of the T2SS.\textsuperscript{114}

Collectively, this evidence suggests that the PilM/N/O complex is equivalent to the GspL/M complex observed in the T4PB system, although the components have unrelated sequences. The predicted actin-like ATPase fold of PilM together with a homology model for the PilN periplasmic domain are analogous to full-length GspL, whereas the ferrodoxin-like fold observed in the periplasmic domain of PilO is structurally equivalent to GspM. As yet, however there has not been an interaction observed between PilM and the T4PB ATPases from the GspE family, as one might expect if PilM represents the functional orthologue of cyto-GspL.

Since the publication of the studies of the inner membrane complex of \textit{Pseudomonas}, the crystal structure for PilM of \textit{Thermus thermophilus} has emerged.\textsuperscript{115} Despite subtle differences in the T4PB system of \textit{Thermus} to non-thermophiles - such as a lack of the otherwise essential lipoprotein PilP, a lack of the secretin pilotin PilW and differences in the predicted domain organization of the secretin - it has been shown that the PilM/N/O components are essential, also localize to the inner membrane and are the equivalent of those in non-thermophilic T4PB systems.\textsuperscript{116} Interestingly, PilM crystals diffracted poorly, until the protein was combined and co-purified by size-exclusion with a synthetic peptide incorporating the conserved consensus sequence from the N-terminus of the PilN protein. This resulted in improved crystals, which diffracted to a limit of 2.2Å, whilst also confirming the complex formed by PilM/N. As expected, PilM forms a two-domain structure belonging to the family of actin-like ATPases. Although ATP could be seen bound within the structure, ATP-hydrolysis by PilM was not observed. The closest structural homologue for PilM was FtsA, as expected due to sequence similarity; but PilM was monomeric in solution, whereas FtsA has been reported to polymerize.\textsuperscript{117} The PilM structure also aligns well with that of cyto-EpsL, which was reported to form monomers as discussed earlier. Cyto-EpsL, however, lacks a portion of the second domain that would be required for it to bind ATP. The PilN peptide was bound within a narrow channel of the first domain, which is mirrored in the EpsL structure by the placing of the C-terminus.\textsuperscript{115}
BfpC is a bitopic inner membrane protein of the bundle-forming pilus (Bfp) system of EPEC.\textsuperscript{118} The Bfp system is a type IVb pilus biogenesis system (T4PBb), and it has been suggested that the T4PBa and T4PBb systems are functionally distinct and cannot be compared directly due to differences in composition and lack of sequence homology across many components.\textsuperscript{112} However, a recent crystal structure of the cytoplasmic domain of BfpC has shown that the protein adopts a similar fold to PilM, despite no sequence similarity, indicating there might be functional equivalents in the two systems after all. However, the structure of PilM is far more closely related to that of cyto-EpsL, which seems unusual considering the T4PBa and T4PBb systems are both involved in producing type IV pili.\textsuperscript{119}

The evidence therefore suggests that PilM is more closely related by sequence and structure to FtsA and actin-like ATPases, yet has adapted functionally to act in a position equivalent to that of cyto-EpsL and BfpC. It also supports the previous hypothesis that the PilM/N complex forms an analogue of full-length EpsL in the T2SS, and that the inner membrane complexes of proteins from the two systems, which share little sequence similarity, are structurally and perhaps functionally equivalent.

- The polytopic inner membrane platform protein

An important inner membrane component found across the T2SS and T4PB system is a highly conserved multi-spanning integral membrane protein, known collectively as the GspF family. There is some controversy over the requirement of GspF proteins for a functional secretion system and their function remains debatable. Early mutational studies showed GspF family members to be indispensable for type IV pilus formation in \textit{Neisseria}\textsuperscript{120} and \textit{Pseudomonas}.\textsuperscript{81} However this has been challenged in the case of PilG, a GspF homologue in the \textit{Neisseria} T4PB system, by analysis of pilG/pilT double mutants, which were still able to produce pili.\textsuperscript{82} Therefore PilG may serve to counteract the retraction ATPase, rather than serve as a canonical assembly factor. This cannot necessarily be extended to the T2SS, however, due to lack of a retraction ATPase in these systems.

As mentioned earlier, in the case of the T2SS, there is evidence to suggest that GspF proteins form part of the multi-protein inner membrane complex, via an
interaction with GspE ATPases and GspL/M.\textsuperscript{103,105} A study in the T2SS of \textit{Pseudomonas aeruginosa} exploited the instability of XcpS (GspF) in the absence of other Xcp gene products to determine complex partners. Co-production with XcpR (GspE) and XcpY (GspL) simultaneously stabilized XcpS, indicating it forms part of a ternary complex or an assembly platform for pseudopilus formation.\textsuperscript{121}

A role for GspF proteins in the inner membrane complex has also been described for the T4PBb system of EPEC. Yeast two-hybrid studies were implemented to show that the GspF homologue BfpE forms interactions with the cytoplasmic domain of the bitopic inner membrane protein (BfpC), and both the assembly and retraction ATPases (BfpD and BfpF respectively). Furthermore, the group showed that the complex formed with BfpF was via a cytoplasmic loop of BfpE, the residues of which are conserved across GspF members in the T2SS and T4PBa systems. Disruption of this interaction by mutation caused cells to form irregular aggregates, which failed to disaggregate, indicating aberrant bundle-forming pilus retraction. Therefore, this implies a direct role for BfpE in BfpF mediated pilus retraction and therefore a direct role in pilus dynamics.\textsuperscript{118,122}

There are contrasting lines of evidence regarding the topology of GspF proteins. An in-depth study of BfpE (GspF) topology used an alkaline phosphatase and β-galactosidase fusion plasmid with a variety of BfpE truncations covering several BfpE end points over each of the predicted hydrophilic regions, in order to determine cytoplasmic and periplasmic segments. This revealed the protein adopts a four-pass transmembrane helix topology, with a cytoplasmic N-terminal domain, followed by three transmembrane helices, a large periplasmic loop, a fourth transmembrane helix and a periplasmic domain. By contrast, OutF (GspF) of the \textit{Erwinia carotovora} T2SS revealed a three-pass transmembrane helix topology in a similar fusion protein study, which connects a cytoplasmic domain, a periplasmic loop and a second cytoplasmic domain.\textsuperscript{123} Sequence analysis across the GspF family shows most members to have three predicted transmembrane helices, therefore the alternative topology of BfpE may be a novel variation.\textsuperscript{124}

GspF proteins have been shown to associate with each other\textsuperscript{118} and to form both dimers and tetramers. The crystal structure of the cytoplasmic domain of EpsF (cyto-EpsF), the GspF homologue in the T2SS of \textit{Vibrio cholerae} favoured the
dimeric arrangement. Both crystal packing and in vivo crosslinking experiments suggested a physiological dimer with a substantial, conserved interface.\textsuperscript{125} Low-resolution EM reconstructions of the PilG (GspF) protein from \textit{Neisseria meningitidis} favoured a tetrameric arrangement; however, separation of the purified protein by PFO-PAGE showed that it could form both dimers and tetramers. 3D-reconstructions with both imposed C2 and C4 rotational symmetry produced essentially similar models, although the volume of density would accommodate a molecular mass consistent with a PilG tetramer. It is therefore difficult to distinguish between a dimer of dimers, or a tetramer with a true four-fold symmetrical arrangement in this case.\textsuperscript{126} The subsequent crystal structure of the cytoplasmic domain from PilC (cyto-PilC), a GspF homologue of the T4PB system of \textit{Thermus thermophilus}, has a similar structure to that of cyto-EpsF with both forming a six-helix bundle. The cyto-PilC structure also reveals a dimer, although the arrangement of monomers differs. However, purification of the protein by size-exclusion suggests that it forms tetramers. Mutating the residues of the dimer interface caused the protein to dissociate from tetramers into dimers, suggesting that the dimer interface is physiologically relevant.\textsuperscript{127} This implies that the dimer of dimers arrangement may be the case \textit{in vivo} for GspF proteins.
1.9 T2SS and T4PB system ATPases

The traffic ATPases involved in the T2SS and T4PB systems have been briefly referred to so far in the context of their potential interactions with other members of the inner membrane complex, they are discussed here in terms of structure and function. ATPases generally have low sequence similarity, but possess a highly conserved nucleotide-binding sequence that is vital for protein secretion and pilus biogenesis.\(^{128}\) The “traffic ATPases” of the T2SS and T4SS possess additional conserved regions for function: the aspartate “Asp” box, which may form the nucleotide-binding fold by interacting with magnesium, the histidine “His” box, of unknown function and a tetracysteine motif, which appears to be essential for secretion.\(^{129,130}\) Although they are cytoplasmic components, their notable involvement in complex formation at the inner membrane and the obvious requirement for energy transduction across the membrane for the secretion process make them a vital component of the inner membrane subassembly. They are generally referred to as the GspE family, although ATPases involved in pilus retraction belong to a subgroup, which will be referred to as the PilT family; these are not found in the T2SS, which possess only a single assembly ATPase.\(^{131}\) Members of both, however, are also found in the T4SS. The ‘traffic’ ATPases are believed to form part of the complex at the inner membrane, as the ATP-hydrolysis activity of EpsE of *Vibrio cholerae* is stimulated by EpsL, the bitopic inner membrane protein.\(^{132,133}\) In addition, as noted earlier, an interaction between BfpF (PilT family) and the polytopic inner membrane protein is vital for pilus retraction,\(^{118}\) which enables bacterial twitching motility. As yet, however, there have not been any specific interaction partners identified for ATPases of the T4PBa system.

Traffic ATPases invariably appear to function as hexamers. The crystal structure of the T4SS ATPase from HP0525 from *Helicobacter pylori* formed a bilobed structure with the distinct N-terminal and C-terminal domains. These arranged into two hexameric rings with ADP bound at the interface of the two domains of mixed alpha and beta topology. The C-terminal domain also adopted a fold similar to that of the RecA-like proteins.\(^{134}\) A hexameric arrangement was then verified by ultracentrifugation.\(^{135}\) When the structure was solved in the apo-form, it became apparent that the N-terminal domains adopted a variety of rigid body
conformations about the linker, which were rotated by anything from 2-16° relative to the C-terminal domain. This suggests that ATP-binding may induce a closed conformation.\textsuperscript{135} Alternating open and closed conformations of a hexameric ring were also observed in small angle X-ray scattering solutions of GspE from \textit{Archaeoglobus fulgidus}, with the clamped, closed conformation the catalytically active form depending on ATP.\textsuperscript{136}

The crystal structure of EpsE from \textit{Vibrio cholerae} was also similar to that of HP0525, with a distinct N-terminal domain and RecA-like C-terminal domain. The C-terminal domain contains the characteristic Walker A, Walker B, Asp box and His box sequence motifs, which line the nucleotide-binding groove.\textsuperscript{109} A conformational change model has also been proposed for hexameric EpsE, where binding of nucleotide to one subunit causes drastic conformational changes that are relayed through the other subunits and across the interacting inner membrane components.\textsuperscript{137,138}

A crystallographic study on the retraction ATPase PilT of \textit{Aquifex aeolicus} revealed a hexameric structure, with a PAS-like N-terminal domain and a RecA-like C-terminal domain, which superimposed well onto the EpsE and HP0525 structures possessing all the conserved traffic ATPase motifs. In addition, the base of the C-terminal domain was decorated with seven short alpha helices one of which contained the AIRNLIRE motif, which is a signature sequence necessary for pilus retraction.\textsuperscript{139} Crystals were obtained in several packing contexts. The six-fold rotation axis in one crystal generates a symmetric hexamer, which diffracts to 2.2 Å resolution and clearly defines the subunit-subunit interface between the N- and C-terminal domains of neighbouring subunits. This does not, however, form a complete three-dimensional ATPase active site; therefore it is probably not the enzymatically active conformation. A second, lower resolution, crystal form with six monomers in each \textit{C2} asymmetric unit (\textit{figure 1.9.1}) was also solved. Monomers A,C,D and F formed a closed conformation via the linker between the N- and C-terminal domains. The remaining two have undergone a large domain rotation of 65° about the linker to form an open conformation and pull the hexamer into an ellipsoid shape.\textsuperscript{139} Such domain movements have also been observed in PilT from \textit{Pseudomonas aeruginosa}, with and without nucleotide bound.\textsuperscript{140}
Structural information for the traffic ATPases has shed light on the dynamic nature of the energy motors for protein secretion. Large movements of the N-terminal domain, which is believed to connect ATPases to inner membrane components such as GspL, produce a substantial conformational change in the hexamer, which could be transferred through proteins across the inner membrane to supply the energy to drive secretion processes.

**Figure 1.9.1: Crystal structure of PilT from Aquifex aeolicus**

Subunits B and E in the PilT hexamer (blue) adopt a substantially different domain orientation to the other four subunits (magenta), breaking the six-fold symmetry and perhaps representing catalytically active and inactive states (PDB ID 2GSZ). Figure produced using CCP4MG.
1.10 GspC and PilP proteins

PilP proteins are inner membrane anchored lipoproteins essential for T4PB. Due to association of PilP with the outer membrane secretin, previous reports have speculated the protein functions as a pilotin,\(^1\) as a regulator of secretin dynamics, or as an anchor for the secretin,\(^2\) however lack of evidence for these hypotheses mean that its function remains unclear. The protein is unlikely to function as a pilotin, as evidence shows the secretin PilQ can form outer membrane multimers without PilP.\(^3\) PilP proteins are esterified at a conserved cysteine residue at their C-terminus by the addition of a fatty acid for membrane anchoring. This is followed by a long stretch of ~80 amino acids with low conservation and no secondary structure.\(^4\) A folded domain is found at the C-terminus, comprised of a seven-stranded beta-sandwich variant which opens by approximately 45° and is lined with hydrophobic residues.\(^5\) Far western analyses have shown that PilP interacts with the secretin,\(^6\) and co-purification studies have shown that it forms a stable ternary complex with PilN and PilO,\(^7\) so it may serve to bridge the outer and inner membrane components, yet little more is known about its function. PilP therefore forms a large part of the focus of this thesis, and is discussed in more detail in the introductory sections to relevant chapters.

The remaining inner membrane component of the T2SS is a bitopic membrane protein, often referred to as the trans-periplasmic protein, and designated GspC. GspC proteins have no known sequence homologue in the T4PB system, and appear to be the least conserved component of the T2SS.\(^8\) GspC forms part of the GspE/L/M subcomplex, which has been demonstrated in the case of PulC of *Klebsiella oxytoca*,\(^9\) and also forms an interaction with the GspD secretin, providing a connection between components of both membranes.\(^10\)

GspC family members are organized into several domains. Typical secondary structure elements at their C-terminus indicate the presence of either a coiled-coil domain, in the case of *Pseudomonas aeruginosa* and *Pseudomonas alcaligenes*, or a PDZ-domain in *Vibrio cholerae*, *Klebsiella oxytoca* and *Erwinia chrysanthemi*.\(^11\),\(^12\),\(^13\) Central to the peptide is a highly conserved domain, known as the homology region (HR), followed by a transmembrane segment and an N-terminal cytoplasmic extension.\(^14\) GspC proteins are the least conserved of the
T2SS, and they are species-specific and not interchangeable between even closely related organisms.\textsuperscript{143,145} It has been proposed that GspC proteins maybe involved in recognizing exoprotein substrates, which would explain their species specificity. A recent set of \textit{in vitro} SPR experiments exploited the presence of two T2SS, the Xcp and Hxc systems, present in \textit{Pseudomonas aeruginosa} to test this. The results indicated an interaction between purified XcpP and secreted substrates elastase and lipase, and no interaction between XcpP and alkaline phosphatase, the substrate of the Hxc T2SS.\textsuperscript{146} Another study of the Xcp secretion machineries of \textit{Pseudomonas aeruginosa} and \textit{Pseudomonas alcaligenes} demonstrated that while XcpP (GspC) and XcpQ (GspD) are not individually interchangeable, partial complementation of an \textit{xcpPQ} mutant could be achieved when the two components are interchanged pairwise.\textsuperscript{145} Therefore the species specificity of GspC proteins may be attributed in part to a species-specific interaction with the secretin, as well as the recognition of exoproteins.

A separate study aimed to determine specific non-exchangeable domains within the GspC members XcpP of \textit{Pseudomonas aeruginosa} and OutC of \textit{Erwinia chrysanthemi} by producing chimera proteins and monitoring exoprotein secretion. This determined that the transmembrane domain of XcpP could be replaced by that of OutC without a complete loss of function, although previous work had shown it could not be replaced by that of an unrelated protein. Therefore it functions as more than a membrane anchor but is unlikely to confer specificity. This study also narrowed down a completely non-exchangeable region to 35 residues in the periplasm preceding the HR-domain, and surprisingly indicated that the respective coiled-coil and PDZ-domains were exchangeable despite their structural differences.\textsuperscript{143} In addition, a C-terminal truncation of XcpP (GspC) from \textit{Pseudomonas} lacking the coiled-coil domain was able to complement an \textit{xcpP} mutant.\textsuperscript{147} In contrast, deletion and domain exchange studies of \textit{Erwinia} OutC homologues indicated that the PDZ-domain was necessary to confer the specificity of some, but not all, proteins to be secreted,\textsuperscript{148} therefore the function of the C-terminal domain remains unclear or may differ between species or between substrates, some of which may be PDZ-independent. A bacterial two-hybrid study using the OutC protein from \textit{Erwinia chrysanthemi} showed that the transmembrane helix tightly self-interacts irrespective of the periplasmic region,
which does not form multimers alone. Therefore the transmembrane segment is vital for GspC multimerization and also for OutC function, as periplasmic forms lacking this segment are non-functional.\textsuperscript{147,148} Hydrodynamic studies revealed OutC forms stable elongated dimers with a globular C-terminal end formed by the PDZ-domain;\textsuperscript{149} the transmembrane segment may, therefore, serve just to produce active dimers, or may also be involved in complex formation with other inner membrane components.

Co-purification studies on EpsC and EpsD proteins from \textit{Vibrio vulnificus} indicated that the HR-domain of EpsC binds to the periplasmic domain of the secretin, but the PDZ-domain does not.\textsuperscript{144} Furthermore, deletions in the HR-domain have been shown to inactivate the secretion of exoproteins,\textsuperscript{148} indicating its importance for GspC function. Login \textit{et al} used NMR spectroscopy to characterize the interaction between OutC and OutD from \textit{Erwinia chrysanthemi}. Comparison of the HSQC spectra for the full periplasmic region of OutC and the PDZ domain alone indicated that the HR and PDZ-domains are independent of one another, consistent with a segmental domain structure. In addition the PDZ-domain alone was unaffected by addition of OutD, whereas the construct including the HR-domain was affected. The interaction domain of OutC was further refined to a 20-residue fragment, designated the ‘secretin interacting peptide’ (SIP), at the C-terminal end of the HR-domain. This formed an interaction with two distinct sites on the periplasmic region of the secretin, one in the N0-domain and one sitting astride the N2- and N3-domains. Furthermore, the N0-domain of the secretin recognized a second region of OutC close to the transmembrane segment. The authors proposed that binding of the SIP to the first site in the N0-domain may increase its affinity for the second by causing a conformational change in the secretin, thereby provoking dissociation of the SIP from the N0-domain, which is free to form an interaction with the second region on OutC.\textsuperscript{150} This describes an allosteric switch mechanism, which directly implicates GspC in secretin dynamics perhaps by opening the secretin pore.

A subsequent study from the same group again used NMR spectroscopy to determine both the structure of the HR-domain and also residues involved in the interaction with the N0-domain.\textsuperscript{151} Similarly, a crystal structure emerged, detailing the interaction between the HR-domain of GspC and the N0-domain of GspD from ETEC.\textsuperscript{152} Both studies establish an interaction between the HR-domain and N0-
domain, however they contradict the previous finding that the SIP at the C-terminal domain mediates this interaction. Instead, residues at the HR:N0 interface are located at the N-terminal end of the HR-domain, within the first β-strand. The two HR-domain structures are in agreement, a sandwich of two three-stranded antiparallel beta sheets, which open at an approximate 45° angle.\textsuperscript{151,152} Remarkably the structures bear striking similarity to that of the conserved domain of PilP from the T4PB system despite a lack of sequence similarity. Most HR-domain structures and predicted structures appear to lack the secondary structure found in strand β4 of PilP, which is also more conserved in PilP proteins.\textsuperscript{151}

The HR:N0 complexes presented in the two papers differ, with regards to the interacting beta-strand on the secretin. In the computational HADDOCK\textsuperscript{153} complex model that was produced using the solution NMR structure of HR from OutC, with an OutD N0-domain homology model, strand β3 of the N0-domain is placed in an antiparallel arrangement with strand β1 of HR.\textsuperscript{151} However, residues most affected by the interaction reside within strand β1 and helix α2 of N0.\textsuperscript{151} In the crystal structure, strand β1 of the HR-domain forms an antiparallel arrangement with strand β1 of the N0-domain. Residues on helix α1 are also contributors to the interface.\textsuperscript{152} In addition, the interaction between OutC and OutD was not sufficiently tight to isolate by size-exclusion or co-crystallize.\textsuperscript{151} The differences could result from species-specific differences in the GspC/D complexes, however both indicate that the proteins interact by beta-strand augmentation, and highlight the similarity of the GspC HR-domain structure to PilP.
1.11 Aims and Objectives of the PhD project

An understanding of the structure and function relationships of proteins within macromolecular systems such as the type IV pilus biogenesis system is a fundamental step towards understanding their mode of operation and how they might provide future antimicrobial targets. The recent advances with respect to the individual components of secretion systems which utilize fimbrial structures for their purpose have been outlined in this introduction. There is, of course, still much to be discovered.

The aim of this project was to investigate the secretin component of the T4PB system which forms the outer membrane pore, a vital requirement for pilus extrusion. Electronmicroscopy structures have indicated that this pore can adopt open and closed states in several secretion systems therefore must be regulated in some manner. Recent atomic-resolution data has provided greater insight into the structures of secretin periplasmic domains for the type II and type III systems but as yet none has become available for the T4PB system. Therefore the initial aims of the project were:

- To screen a variety of PilQ homologues from several different organisms for overproduction and purification of N-terminal periplasmic domains.

- To produce soluble, recombinant PilQ domains in samples suitable for structural studies by X-ray crystallography or NMR, and to determine atomic-resolution structures for these.

Another component of the T4PB system about which relatively little is known is PilP. PilP is essential for the system, and believed to interact with both the inner and outer membrane components. At the start of the project, an NMR solution structure was available for PilP but as yet no crystal structure had been published. Therefore the aims of the project were:

- To overproduce and purify the recombinant folded domain of PilP for structural studies by X-ray crystallography.

- Conduct an in-depth investigation into the interaction formed between PilP and PilQ, and to characterize this by biophysical methods and co-crystallography.
Chapter 2: Materials and Methods

2.1 Cloning and expression materials

The tables presented in the following pages of section 2.1 reference the cloning and gene expression materials used throughout this study. These include the oligonucleotide primers used for amplifying the required DNA sequences from organism genomes by PCR (Table 2.1.1 and Table 2.1.2), the bacterial strains which were used as expression hosts for protein overproduction (Table 2.1.3), gene expression vectors used for protein over production (Table 2.1.4) and the organism and strain names of genomic DNA templates used (Table 2.1.5).
<table>
<thead>
<tr>
<th>Construct</th>
<th>Forward primer 5’ to 3’</th>
<th>Reverse primer 5’ to 3’</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>22-478</td>
<td>AGGTTCTGTTTCAGGGCCCGCTAAGCGTTTCCAAAATCGATTTCAAACG</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>161-478</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>22-195</td>
<td>AGGTTCTGTTTCAGGGCCCGTGCAATTCCGTGAAGGGGGTCG</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>119-308</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>199-507</td>
<td>AGGTTCTGTTTCAGGGCCCGTGCAATTCCGTGAAGGGGGTCG</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>22-266</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>22-460</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>314-507</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>209-460</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>208-468</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>19-198</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>113-278</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>64-169</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>173-493</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>292-408</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>53-160</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>27-132</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>1-132</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>227-545</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>350</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>425</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
<tr>
<td>480</td>
<td>AGGTTCTGTTTCAGGGCCCGGTCCCCAAGACCGTCCGCC</td>
<td>ATGGTCTAGAAGAACTTTAGCCGCCAATGCCCTTTGGC</td>
<td>OPPF</td>
</tr>
</tbody>
</table>

**Table 2.1.1: Primers for constructs produced at OPPF**

Chapter 2

**OPPF**
### Table 2.1.2: Primers for constructs produced at the home laboratory

<table>
<thead>
<tr>
<th>Construct</th>
<th>Forward primer 5’ to 3’</th>
<th>Reverse primer 5’ to 3’</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>NmPilQ24-131</td>
<td>CGCCTTTCAGACAGGATCCGCAGGAAACATTACA</td>
<td>GCGGGGCGTGCCTCGAGTTACACGGTATCGTCCGA</td>
<td>pET28a</td>
</tr>
<tr>
<td>NmPilQ224-329</td>
<td>...</td>
<td>...</td>
<td>pET28a</td>
</tr>
<tr>
<td>NmPilQ24-329</td>
<td>...</td>
<td>...</td>
<td>pET28a</td>
</tr>
<tr>
<td>PaPilQ24-128</td>
<td>GTTCGCGCCGGCACATATGGCCGCGGACCTG</td>
<td>CGGCGGAAGCGACGGATCCTCAGGCCGGCGAATTGCC</td>
<td>pET15b</td>
</tr>
<tr>
<td>PaPilQ159-267</td>
<td>...</td>
<td>...</td>
<td>pET15b</td>
</tr>
<tr>
<td>PaPilQ24-267</td>
<td>...</td>
<td>...</td>
<td>pET15b</td>
</tr>
<tr>
<td>NmPilP</td>
<td>...</td>
<td>...</td>
<td>pET15b</td>
</tr>
</tbody>
</table>

### Table 2.1.3: \textit{E. coli} cloning and protein overproduction strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shott Tsp10</td>
<td>F–, mcrA Δ(mrr-hsdRMS-merBC) Δ(biuA51) Δ(prr) Δ(srfA) Δ1056 (DE3)</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>Omnimax II</td>
<td>F’ (acmAB ΔiacIq ΔZAM15 ΔiacZB ΔfopR) Δ(oriC) (DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td>B834 (DE3)</td>
<td>Δ-ampR ΔhslSΔrscU ΔrpoB ΔrpoN ΔrpoD (DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td>Rosetta (DE3) plycD</td>
<td>Δ-ampR ΔhslSΔrscU ΔrpoB ΔrpoN ΔrpoD (DE3) plycD (CamR)</td>
<td>Novagen</td>
</tr>
<tr>
<td>Origami B (DE3)</td>
<td>Δ-ampR ΔhslSΔrscU ΔrpoB ΔrpoN ΔrpoD (DE3) upaC (CamR)</td>
<td>Novagen</td>
</tr>
<tr>
<td>BL21 (DE3)</td>
<td>Δ-ampR ΔhslSΔrscU ΔrpoB ΔrpoN ΔrpoD (DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td>T7 Express</td>
<td>Δ-ampR ΔhslSΔrscU ΔrpoB ΔrpoN (CamR)</td>
<td>NEB</td>
</tr>
</tbody>
</table>

### Table 2.1.4: Protein overproduction vectors

<table>
<thead>
<tr>
<th>Vector</th>
<th>Cloning sites used</th>
<th>Purification tag</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET15b</td>
<td>BamHII; Ndel</td>
<td>N-terminal histidine</td>
<td>Novagen</td>
</tr>
<tr>
<td>pET28a</td>
<td>BamHII; Xhol</td>
<td>N-terminal histidine; C-terminal histidine</td>
<td>Novagen</td>
</tr>
<tr>
<td>pOPIN-F</td>
<td>KpnI, HindIII</td>
<td>N-terminal histidine</td>
<td>Novagen</td>
</tr>
</tbody>
</table>

### Table 2.1.5: Genomic DNA templates

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neisseria meningitidis</td>
<td>MC58</td>
</tr>
<tr>
<td>Pseudomonas aeruginosa</td>
<td>PAO1</td>
</tr>
<tr>
<td>Xylella fastidiosa</td>
<td>Temecula1</td>
</tr>
<tr>
<td>Legionella pneumophila</td>
<td>Philadelphia 1</td>
</tr>
<tr>
<td>Xanthomonas campestris</td>
<td>ATCC 33913</td>
</tr>
<tr>
<td>Aeromonas hydrophila</td>
<td>ATCC 7966</td>
</tr>
<tr>
<td>Ralstonia solanacearum</td>
<td>GM1000</td>
</tr>
<tr>
<td>Thermus thermophilus</td>
<td>HB8</td>
</tr>
</tbody>
</table>
2.2 Composition of buffers and reagents

2.2.1 Chemicals

All analytical grade purity chemicals were purchased from BDH Ltd., Difco Co., Sigma Aldrich or Melford Laboratories Ltd.

2.2.2 Cloning enzymes

All enzymes used during cloning were purchased from New England Biolabs, Roche or Fermentas and used according to the supplier's guidelines.

2.2.3 Antibiotics and inducers

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Stock solution</th>
<th>Working concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin/ carbenicillin</td>
<td>75 mg ml$^{-1}$ in sterile water</td>
<td>75 µg ml$^{-1}$</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>40 mg ml$^{-1}$ in sterile water</td>
<td>40 µg ml$^{-1}$</td>
</tr>
<tr>
<td>Tetracycline-HCl</td>
<td>5 mg ml$^{-1}$ in ethanol</td>
<td>50 µg ml$^{-1}$</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>34 mg ml$^{-1}$ in ethanol</td>
<td>170 µg ml$^{-1}$</td>
</tr>
<tr>
<td>IPTG</td>
<td>1 M in sterile water</td>
<td>0.1-1 mM</td>
</tr>
</tbody>
</table>

All solutions were passed through a 0.22 µm syringe filter

2.2.4 Growth media

<table>
<thead>
<tr>
<th>Media</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYT (1 L)</td>
<td>10 g yeast extract, 16 g tryptone, 5 g NaCl</td>
</tr>
<tr>
<td>LB (1 L)</td>
<td>10 g tryptone, 5 g yeast extract, 10 g NaCl</td>
</tr>
<tr>
<td>SOC (1 L)</td>
<td>20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 20 mM glucose</td>
</tr>
<tr>
<td>DYT agar (1 L)</td>
<td>DYT plus 15 g bacto-agar</td>
</tr>
<tr>
<td>LB agar (1 L)</td>
<td>LB plus 15 g bacto-agar</td>
</tr>
</tbody>
</table>

All growth media solutions were sterilized by autoclaving
2.2.5 Composition of M9 minimal media

- 42 mM Na$_2$HPO$_4$, 24 mM KH$_2$PO$_4$, 9 mM NaCl, 19mM $^{15}$N-NH$_4$Cl, 1 mM MgSO$_4$, 0.1 mM CaCl$_2$, sterilized by autoclaving

- 4 g L$^{-1}$ of glucose for $^{15}$N labelling OR $^{13}$C-glucose for a $^{15}$N$^{13}$C labelling, added after autoclaving

- 0.2 mg ml$^{-1}$ $^{15}$N-Isogro (Sigma) OR $^{15}$N$^{13}$C-Isogro added after autoclaving

2.2.6 Buffers and reagents for agarose gel electrophoresis

**Loading dye (6x)**
50 mM EDTA, 0.25% (w/v) bromophenol blue, 30% (v/v) glycerol, 0.25% (w/v) xylene cyanol

**Ethidium bromide solution**
Ethidium bromide 10 mg ml$^{-1}$ (w/v), ethanol

**TAE buffer (50x)**
Per 1 litre aqueous volume: 25 mM EDTA pH 8.0, 57 g glacial acetic acid, 142 g Tris-base

**TE buffer (6x)**
10 mM Tris-base, 1 mM EDTA, pH 8.0

2.2.7 Protein purification buffers

**Binding buffer**

- 50 mM NaH$_2$PO$_4$/ Na$_2$HPO$_4$ pH 7.8, 100 mM NaCl in most cases

- 50 mM Tris-HCl pH 8.0, 100 mM NaCl if the protein was intended for crystallography trials

**Lysis buffer**

- Binding buffer plus 1 mg ml$^{-1}$ lysozyme, 40 μg ml$^{-1}$ DNase I (Sigma), and 1x EDTA-free complete protease inhibitor cocktail (Roche) for a 2 litre cell culture preparation

68
Chapter 2

Wash buffer

- Binding buffer plus 20 mM imidazole

Elution buffer

- Binding buffer plus 200 mM imidazole

Protein standard buffer

Proteins were exchanged into a variety of standard buffers for storage or specific experiments. This was typically elution buffer minus imidazole after desalting, although a range of other buffers were used and these are specified where necessary in the results:

- 50 mM Tris-HCl pH 9.0, 200 mM NaCl for PaPilP crystallography
- PBS: 137 mM NaCl, 10 mM Phosphate, 2.7 mM KCl, pH 7.4 for SPR
- 50 mM NaH$_2$PO$_4$/ Na$_2$HPO$_4$ pH 6.8, 50 mM NaCl for most NMR experiments
- 50 mM NaH$_2$PO$_4$/ Na$_2$HPO$_4$ pH 7.0, 100 mM NaCl for ITC experiments

2.2.8 Qiagen buffers for robotic protein overproduction screening

Qiagen lysis buffer

- 50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, 1%(v/v) Tween 20, adjust pH to 8.0 using NaOH and sterile filter

Qiagen wash buffer

- 50 mM NaH$_2$PO$_4$, 300 mM NaCl, 20 mM imidazole, 0.05%(v/v) Tween 20, adjust pH to 8.0 using NaOH and sterile filter

Qiagen elution buffer

- 50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM Imidazole, 0.05%(v/v) Tween 20, adjust pH to 8.0 using NaOH and sterile filter
### 2.2.9 SDS-PAGE buffers and reagents

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Loading buffer (4x)</strong></td>
<td>200 mM Tris-HCl pH 6.8, 40% (v/v) glycerol, 8% (w/v) SDS, 0.02% (w/v) bromophenol blue, 20% β-mercaptoethanol</td>
</tr>
<tr>
<td><strong>MES running buffer (1x)</strong></td>
<td>50 mM MES, 50 mM Tris-base, 0.1% SDS, 1 mM EDTA, pH 7.3</td>
</tr>
<tr>
<td><strong>MOPS running buffer (1x)</strong></td>
<td>50 mM MOPS, 50 mM Tris-base, 0.1% SDS, 1 mM EDTA, pH 7.7</td>
</tr>
<tr>
<td><strong>SDS running buffer (10x)</strong></td>
<td>25 mM Tris-base, 192 mM glycine, 0.1% SDS, pH 8.3</td>
</tr>
<tr>
<td><strong>Resolving buffer</strong></td>
<td>1.5 M Tris-HCl pH 8.8</td>
</tr>
<tr>
<td><strong>Stacking buffer</strong></td>
<td>0.5 M Tris-HCl pH 6.8</td>
</tr>
<tr>
<td><strong>SDS solution</strong></td>
<td>10% (w/v) SDS</td>
</tr>
<tr>
<td><strong>APS solution</strong></td>
<td>10% (w/v) APS, dH₂O</td>
</tr>
</tbody>
</table>

**Stacking gel:**

- 2.6 ml dH₂O, 1.3 ml stacking buffer, 1 ml acrylamide, 50 μl SDS solution, 50 μl APS solution, 4 μl TEMED.

**Resolving gel:**

- 0.6 ml dH₂O, 1.3 ml stacking buffer, 3 ml acrylamide, 50 μl SDS solution, 50 μl APS solution, 4 μl TEMED.
2.2.10 Western blotting buffers and staining solution

**Western transfer buffer**
25 mM Tris, 150 mM glycine, 10% (v/v) methanol, pH 8.3

**TBS**
10 mM Tris-HCl, pH 7.5, 150 mM NaCl

**TBS-Tween/ Triton**
20 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.05% (v/v) Tween 20, 0.2% (v/v) triton X-100

**Blocking buffer**
3% (w/v) BSA in TBS buffer

**Staining solution**
1 tablet of Sigma Fast™ BCIP/NBT alkaline phosphatase in 10 ml of sterile water
2.3 Cloning, protein production and purification protocols

2.3.1 Transformation procedure

Plasmid DNA was transformed into the necessary *E. coli* cells for cloning purposes or for protein overproduction. Strains used for different purposes are indicated where appropriate. In most cases, ~50-100 ng of plasmid DNA was added to a 50 µl aliquot of competent cells. Cells were then incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds, and placed back onto ice for 2 minutes, unless otherwise indicated by the supplier. Cells were grown for one hour by adding 950 µl DYT media (2.2.4) and shaking at 37°C. Depending on the cloning efficiency, which was judged on a trial and error basis, 10-200 µl of the cell suspension was spread onto DYT agar plates (2.2.4) containing the necessary antibiotic at working concentration (2.2.3) and left to incubate overnight at 37°C. Three plates were often prepared, each with a different volume of cells applied to them in order to ensure single colonies were obtained.

2.3.2 PCR amplification of genomic sequences

Target coding sequences were amplified from template genomic DNA by PCR. Reactions were prepared in total aqueous volumes of 50 µl, containing 50% FailSafe™ PCR system pre-mix D (EPICENTRE® Biotechnologies), ~10⁵ molecules of genomic DNA, 50pmol of the appropriate forward and reverse primer and 1.25 U FailSafe™ PCR enzyme mix. 30 cycles were conducted at the following parameters:

- 95°C for 2 minutes
- 50°C for 2 minutes
- 70°C for 3 minutes.

Products were analysed and purified by agarose gel (1%) electrophoresis and gel extraction (QIAGEN) according to the manufacturer’s protocol. Where amplification was successful, four additional 50 µl reactions were prepared to increase the yield of DNA.
2.3.3 Agarose gel electrophoresis

Gels were prepared by dissolving 0.5 g agarose into 50 ml TAE buffer (1x) by gentle microwave heating in a 250 ml conical flask for approximately 1 minute. Ethidium bromide (2 µl) was then added to a final concentration of 0.4 µg ml\(^{-1}\). Gels were allowed to set for 30 minutes, placed into the electrophoresis tank and submerged in 1x TAE buffer (2.2.6). Loading dye was mixed with the DNA samples (1x final concentrate) before loading into the gel, and electrophoresis was conducted at 100 V for approximately 30-40 minutes. DNA markers (Fermentas) were also loaded in known amounts for size comparison, and to estimate the concentration of nucleic acid in the samples if necessary. Gels were visualised under a UV light source. Where DNA needed to be recovered, it was purified using a gel extraction kit (Qiagen) according to the manufacturer’s protocol.

2.3.4 Preparation of vector and target sequence

The required overproduction vector was transformed into OneShot® Top10 chemically competent E. coli cells for cloning (2.1.3). Around 0.5 µl of plasmid from the manufacturer’s stock was used in each case, according to their recommendation. Transformants were selected as single antibiotic-resistant colonies on DYT agar plates, and grown overnight at 37°C with shaking in 5 ml DYT media plus antibiotic at working concentration. Plasmid DNA was harvested from each overnight culture using QIAprep Spin Miniprep Kit (Qiagen).

The recovered plasmid DNA and the corresponding PCR product of the target sequence were treated with the appropriate restriction endonucleases for vector ligation, as indicated in Table 2.1.4. Digestions were set up in 50 µl reaction volumes, each containing 5 µg of DNA, with ~200 U of each enzyme and the supplied buffer (1x concentrate) (all New England Biolabs). Where necessary, BSA was added according to the manufacturer’s recommendation. Digestion reactions progressed for 4 hours at 37°C; those containing vector DNA were also treated for 2 hours with 2 U CIAP and the supplied buffer (1x concentrate) (Fermentas) to prevent re-annealing of the plasmid ends.

DNA was extracted by addition of 1/10 volumes of 3 M sodium acetate pH 5.4 and 4 volumes of ethanol, stored at -80°C for one hour, then centrifuged at 13000 x
rpm for 10 minutes to pellet the DNA precipitant. The pellets were washed in 70% ethanol, centrifuged for a second time and allowed to air-dry at room temperature after removal of the supernatant. The DNA was then re-suspended in 30 µl of TE buffer (2.2.6) and purified by agarose gel electrophoresis and gel extraction (Qiagen). DNA was eluted in sterile water for long-term storage at -20°C. Yields were measured by nanodrop spectrophotometry.

2.3.5 Ligation of constructs

Approximately 70 ng of vector was used for ligation with a 5x molar excess of each insert. Reactions also contained 400 U T4 DNA ligase and its supplied buffer (1x concentrate) (Fermentas), made up to 20 µl with sterile water. A control reaction was also conducted, containing restricted vector only, with no insert sequence. The ligation reaction was allowed to run for 2 hours at room temperature, and tubes were periodically cooled on ice for 1 minute every 10 minutes to promote annealing of the DNA inserts to the vector. Reaction components were removed from the resulting constructs by ethanol precipitation as earlier described (2.3.4). The resulting plasmids were then transformed into OneShot® Top10 chemically competent E. coli cells (2.1.3) and plated onto DYT agar containing antibiotic at working concentration. Ten colonies were selected as putative positive clones for each of the constructs after overnight growth at 37°C. These were then grown into 5 ml cultures overnight at 37°C with shaking in the presence of antibiotic. Plasmids were recovered from each culture by mini-prep (Qiagen) and eluted into sterile water, then samples were analysed for the presence of the DNA insert by restriction digestion and agarose gel electrophoresis. Those containing the insert sequence were checked for errors by DNA sequencing, which was performed using T7 forward and reverse primers at Geneservice, London (Source Bioscience).
2.3.6 Protein overproduction trials

Overproduction trials were conducted to determine the optimal IPTG concentration for induction. The resulting constructs were each transformed into the appropriate overproduction strain (2.3.1), which is indicated for individual preparations in the results sections. A single colony per construct was inoculated in 5 ml DYT media, and incubated with shaking at 37°C until the culture reached an OD<sub>600</sub> of 0.6 (3-4 hours in most cases). At this stage, a 200 µl sample of non-induced culture was taken for analysis by SDS-PAGE and western blotting. For each construct being tested, the 5ml culture was then split and induced with either 0.1 mM or 1 mM IPTG. Samples of each were taken again after one hour and three hours of induction. The sample volume was determined as 200 µl for an OD<sub>600</sub> of 0.6, or a proportion thereof, to give an approximate equal density of cells per sample. For example, if the OD<sub>600</sub> reached 1.2, 100 µl was taken. Samples were centrifuged at 6000 x rpm for ten minutes, and pellets were re-suspended in 20 µl 2x SDS loading buffer for SDS-PAGE and western blotting.

2.3.7 Large-scale protein overproduction

Constructs were transformed into the appropriate overproduction strain as described earlier (2.3.1). Starter cultures were grown by selecting a single colony from a fresh transformation plate, and inoculating cells into 50 ml of DYT broth plus antibiotic at working concentration. These were incubated in an orbital shaker for ~16 hours at 37°C. For a 2 L protein preparation, 1-10 ml of starter culture was added to four 2 L conical flasks containing 500 ml of DYT broth plus antibiotic; then the cultures were allowed to grow at 37°C in an orbital shaker until they reached mid-log phase, as measured by an OD<sub>600</sub> of 0.6-0.8. Protein overproduction was induced by adding the working concentration of IPTG as determined in small-scale trials. For most constructs, this was 1 mM IPTG for 3 hours at 37°C. Alternatively, slow protein overproduction was conducted using 0.1 mM IPTG for 16 hours at 20°C.

2.3.8 Cell lysis

The following protocol refers to lysis of a 2 L cell culture and where smaller cultures were prepared, volumes were adjusted accordingly. Cells were harvested
from overproduction cultures by centrifugation at 8000 x rpm and 4°C. The supernatant was decanted and the cell pellets from were resuspended in 40 ml ice-cold lysis buffer, which contained the appropriate binding buffer for the protein \((2.2.7)\) plus lysozyme, DNase I and protease inhibitors. Cells were disrupted by sonication using a probe sonicator, with the cell suspension kept on ice throughout. Typically, 2-5 minutes of sonication pulses at 40% amplitude, with 5 seconds on and 10 seconds rest, was sufficient to lyse the cell suspension; however this was judged on an individual basis depending on volume and viscosity and taking care not to let the suspension warm considerably. A key indication of successful lysis was a homogenous, less turbid and less viscous solution. Cell debris was removed by two rounds of centrifugation at 12000 x rpm and 4°C, and the supernatant was then decanted and passed through a 0.45 µm syringe filter.

2.3.9 Immobilized metal-ion affinity chromatography (IMAC)

All recombinant proteins produced in this study were engineered as N-terminal His-tag fusions so they could be purified from cell lysates by IMAC. Charged Ni\(^{2+}\)-NTA resin (Qiagen) was washed thrice in binding buffer for equilibration, then added to the cell lysate and left to incubate for 2 hours at 4°C. For a 2 L overproduction culture, 1.2 ml of Ni\(^{2+}\)-NTA resin was sufficient. Falcon tubes containing the lysate were left on a rotator during incubation to keep the resin suspended in solution and allow as much protein to bind as possible. After incubation, the suspension was transferred to a 10 ml gravity flow disposable column (Bio-Rad) and unbound proteins were eluted in the flow-through. The column resin was washed for 2 minutes in 4 ml wash buffer (20 mM imidazole) \((2.2.7)\), which was eluted to remove non-specific proteins that may be interacting weakly with the column. This step was repeated 3 times. The protein of interest was then eluted by washing the column resin in 1.5 ml elution buffer (200 mM imidazole) \((2.2.7)\) for 2 minutes and collecting the flow-through, and this step was repeated 4 times to ensure all of the protein had been eluted. Samples from each fraction were analysed by SDS-PAGE, and elution fractions containing the protein of interest were pooled together.
2.3.10 Dialysis buffer-exchange

Desalting and buffer-exchange were used both during the purification process to remove imidazole, and also at later stages when proteins needed to be transferred into an alternative buffer for specific experiments. After IMAC purification, imidazole was removed by dialysis against the appropriate standard buffer for the protein (2.2.7). Visking tubing was first soaked in distilled water at ambient temperature until it could easily be opened up. One end was secured with a clip, and the pooled IMAC eluate was transferred carefully to the tube with a Pasteur pipette. The tube was sealed at the remaining end with a clip and placed into a beaker containing 4 L of chilled standard protein buffer. The beaker was covered with cling film and left for 16 hours at 4°C, being constantly stirred with a magnetic stirrer. After incubation, the buffer was replaced with 2 L of fresh buffer and left for a further 4 hours.

2.3.11 Desalting column buffer-exchange

In some instances protein buffers were exchanged using PD-10 desalting columns (GE Healthcare) according to the manufacturer’s instructions. The column was equilibrated by washing several times in the required buffer before applying the sample, and the protein was then eluted using the required volume of buffer indicated by the column size.

2.3.12 Thrombin treatment

In some instances, the His-tag was removed from the protein of interest and this was conducted using thrombin protease (GE Healthcare). Protein concentration in the eluate was measured approximately by reading the absorbance at 280 nm on a nanodrop spectrophotometer, taking into account the predicted extinction coefficient. For 100 µg of protein, 1.5 U of thrombin was added and left to incubate for 16 hours at 4°C. For all proteins which underwent thrombin treatment in this study this was found to be sufficient for near complete cleavage. Thrombin treatment was often conducted simultaneously with the dialysis step, therefore added to the visking tubing with the eluate. After incubation, the cleaved protein was separated from any remaining uncleaved material by a reverse IMAC step. This essentially followed the procedure highlighted above, however used half the
Chapter 2

amount of Ni\textsuperscript{2+}-NTA resin. Protein with the His-tag successfully removed did not bind the column and was eluted in the first flow-through fraction, and the contents of each fraction were checked by SDS-PAGE analysis. The flow-through was incubated with Benzamidine Sepharose 6B (GE Healthcare) for 1 hour at 4°C, which was then removed by retention in a 10 ml disposable gravity flow column. This resin binds trypsin-like serine proteases such as thrombin, thus removing them from the sample and preventing further cleavage.

2.3.13 Concentration of purified protein

Protein samples were concentrated using VIVASPIN centrifugal concentrators according to the manufacturer’s guidelines (Sartorius). The concentrator membrane was first equilibrated with protein standard buffer, by 5 minutes centrifugation at the appropriate speed. The protein samples were kept at 4°C during concentration and monitored throughout to avoid precipitation. The samples were also mixed with a pasteur pipette every 5 minutes to avoid a local concentration of protein at the membrane. Generally, IMAC eluates from a 2 L culture could be concentrated to 0.5 ml volumes ready for size exclusion chromatography without precipitating, however on occasions where larger cell pellets had been harvested or the sample appeared more concentrated this was sometimes increased to 1 ml.

2.3.14 Size-exclusion chromatography (SEC)

SEC was performed by gel filtration using a Superdex 200 10/300 GL or a Superdex 75 10/300 GL column connected to an AKTA FPLC system. SEC was used as an extra purification step and to analyse the homogeneity of the protein sample and remove any aggregates. It was also used to estimate the oligomeric state of the protein, although this was often clarified by light scattering (2.3.15). Gel filtration columns were stored in 20% sterile-filtered ethanol; therefore, columns were washed with 40 ml of sterile water before use at a flow rate of 0.5 ml min\textsuperscript{-1}. This wash step was repeated with standard protein buffer for column equilibration. The protein sample was centrifuged for 30 minutes at 14000 x \textit{rpm} and 4°C in a microcentrifuge to remove any precipitant. This was then injected into a 500 µl loop connected to the AKTA FPLC, for loading onto the column. Fractions were collected in 0.4 ml aliquots and those containing protein according to the
chromatogram were analysed by SDS-PAGE where necessary. The required fractions were pooled and re-concentrated with centrifugal concentrators as before.

2.3.15 SEC with multi-angle laser light scattering (SEC-MALLS)

The molecular mass, oligomeric state and homogeneity of protein samples was analysed by SEC-MALLS where required. A sample of purified protein was passed through a Superdex 75 10/300 column (GE healthcare) connected to an AKTA FPLC. The column eluate was directed through an in-line DAWN HELEOS laser photometer at 658 nm and Optilab rEX refractometer (Wyatt technologies), and the scattered light intensity and refractive index were recorded. ASTRA software provided with the instrument was used to calculate the molecular weight and homogeneity of the sample.


2.4 High-throughput cloning and protein overproduction at OPPF

The majority of PilQ constructs were cloned and tested for small-scale protein overproduction using equipment and reagents at the Oxford Protein Production Facility (OPPF), with a specialized high-throughput protocol. Details and contacts can be found at www.oppf.ox.ac.uk. These methods are outlined in this section, separately from those cloning and overproduction trials performed at home. Some constructs produced at OPPF expressed well and were suitable for higher-level protein overproduction, which then followed the purification protocols outlined in sections 2.3.7-2.3.15.

2.4.1 High-throughput PCR

Oligonucleotide primers (2.1.1) were first designed for each construct for In-Fusion™ cloning into the pOPINF overproduction vector using the OPTIC bioinformatics tool (www.oppf.ox.ac.uk). The vector introduces an N-terminal histidine purification tag and a 3C-protease cleavage site. Primers were stored in a 96-well plate as 10 μM stocks in 10 mM Tris-HCl pH 8.0. Initial PCR reactions for all 47 constructs were conducted using the KOD Xtreme™ Hot Start DNA Polymerase master mix (Novagen), which was set up for a total of fifty 50 μl reactions as follows:

- KOD Hot Start Buffer (2x): 1250 μl
- dNTP mix (2 mM): 500 μl
- KOD Hot Start Polymerase (1 U μl⁻¹) 50 μl
- Template DNA 100 μl
- Sterile water 300 μl

Using a multi-dispense pipettor 42 μl of this master stock was dispensed into each well of a PCR plate on ice. Also, using a multi-dispense pipettor 3 μl of each of the 10 μM forward and reverse primer stocks was transferred to the appropriate wells of the PCR plate. The plate was sealed with a clear adhesive film and load into a Veriti PCR machine (ABI). Thermal cycling was performed with the following parameters:
Hot start:
- 94°C for 2 minutes

Followed by 29 cycles of:
- 98°C 10 seconds
- 60°C 30 seconds
- 68°C 4 minutes

Hold:
- 68°C 2 minutes
- 4°C Hold

PCR products were analysed by SDS-PAGE. A second round of PCR was then conducted for a few problematic targets with altered parameters using Phusion Flash master mix (NEB). A master stock was set up depending on the number of reactions required, so that each 50 µl reaction contained:

- Phusion Flash Master mix (2x) 25 µl
- Sterile water 18 µl
- Template DNA 3 µl

Reactions were dispensed into a PCR plate and primers were added as before. The parameters for the Phusion Flash PCR were set to:

Hot start:
- 98°C for 10 seconds

Followed by 29 cycles of:
- 98°C 1 second
- 60°C 5 seconds
- 72°C 1 minute

Hold:
- 72°C 2 minutes
- 4°C Hold
2.4.2 AMPure magnetic bead PCR purification

PCR products were then purified from the template by AMPure XP magnetic bead-based purification (Agencourt/ Beckman). The reactions were first treated with 5U Dpn I and incubated at 37°C for 1 hour to digest the template DNA. A multi-dispense pipettor was used to add 90 µl of AMPure beads to each PCR reaction in the 96-well plate. Each well was mixed by pipetting and left to sit for 5 minutes at ambient temperature, so that PCR fragments ≥100 bp could bind. After incubation, the reaction plate was placed onto a SPRIPlate 96R magnet (Beckman) for around 5 minutes to separate beads from solution. When the solution appeared clear, with the reaction plate still situated on the magnet a multi-dispense pipettor was used to aspirate the cleared solution from the wells and this was discarded. The beads were washed twice with 200 µl of 70% ethanol, which was dispensed into each well and left for 30 seconds at ambient temperature before being aspirated and discarded. The plate was left to air-dry for 20 minutes to allow complete evaporation of any residual ethanol. Clean PCR products were eluted from the beads by addition of 30 µl of 10 mM Tris-HCl pH 8.0 and mixing. The reaction plate was placed back on the magnet and the supernatant was carefully removed and transferred to a fresh plate. PCR products were checked by agarose gel electrophoresis.

2.4.3 InFusion™ ligation reactions

Ligation reactions for all 47 constructs were set up simultaneously in 96-well plates. First, the pOPINF vector had been cut with the appropriate restriction enzymes, and 1 µl was dispensed into each required well of a 96-well plate. 10 µl of PCR product (10-250 ng DNA) was transferred to a dry-down In-Fusion™ plate (Clontech) containing lyophilized enzyme and buffer and mixed well by pipetting to resuspend the dry pellet. The reactions were transferred to a fresh PCR plate, sealed with adhesive foil and incubated for 30 minutes at 42°C in a Veriti® thermocycler. After incubation, the plate was transferred onto ice and diluted with 40 µl of chilled TE buffer.
2.4.4 High-throughput transformation

Loose eppendorfs containing 50 µl aliquots of OmniMAX™ 2 cloning grade *E. coli* (2.1.3) were placed in racks following the format of the ligation reactions in the 96-well plate. 5 µl of each InFusion™ reaction was placed into each cell aliquot and left on ice for 30 minutes. Cells were then heat shocked for 30 seconds at 42°C, and rested on ice for a further 2 minutes. 300 µl of SOC medium (2.2.4) was dispensed into each tube and they were incubated at 37 °C for 1 hour.

Ligations were verified by blue/white screening. For this, LB Agar plates (2.2.4) were prepared in 24-well plates containing antibiotic at working concentration, 0.5 mM IPTG and 0.02% X-Gal. 1 ml of the agar was dispensed using a repeat pipettor into 94 wells to give two replicates of each experiment. Once solidified, 25 µl of each transformation reaction was dispensed onto half of the agar plates, and spread by gentle tilting of the plate. Cells in the transformation mixtures were then diluted 1:5 by addition of SOC medium and 25 µl was dispensed onto the remainder of the plates in the same manner. Plates were incubated overnight at 37°C. White, recombinant colonies were selected for high-throughput culturing.

2.4.5 High-throughput culture and plasmid recovery

Two white colonies per construct were picked which will normally give positive clones for ~95% of the PCR products. A 96-well deep-well block (BD Falcon) was prepared by addition of 1.5 ml of Power Broth™ (Molecular Dimensions) supplemented with the appropriate antibiotic at working concentration to each well. Individual colonies were picked with sterile pipette tips and placed into each well. When the block was complete tips were removed and the block was sealed with a gas-permeable adhesive seal (ABgene). The filled block was placed in an incubator at 200-225 x rpm and 37°C overnight. The gas-permeable seal was replaced with a foil seal and the cells were harvested by centrifugation at 5000 g for 15 minutes at 4°C, using a Beckman JS5.3 rotor and a Beckman Avanti centrifuge. The medium was decanted off the cell pellets, and the plasmid mini-preps were performed using a Qiagen BioRobot 8000. Plasmids were verified by another round of PCR as described earlier to check for the insert. This resulted in positive clones for the full 47 target sequences.
2.4.6 High-throughput protein overproduction trials

All 47 constructs plus a control expressing GFP were transformed into both *E. coli* B834 (DE3) (Novagen) and *E. coli* Rosetta (DE3) LysS (Invitrogen) for small-scale overproduction trials using the high-throughput transformation procedure described above (2.4.4) with the following alterations: 3 µl of mini-prepped plasmids was added to the competent cell aliquots, and 1200 µl of Power Broth™ was added during the 1 hour recovery period. LB agar plates were prepared containing the appropriate antibiotic for the plasmid, with the addition of 170 µg ml⁻¹ chloramphenicol to the Rosetta strain, in order to retain the pRARELysS plasmid.

For IPTG induction, single colonies were grown in 0.5 ml Power Broth™ starter cultures supplemented with the appropriate antibiotic(s) in deep-well blocks with gas-permeable seals in an orbital shaker overnight at 37°C. 62.5 µl of overnight culture was transferred to fresh deep-well blocks containing 2.5 ml of Power Broth™ plus antibiotic(s) for protein overproduction cultures. These were grown in an orbital shaker at 37°C to OD⁶₀₀ of ~0.5 (3-5 hours) before induction. Cultures were induced by addition of 0.5 mM IPTG, and left to produce protein overnight in an orbital shaker at 20°C. All constructs were also tested for protein production in both strains using Overnight Express™ Instant TB (TB-ONEX) auto-induction medium (Novagen), in the same manner but without addition of IPTG. These were induced for 20-24 hours at 25°C. 1 ml of each culture was transferred to a fresh deep-well block, taking care to keep to the matrix format of the plates, and sealed with a solid foil seal. Cells were harvested by centrifugation at 5000 g for 15 minutes at 4°C within the block plates, and pellets were frozen at -80°C.

2.4.7 Ni²⁺-NTA robotic protein production screening

Frozen cell pellets within deep-well blocks were broken by thawing and resuspension in 210 µl Qiagen Lysis Buffer (2.2.8) supplemented with 1 mg ml⁻¹ lysozyme and 40 µg ml⁻¹ DNase Type I. After 30 minutes, the lysate was cleared by centrifuging the deep-well blocks for 15 minutes at 5000 g and 4°C. The soluble fractions were transferred to magnet-compatible micro-titre plates (MTP), containing 20 µl Ni-NTA magnetic bead suspension (Qiagen) in each well. The plates were then left on a MTP-shaker for 30 minutes at room temperature, before being placed on a 96-well magnet (Qiagen, type A or B magnet) so that the lysate
could be pipetted carefully off the beads. The beads were washed with 200 µl Qiagen Wash Buffer (2.2.8), and mixed on the MTP-shaker for 5 minutes. The plates were transferred back to the magnet for 1 minute and the buffer was carefully aspirated. This wash step was repeated twice. The protein of interest was eluted from the beads by addition of 50 µl Qiagen Elution Buffer (2.2.8) to each well, 1 minute mixing on the MTP-shaker and 1 minute on the magnet. The eluate was carefully removed using a multi-channel pipette and transferred to a fresh MTP, maintaining the matrix format of the plates. Proteins were analysed by SDS-PAGE as outlined in section 2.5.1.
2.5 Analysis of proteins

2.5.1 SDS-PAGE

Protein samples were analysed by SDS-PAGE multiple times throughout the project to assess protein production, purity and approximate yield of proteins wherever necessary.

Protein samples were first mixed with sample loading buffer (1x concentrate) (2.2.9) and heated to 90°C for 10 minutes before loading onto the gel. For the majority of cases, pre-cast 10% or 12% NuPage Bis-Tris gels (Invitrogen) were used in a vertical electrophoresis tank filled with MES or MOPS running buffer (2.2.9), and run at 180 mV for approximately 45 minutes, or until the dye had migrated through the gel. In some instances, gels were prepared manually using commercially available cassettes (BioRad). An 18% polyacrylamide resolving gel (2.2.9) was poured into the cast to about ¾ full, which was allowed to set before applying a 6% stacking gel (2.2.9) on the surface and placing a plastic comb in the top of the stacking gel. For these gels 1 x SDS running buffer (2.2.9) was used, and gels were allowed to run at 100 mV until the dye had reached the base of the stacking gel, then increased to 200 mV for the remainder of the run. Gels were stained with commercially available Instant Blue (Expedon) coomassie-based stain for approximately 1 hour, or until the bands were clearly visible. Stain was rinsed away with distilled water and gels were photographed using a standard PC scanner or a digital camera.

2.5.2 Western blot analysis

Protein samples were analysed for the presence of His-tagged protein by western blot analysis. After SDS-PAGE without coomassie staining, a nitrocellulose membrane was placed over the gel, and the two were sandwiched between 6 sheets of filter paper, which had been pre-soaked in western transfer buffer (2.2.10). The gel, membrane and papers were placed onto the transfer apparatus cathode so that the gel faced the cathode with the nitrocellulose on top. The anode was clamped into place and the proteins were transferred onto the nitrocellulose membrane for one hour at 220 mA. The membrane was washed twice for ten minutes in TBS buffer (2.2.10), incubated in blocking buffer (2.2.10) for one hour at
room temperature then washed a further two times in TBS. The membrane was then incubated in Anti-TetraHis antibody (Qiagen) which had been diluted 1:2000 with blocking buffer. After one hour at room temperature, it was washed twice for ten minutes in TBS-Tween/ Triton (2.2.10) and once for ten minutes in TBS. The secondary antibody, Anti-Mouse IgG Alkaline Phosphatase Conjugate (Sigma), was diluted 1:3000 in blocking buffer and the membrane was incubated with this for one hour at room temperature. The membrane was then washed four times, each for ten minutes in TBS-Tween/ Triton, and stained with alkaline phosphatase solution (2.2.10). The nitrocellulose membrane was rinsed thoroughly with sterile water to stop the colouring reaction.

2.5.3 Measuring protein concentration

Protein concentration was measured using the BCD protein assay kit (Pierce) according to the manufacturer’s recommendations. In summary, 1 ml of a mixed BCA and copper sulphate reagent was added to 50 μl of protein solution, including the sample to test and albumin samples of known concentration. These were incubated for 1 hour at 37°C, then cooled, and the absorbance was measured at 562 nm. The absorbance readings for standards were plotted against the concentration value and this was used to determine the concentration of the unknown sample.

Alternatively, concentrations were measured with a nanodrop spectrophotometer, which measures the absorbance at 280 nm taking into account the theoretical extinction coefficient for the protein based on its amino acid sequence.
2.6 Protein structural studies

2.6.1 Preparation of crystal trials and cryoprotection of crystals

Preliminary crystal trials were conducted using the sitting-drop vapour diffusion technique in 96-well MRC plates (Molecular Dimensions). A mosquito™ nanolitre pipetting robot (TTP Labtech) was used to dispense and mix 100-300 nl of protein sample, or control sample containing only protein buffer, with 100-300 nl of liquor solution from the following commercially available crystallization screens (Qiagen):

| Classics suite | pH clear suite | Cryos suite |
| Classics II suite | JCSG core I suite | Cations suite |
| PEG suite | JCSG core II suite | Opti-salts suite |
| PEG II suite | Anions suite |

The wells were sealed with ClearSeal Film (Hampton Research) and analysed immediately for precipitate using a light microscope. Drops were monitored for progress intermittently over 14 days for any initial ‘hit’ solution, which may have formed crystals, micro-crystals or needle crystals for instance; however, trays were retained until the drops had dried out. Crystal trials were initiated at 4°C and 20°C.

Successful hits were obtained when 300 nl of PaPilP\(^{c}\) (10 mg ml\(^{-1}\)) in 50 mM Tris-HCl pH 9.0, 200 mM NaCl was mixed with 300 nl of the well solution. Two crystal forms were grown, using different conditions; for the \(C2\) crystal form, the well solution comprised 0.1 M Tris-HCl pH 8.5 and 0.6 M zinc acetate. For the \(P2_1\) crystal form the well contained 0.1 M HEPES pH 7.5, 8% (v/v) ethylene glycol, 20% (w/v) PEG 10,000. Each crystal was cryoprotected by careful removal from the well using a purpose made 0.2 mm CryoMount set loop (Molecular Dimensions), and sequential washing into crystallization buffer (well solution) plus 5%, 10%, 15% and finally 20% (v/v) ethylene glycol (\(P2_1\) crystal) or glycerol (\(C2\) crystal), allowing 2-3 minutes between each wash. Crystals were rapidly frozen in liquid nitrogen.

2.6.2 Data acquisition

Data for the \(P2_1\) crystal form were collected on a Rigaku MicroMax007 rotating anode X-ray generator and R-AXIS IV++ image plate detector, and data for the \(C2\)
crystal form were collected at a synchrotron source (Diamond Light Source, Oxfordshire, UK). Data were processed as detailed in the results (5.6).

2.6.3 Preparation of $^{15}$N- and $^{13}$C-labelled protein samples for NMR

After transformation, starter cultures were grown as normal (2.3.7) in DYT plus antibiotic. After 16 hours, 5 ml of starter culture was gently centrifuged for ten minutes at 5000 x rpm and 4°C to pellet the cells. The cell pellet was washed carefully with sterile water to remove any residual media. It was then used to inoculate 500 ml of M9 minimal media (2.2.5) plus antibiotic, by gently resuspending in about 10 ml of the media first and then transferring it to the flask under aseptic conditions. The cultures were grown to mid-log phase before induction, which was usually with 1 mM IPTG overnight at 37°C, as cell growth can be slower in minimal media.

2.6.4 Data collection$^{154}$

NMR was used to analyse purified protein samples consisting of either natural isotopic abundance (2.3.7) or 98% $^{15}$N and 99% $^{13}$C labelled protein (2.6.3) at a concentration range of 250 μM-1 mM. Proteins were exchanged if necessary by desalting into 50 mM NaH$_2$PO$_4$/ Na$_2$HPO$_4$, 50 mM NaCl pH 6.8, 90% $^1$H$_2$O 10% $^2$H$_2$O.

NMR experiments were conducted at 298 K using Bruker Avance III 600 MHz and 800 MHz spectrometers equipped with TCI triple resonance cryoprobes. Spectra were processed using Topspin2.1 (Bruker) and the Azara processing package provided as part of the CCPN suite, and assignment was carried out using CCPN Analysis.$^{155}$

Triple resonance assignment was obtained utilising two-dimensional $^1$H$^{13}$C and $^1$H$^{15}$N HSQCs in conjunction with 3D HNCA, HN(CA)CB, HN(CO)CA, HNCO, CBCA(CO)NH, HBHANH, HBHA(CO)NH, CB(CGCD)HD and HCCH-TOCSY experiments. Distance restraints were obtained from 3D $^1$H$^1$H$^{13}$C NOESY and $^1$H$^1$H$^{15}$N NOESYs using a mixing time of 100 ms.$^{154}$
2.6.5 NMR solution structure calculation & validation\textsuperscript{154}

Automated NOESY assignment and preliminary structure calculations of \textit{NmPilQ}\textsuperscript{224-329} and \textit{NmPilQ}\textsuperscript{343-442} were performed using CYANA 2.1 software,\textsuperscript{156,157} with input data of shift lists derived from $^{15}$N- and $^{13}$C-HSQC spectra to assign peaks. CYANA 2.1 calculations ran with standard protocols using 7 cycles of automated NOE assignment and structural calculations, producing 100 structures per cycle. Structures were subsequently water-refined using CNS1.2\textsuperscript{158} with unambiguous inter-proton distance restraints and dihedral restraints. Dihedral $\phi$ and $\psi$ torsion angles were produced by TALOS+\textsuperscript{159} and the final ensembles of the best 20 water-refined structures were selected on the basis of low total and NOE energies, and validated with PROCHECK-NMR\textsuperscript{160} using the iCing interface (http://nmr.cmbi.ru.nl/icing/iCing.html). Secondary structure was calculated using the STRIDE webserver.\textsuperscript{161,162} Random coil index (RCI) analysis was carried out using the RCI webserver.\textsuperscript{163,154}

2.6.6 \textit{NmPilQ}\textsuperscript{343-545} model\textsuperscript{154}

The structure of the PilQ tandem N0/N1-domain construct, \textit{NmPilQ}\textsuperscript{343-545} presented in chapter 4, could not be determined using conventional methods due to the low number of unambiguous NOEs extracted using the methods above. Therefore the CS-ROSETTA webserver\textsuperscript{164} was used for chemical shifts-based structure determination of residues 419-515. The $^{1}$C, $^{\mathrm{H}_{a}}$, $^{\mathrm{C}_{a}}$, $^{\mathrm{C}_{b}}$, N, and $^{\mathrm{H}_{N}}$ assigned chemical shifts were submitted to CS-ROSETTA\textsuperscript{164} and the mid-protein domain boundary was identified using RCI;\textsuperscript{163} information which was useful for design of the shorter construct which could be determined by conventional methods (\textit{NmPilQ}\textsuperscript{343-442}). The tandem N0/N1-domain structure model at this stage therefore comprised an NOE-derived structure for residues 343-442 and a CS-ROSETTA-derived model for the remainder of the construct.\textsuperscript{154}

A complete final model of the \textit{NmPilQ}\textsuperscript{343-545} tandem domain was assembled using MODELLER\textsuperscript{165} in multiple template mode. For the N0-domain the \textit{de novo} \textit{NmPilQ}\textsuperscript{343-442} NMR structure was used as a template and for the linker region the CS-ROSETTA model structures spanning PilQ residues 419-437 were used. For the N1-domain, homologous domains from the EscC and GspD structures (PDB ID 3GR5 residues 105-173 and PDB ID 3EZJ residues 102-168 respectively),
together with the CS-ROSETTA model, were used. Each template consisted of only one folded domain. As no restrictions were placed on inter-domain orientation, 100 structures were calculated and clustered accordingly. These 100 structures were divided into 7 clusters based on the criterion that the RMSD of the cluster be no more than 3 Å and each cluster must comprise four or more structures. For each cluster a representative model was selected as the closest to mean structure. The best approximate orientation was then selected from these structures based on quality of fit to the electron microscopy density map using MULTIFIT, as described in the main discussion, chapter 7.

Molecular graphics, including images of protein structures, analysis of superimposed structures and analysis of surfaces, were produced using COOT, CCP4MG and UCSF CHIMERA softwares and these are indicated with reference to specific figures in the legend.
Chapter 2

2.7 Protein interaction studies

2.7.1 Isothermal titration calorimetry (ITC)

ITC was used in this project to probe an interaction between NmPilQ\(^{343-545}\) and NmPilP\(^C\). Freshly purified samples of both proteins were dialysed simultaneously against 50 mM NaH\(_2\)PO\(_4\) 100 mM NaCl pH 7.0, and centrifuged at 14000 x rpm for 30 minutes at 4\(^\circ\)C to remove any precipitant. Experiments were conducted using an ITC200 MicroCalorimeter instrument (MicroCal, GE Healthcare). The instrument syringe component was filled with a 1 mM NmPilP\(^C\) sample, taking care not to introduce any bubbles. Using a glass syringe, 200 \(\mu\)l of the NmPilQ\(^{343-545}\) sample was carefully injected into the cell chamber of the instrument, again without producing air bubbles. Experiments were conducted with the concentration of this sample at both 200 \(\mu\)M and 100 \(\mu\)M, and the temperature was maintained at 25\(^\circ\)C throughout. The instrument was programmed so that NmPilP\(^C\) was added into the cell in sequential 1 or 2 \(\mu\)l injections at 90-second interludes. The contents of the cell were continually stirred and the enthalpy changes were measured and processed using the ITC200 supplied software with Origin\textsuperscript{®}7.0.

2.7.2 Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were conducted using the ProteOn\textsuperscript{TM} XPR system (BioRad) to verify an interaction between NmPilQ\(^{343-545}\) and NmPilP\(^C\). Purified samples of both proteins were buffer exchanged using a desalting column into PBS (2.2.7). 2500 response units of NmPilP\(^C\) were immobilized to a PRoteon\textsuperscript{TM} GLC sensor chip by amine coupling via lysine residues. Six concentrations, ranging from 10-60 nM, of NmPilQ\(^{343-545}\) were loaded into separate lanes on the instrument and passed over the surface of the sensor chip.

2.7.3 NMR chemical shift perturbation (NMR-CSP)\(^{154}\)

NMR-CSP was used to probe the interaction between NmPilP\(^C\) and two recombinant domains from NmPilQ: NmPilQ\(^{343-545}\) and NmPilQ\(^{343-442}\). For the first titration experiment a 50 \(\mu\)M sample of \(^{15}\)N-labelled PilP\(^{77-164}\) was prepared in 50 mM NaH\(_2\)PO\(_4\)/ Na\(_2\)HPO\(_4\) pH 6.8, 50 mM NaCl, 90% \(^1\)H\(_2\)O, 10% \(^2\)H\(_2\)O. Purified
natural isotopic abundance NmPilQ\textsuperscript{343-545} in the same buffer was titrated into the NMR tube containing the NmPilP\textsuperscript{C}, and spectra were recorded at the following ratios of [PilP]:[PilQ]: 1:0.1, 1:0.2, 1:0.5, 1:0.8 and 1:1 with a final concentration of 40 μM NmPilP\textsuperscript{C}. 2D \textsuperscript{1}H\textsuperscript{15}N-HSQC experiments were carried out with 2 scans and 256 increments. This identified a set of peaks which attenuated during the titration. To ensure peaks were not lost due to dilution effects at higher titration points two datasets were acquired, one with low and the other with higher numbers of scans; in all cases, it was possible to confirm that the loss of peaks was due to binding rather than dilution.\textsuperscript{154}

To investigate the binding site on NmPilQ, natural isotopic abundance NmPilP\textsuperscript{C} was titrated into \textsuperscript{15}N\textsuperscript{13}C-NmPilQ\textsuperscript{343-442} or NmPilQ\textsuperscript{343-545} for the reverse experiment. Spectra were collected at titration points of [NmPilQ]:[NmPilP\textsuperscript{C}]; 1:0.1, 1:0.2, 1:0.3, 1:0.5 and 1:1. All spectra were collected and processed using Topspin 2.1 (Bruker, Biospin).

The NmPilP\textsuperscript{C} \textsuperscript{1}H\textsuperscript{15}N-HSQC was assignment was obtained by transfer from BMRB star file 7209\textsuperscript{113} using CCPN format converter and CCPN analysis.\textsuperscript{155} 93\% of the amide backbone assignment was achieved for residues 79-163 based on closest singly matched peaks between BMRB reference assignment and experimental data. Peaks affected by the binding of NmPilQ were mapped onto the previously determined NmPilP structure\textsuperscript{113} and colour-coded according to the concentration ratio where the backbone amide peak was attenuated.
Chapter 3: High-throughput protein production, purification and characterization of individual subdomains from the type IV pilus dependent secretin PilQ

3.1 Introduction

The secretins form a large and unique group of integral outer membrane proteins which mediate a wide variety of translocation functions for at least four multi-protein secretion systems in gram-negative bacteria. In comparison to other OMPs, they differ with respect to the complexity of their overall structure and domain architecture, most likely owing to the requirement for passage of folded proteins and assembled fibres. For example, other bacterial OMPs such as the porins, autotransporters such as NalP, the Omp85 family and the ‘usher’ components of the chaperone/ usher pathway such as FimD and PapC, all adopt rigid β-barrel structures which impose functional constraints restricting the size of substrates according to the diameter of the pores they form. These barrels form the transmembrane section, with minimal periplasmic components such as small functional domains or loop regions. In contrast secretins form large, homo-multimeric, cylindrical vestibules of 12-15 subunits, which reach far into the periplasmic space. A large, central cavity of 50-80 Å would be sufficient for the passage of most protein substrates, but whether this cavity forms a continuous channel remains debatable. They also often require additional accessory proteins, such as pilotins, for correct assembly within their specific multi-protein secretion system. There is little structural information available, and none at atomic resolution, to provide clues as to how secretin monomers associate within the membrane to form multimeric assemblies. The relatively low number of predicted β-strands contained within a conserved transmembrane domain suggests that individual monomers may contribute to the wall of a large continuous barrel, rather than by the lateral organization of individual barrels. A similar assembly may be seen in the published structure of TolC from E. coli. The transmembrane β-domain of the TolC trimer forms a 12-stranded antiparallel β-barrel with four strands contributed by each monomer. The periplasmic domains then form an α-helical barrel which reaches down into the periplasm.
3.2 Domain organization of secretins

The secretin family comprises a variable array of members across several secretion systems, yet there is common, modular domain architecture across the group. Figure 3.2.1 gives a schematic representation of the typical domain organization of secretins from different systems. Bioinformatic analysis of secretin sequences indicates that they share a canonical C-terminal domain, referred to as the secretin domain or C-domain, which has been established as the transmembrane region and is also required for multimerization.\textsuperscript{31,172} This region is predicted to contain solely amphipathic beta secondary structure, which has been supported by experimental evidence,\textsuperscript{31} and is therefore likely to be embedded within the membrane. Proteolysis of the full length, purified, dodecameric PulD secretin from the T2SS leaves the C-domain as an intact multimer,\textsuperscript{33} and this has similarly been demonstrated for its homologue XcpQ.\textsuperscript{31} There is high sequence conservation within the C-domain, which diverges across the family towards the N-terminal regions where there is little homology. Sequence similarity in these regions tends only to be found between secretins belonging to the same secretion systems.

There is structural conservation within the central and N-terminal regions of some secretins, which are predicted to contain a similar arrangement of mixed alpha and beta secondary elements. Recent crystal structures of these domains from the T2SS\textsuperscript{34} and T3SS\textsuperscript{35} secretins have shown that, despite lack of sequence homology in these N-terminal regions, they do indeed share a similar alpha/beta fold. In the case of GspD from the T2SS, two of these domains (N1 and N2) which are central within the modular arrangement share close structural homology and 25% sequence identity with each other.\textsuperscript{34} In addition, they share a similar alpha/beta fold to that of the eukaryotic hnRNP K-homology domain (KH-type domain), which interestingly functions as a nucleic acid recognition motif.\textsuperscript{36} The PilQ secretin of the T4PB system is important for the uptake of DNA from the environment, and it has been shown to bind DNA in the case of both Neisseria\textsuperscript{176} and the naturally highly competent Thermus\textsuperscript{177,178} bacteria. In addition, some components of the T4PB system are actually shared with the macromolecular competence system for DNA uptake, which has been characterized in Haemophilus influenzae\textsuperscript{179,180} and Actinobacillus actinomycetemcomitans.\textsuperscript{181} The
extra-membranous domains of HofQ, the secretin component of the competence system of *A. actinomycetemcomitans*, have been structurally characterized, and this revealed that they share the KH-type domain fold and also include a GxxG DNA-binding motif, which binds double-stranded DNA.\(^{182}\)

**Figure 3.2.1: Schematic representation of the domain organization of secretins**

Members of the secretin family share a high degree of conservation at their C-terminus, which decreases towards the N-terminus. Only secretins found within the same secretion systems tend to share significant homology in their N-terminal domains. Secretins have a multi-domain organization, with a C-terminal transmembrane domain predicted to be of beta-strand secondary structure content (grey). Secretins of the T2SS and T3SS also have a region at their extreme C-terminus, sometimes referred to as the S-domain (blue), which binds to its cognate pilotin in the periplasm during outer membrane targeting. Secretins contain repeat domains of predicted alpha/ beta secondary content in their central and N-terminal regions, some of which have been characterized as KH-type (N1-N3) or N0-domains, depending on the arrangement of alpha and beta elements. Secretins of the T4PB system have one or two putative domains of predicted beta-strand content at their N-terminus, which appear to be unique to the type IV pilus biogenesis system (green). Secretins also have an N-terminal periplasmic export sequence, which is cleaved by a signal peptidase (yellow).
The first N-terminal domain in the GspD and EscC crystal structures, the N0-domain, has a βαβαβαβ fold, which differs slightly from that of the KH-type domains. This domain is structurally related to a signalling domain found in TonB-dependent outer membrane receptors. As structural data for the secretin family is limited, it is difficult to unambiguously distinguish N0-domains from KH-type (N1-N3) domains using sequence alignments and secondary structure prediction without relying on some guesswork. For simplicity, constructs used in this chapter that encode domains of mixed alpha and beta elements are collectively described as putative KH-type domains, where structural data is not otherwise available.

Secretins within the T4PB systems, known generically as PilQ, appear to contain additional domains at the extreme N-terminus that are not found by prediction algorithms in secretins from other systems. These domains would potentially reach the furthest into the periplasmic space, towards the inner membrane components of the pilus machinery. Analysis of these regions by secondary structure prediction algorithms suggests they consist of one or more beta-rich subdomains. There is as yet no experimental evidence suggesting a structure or function for these domains, which are unique to the pilus system. PilQ from *Thermus thermophilus* forms an exception whereby bioinformatic analysis of the amino acid sequence suggests it does not contain the N-terminal beta-rich subdomains. Instead, repeats of the more structurally conserved alpha/beta or KH-type domains are apparent. The *Thermus* PilQ complex spans the periplasm and has been found to co-fractionate with the inner membrane as well as the outer membrane, albeit in minor amounts. The differences between *Thermus* PilQ and that of other bacteria may have evolved due to requirements of the thermophilic organisms for a more stable heat-resistant complex or for the DNA translocation function of the protein, which has a dual occupation within the T4PB and competence systems. It has recently been suggested that N-terminal regions of PilQ from *Thermus thermophilus* are important for multimerization, in contrast to other PilQ secretins where multimer formation is believed to be governed only by the C-domain.

Several previous attempts have highlighted a notorious difficulty in expression and purification of full-length secretins due to their propensity to form stable multimers via their transmembrane domain. Most studies have highlighted the heat and SDS-resistance of purified secretin complexes. Also, detergent extraction of
full-length secretins can cause the N-terminal regions to destabilize and become disordered.\textsuperscript{30} Soluble, homogenous samples for atomic resolution structural and \textit{in vitro} biochemical studies are therefore difficult to obtain. As was the case for GspD and EscC, exploiting the modular domain architecture may provide a way to gradually build a full picture of the assembly piece by piece. There has even been some recent progression in extracting the C-domain from inclusion bodies\textsuperscript{186} although as yet structural information is lacking.

This chapter explores the use of bioinformatic analysis where there is little available structural information to predict domain boundaries within the T4PB secretin PilQ. Using a high-throughput method, various individual and combinations of subdomains from a range of PilQ homologues were cloned and tested for soluble protein production. A small selection of constructs, which produced good levels of protein, was advanced to large-scale protein production, purification and characterization.

### 3.3 Selection of targets

PilQ homologues were chosen from a range of organisms that have been shown to express type IV pili on their surfaces. These were \textit{Xyella fastidiosa},\textsuperscript{187} \textit{Legionella pneumophila},\textsuperscript{188} \textit{Xanthomonas campestris},\textsuperscript{189} \textit{Aeromonas hydrophila},\textsuperscript{190} \textit{Ralstonia solanacearum},\textsuperscript{191} \textit{Pseudomonas aeruginosa},\textsuperscript{192} \textit{Neisseria meningitidis}\textsuperscript{193} and \textit{Thermus thermophilus}.\textsuperscript{194} A detailed PROMALS\textsuperscript{195} alignment of these chosen PilQ sequences based on sequence conservation and predicted secondary structure content is shown in figure 3.3.1. The use of a range of homologues maximises the chances of success in later structural studies, as even slight variations in amino acid sequence can affect protein stability and behaviour. In addition, domain boundaries were predicted computationally according to algorithms that are intrinsically inaccurate. Cloned domains that are too short may miss crucial residues that are needed for correct folding and thus render them inactive for biochemical studies. On the other hand, including additional unstructured linker regions may affect solubility, their ability to crystallize or their potential for structure determination by NMR. An assortment of domain combinations were therefore selected, including individual KH-domains, double KH-domains, individual and double beta-domains and combinations of the two. As
an example, the extents of three of the chosen targets from *Neisseria* - namely the beta-1 (B1) domain, the beta-2 (B2) domain and the KH-type double domain - are highlighted in yellow, green and blue respectively on the alignment in *figure 3.3.1*. Two or three residues from putative linker regions either side of the predicted domain boundary were generally included in the target selection to avoid truncating full domains. Some domains were also tested by cloning more than one construct using slightly different domain boundaries, particularly where the secondary structure prediction was less clear, as was the case for *Thermus thermophilus* PilQ.
Conservation:

thermas 1 16469.0 ------------------------------------------------------------------------ 40
neisseria 1 2678.0 -------- RKLTIKSLGLYFA7AA ----------------------------------------------------- 18
pseudomonas 1 3609.0 ------------------------------------------------------------------------ 17
aroanaes 1 1182.0 -------- RICARNSTYFQTYQ ----------------------------------------------------- 20
legionella 1 1566.0 -------- LIAKLIISALI ----------------------------------------------------- 16
raltsonia 1 1479.0 DpvvLQYQTQRQGDQCAAYWLYSSLCHNlvQQLIQQ ----------------------------------------------------- 40
syalla 1 342.0 ------------------------------------------------------------------------ 1
xanthomonas 1 514.0 ------------------------------------------------------------------------ 1

Consensus ss:

h

Conservation:

thermas 49 374DVAYXALADSYVLQdPVYKANDSCQERALK SDPLTVKLKRFQYQ DQKLVRADLFAYAQNYL ----------------------------------------------------- 118
neisseria 1 2678.0 -------- RKLTIKSLGLYFA7AA ----------------------------------------------------- 18
pseudomonas 1 3609.0 ------------------------------------------------------------------------ 17
aroanaes 1 1182.0 -------- RICARNSTYFQTYQ ----------------------------------------------------- 20
legionella 1 1566.0 -------- LIAKLIISALI ----------------------------------------------------- 16
raltsonia 1 1479.0 DpvvLQYQTQRQGDQCAAYWLYSSLCHNlvQQLIQQ ----------------------------------------------------- 40
syalla 1 342.0 ------------------------------------------------------------------------ 1
xanthomonas 1 514.0 ------------------------------------------------------------------------ 1

Consensus ss:

h

Conservation:

thermas 110 386PFVYVYAPFQVITKFGAPEDCTAHILEKTYVGTKQAAPLPSVPV skinaviQGdsvQ ---------------------------------- 106
neisseria 1 2678.0 -------- RKLTIKSLGLYFA7AA ----------------------------------------------------- 18
pseudomonas 1 3609.0 ------------------------------------------------------------------------ 17
aroanaes 1 1182.0 -------- RICARNSTYFQTYQ ----------------------------------------------------- 20
legionella 1 1566.0 -------- LIAKLIISALI ----------------------------------------------------- 16
raltsonia 1 1479.0 DpvvLQYQTQRQGDQCAAYWLYSSLCHNlvQQLIQQ ----------------------------------------------------- 40
syalla 1 342.0 ------------------------------------------------------------------------ 1
xanthomonas 1 514.0 ------------------------------------------------------------------------ 1

Consensus ss:

h

Conservation:

thermas 187 375FLFFLSXALAWGA5WItWEUGELPAVTSVTAT7 ---PROMHSFSLRAGHEFPLEDALQq ---------------------------------- 248
neisseria 1 2678.0 -------- RKLTIKSLGLYFA7AA ----------------------------------------------------- 18
pseudomonas 1 3609.0 ------------------------------------------------------------------------ 17
aroanaes 1 1182.0 -------- RICARNSTYFQTYQ ----------------------------------------------------- 20
legionella 1 1566.0 -------- LIAKLIISALI ----------------------------------------------------- 16
raltsonia 1 1479.0 DpvvLQYQTQRQGDQCAAYWLYSSLCHNlvQQLIQQ ----------------------------------------------------- 40
syalla 1 342.0 ------------------------------------------------------------------------ 1
xanthomonas 1 514.0 ------------------------------------------------------------------------ 1

Consensus ss:

G 5 5 5 5

Conservation:

thermas 160 382PLADBAYQALADPTSNNSOEDPTPKQQALLA9PTBPSVY8FLDF5APQQA SAMPQGALFQ ----------------------------------------------------- 209
neisseria 1 2678.0 -------- RKLTIKSLGLYFA7AA ----------------------------------------------------- 18
pseudomonas 1 3609.0 ------------------------------------------------------------------------ 17
aroanaes 1 1182.0 -------- RICARNSTYFQTYQ ----------------------------------------------------- 20
legionella 1 1566.0 -------- LIAKLIISALI ----------------------------------------------------- 16
raltsonia 1 1479.0 DpvvLQYQTQRQGDQCAAYWLYSSLCHNlvQQLIQQ ----------------------------------------------------- 40
syalla 1 342.0 ------------------------------------------------------------------------ 1
xanthomonas 1 514.0 ------------------------------------------------------------------------ 1

Consensus ss:

---------------
<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Conservation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermus</td>
<td>249</td>
<td>572</td>
</tr>
<tr>
<td>neisseria</td>
<td>210</td>
<td>480</td>
</tr>
<tr>
<td>pseudomonas</td>
<td>186</td>
<td>214</td>
</tr>
<tr>
<td>aeromonas</td>
<td>129</td>
<td>283</td>
</tr>
<tr>
<td>legionella</td>
<td>161</td>
<td>203</td>
</tr>
<tr>
<td>ralstonia</td>
<td>50</td>
<td>226</td>
</tr>
<tr>
<td>myxella</td>
<td>41</td>
<td>137</td>
</tr>
<tr>
<td>Consensus_s:</td>
<td></td>
<td>hhhhh</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Conservation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermus</td>
<td>286</td>
<td>335</td>
</tr>
<tr>
<td>neisseria</td>
<td>277</td>
<td>324</td>
</tr>
<tr>
<td>pseudomonas</td>
<td>270</td>
<td>270</td>
</tr>
<tr>
<td>aeromonas</td>
<td>254</td>
<td>306</td>
</tr>
<tr>
<td>legionella</td>
<td>204</td>
<td>259</td>
</tr>
<tr>
<td>ralstonia</td>
<td>227</td>
<td>259</td>
</tr>
<tr>
<td>myxella</td>
<td>116</td>
<td>260</td>
</tr>
<tr>
<td>Consensus_s:</td>
<td></td>
<td>hhhhh</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Conservation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermus</td>
<td>356</td>
<td>405</td>
</tr>
<tr>
<td>neisseria</td>
<td>386</td>
<td>426</td>
</tr>
<tr>
<td>pseudomonas</td>
<td>332</td>
<td>332</td>
</tr>
<tr>
<td>aeromonas</td>
<td>367</td>
<td>365</td>
</tr>
<tr>
<td>legionella</td>
<td>280</td>
<td>321</td>
</tr>
<tr>
<td>ralstonia</td>
<td>183</td>
<td>342</td>
</tr>
<tr>
<td>myxella</td>
<td>170</td>
<td>247</td>
</tr>
<tr>
<td>Consensus_s:</td>
<td></td>
<td>hhhhh</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Conservation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermus</td>
<td>406</td>
<td>475</td>
</tr>
<tr>
<td>neisseria</td>
<td>387</td>
<td>462</td>
</tr>
<tr>
<td>pseudomonas</td>
<td>366</td>
<td>421</td>
</tr>
<tr>
<td>aeromonas</td>
<td>375</td>
<td>307</td>
</tr>
<tr>
<td>legionella</td>
<td>249</td>
<td>342</td>
</tr>
<tr>
<td>ralstonia</td>
<td>238</td>
<td>303</td>
</tr>
<tr>
<td>Consensus_s:</td>
<td></td>
<td>hhhhh</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates</th>
<th>Conservation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>thermus</td>
<td>476</td>
<td>530</td>
</tr>
<tr>
<td>neisseria</td>
<td>463</td>
<td>520</td>
</tr>
<tr>
<td>pseudomonas</td>
<td>432</td>
<td>462</td>
</tr>
<tr>
<td>aeromonas</td>
<td>388</td>
<td>438</td>
</tr>
<tr>
<td>legionella</td>
<td>314</td>
<td>448</td>
</tr>
<tr>
<td>myxella</td>
<td>304</td>
<td>379</td>
</tr>
<tr>
<td>Consensus_s:</td>
<td></td>
<td>hhhhh</td>
</tr>
</tbody>
</table>
Figure 3.3.1: Alignment and secondary structure prediction of PilQ protein sequences

The PROMALS web server was used to align PilQ sequences from the eight organisms that were used in this study. Sequences are aligned according to both conservation and predicted secondary elements. As an example of some constructs used in this study, for the Neisseria meningitidis PilQ sequence the extent of the B1 (yellow), B2 (green) and α/β or KH-type (blue) regions are highlighted. The predicted transmembrane C-domain is also highlighted in grey. For all sequences, predicted alpha-helical regions are depicted in red lettering and predicted beta-strand elements are given in blue lettering, and indicated below the sequence as (h) and (e) respectively. The degree of conservation of residues is given as a numerical value above the sequences, on a scale of 1-9.
Chapter 3

3.4 High-throughput cloning and protein overproduction

The majority of PilQ constructs were cloned and tested for protein overproduction using facilities at the Oxford Protein Production Facility (OPPF) using a specialized high-throughput method; these are outlined in Table 3.4 on page 110. A small number of Neisseria and Pseudomonas targets were cloned and tested at the home laboratory by other methods (2.3), and these are detailed in Table 3.5 on page 112, and will be discussed later. The high-throughput method implemented at OPPF provided an efficient way to greatly expand the range of chosen targets examined in a limited time. Using this, it was possible to explore a wider range of homologues from various organisms, alternative domain boundaries for chosen targets and soluble protein levels under alternative expression conditions.

The methods used are described in detail in section 2.4. In summary, PCR reactions for all constructs were set up in 96-well plates and analysed by PAGE. A second round of PCR was then conducted for problematic targets with altered parameters (not shown). Ligation reactions were set up simultaneously in 96-well plates and verified by PCR. All PCR products were ligated into the pOPINF vector, to engineer proteins with an N-terminal histidine purification tag. This resulted in positive clones for the full 47 target sequences. All 47 constructs plus a control expressing the GFP gene were transformed into both E. coli B834 and E. coli Rosetta pLysS for small-scale protein production trials. For IPTG induction, single colonies were grown in 0.5 ml starter cultures in 24-well deep-well blocks overnight. These were used to inoculate 2.5ml cultures which were grown to OD<sub>600</sub> of ~0.5 before induction. These were then induced with 0.5 mM IPTG overnight in an orbital shaker at 20°C. All constructs were also tested for protein production in both strains using Overnight Express™ Instant TB (TB-ONEX) auto-induction medium (Novagen), in the same manner but without addition of IPTG. These were induced for 20-24 hours at 25°C. Cells were harvested by centrifugation within the block plates, and pellets were broken by freeze-thawing and resuspension in lysis buffer containing lysozyme and DNase I. The soluble fractions were transferred to microtitre plates and the His-tagged protein of interest was purified by IMAC using a QIAGEN BioRobot 8000 instrument.
Soluble, His-tagged proteins were analysed by SDS-PAGE, and those from IPTG induction are shown in figure 3.4.1 and those from the auto-expression system are shown in figure 3.4.2. Gel lanes are divided into triplet by gridlines for clarity when reading.

- **Soluble protein overproduction with IPTG induction**

The results of the protein production trials are also summarized in Table 3.4. All constructs listed are classified according to soluble protein levels by colour: green indicates a good level of protein production, amber indicates positive production, although this may be at low levels or show multiple bands on the gel, and red indicates no production. The corresponding gel number and lane is also listed in Table 3.4. The first point to note is that constructs expressing regions of PilQ from *Xyella fastidiosa* did not generally produce soluble protein under IPTG induction in either the B834 (figure 3.4.1 gel 1, lanes 1-6) or Rosetta (figure 3.4.1, gel 3, lanes 1-6) cell lines. Constructs expressing recombinant proteins from other organisms produced variable results. At a glance, Table 3.4 indicates that there were slightly greater success rates for protein production in the Rosetta cell line, when compared with B834.

In the B834 cell line (figure 3.4.1, gels 1 and 2) the constructs that expressed and produced good levels of protein generally encoded single or double domain species. For example, B1 domains from *Legionella pneumophilia* (gel 1, lane 7), *Xanthomonas campestris* (gel 1, lane 13), *Aeromonas hydrophilia* (gel 1, lane 20) and *Neisseria meningitidis* (gel 2, lane 14) produced strong bands on the gels, along with the KH-type domain from *Ralstonia solanacearum* (gel 2, lane 3). Two constructs from *Thermus thermophilus*, expressing amino acid residues 310-547 (gel 2, lane 22) and residues 310-480 (gel 2, lane 23) also expressed well. Referring to the PROMALS alignment (figure 3.3.1) these constructs would encode “KH-type” domains, although in *Thermus* PilQ the domain boundaries are less clear. The first construct is predicted to contain a βαββαβαβαβαβα arrangement of secondary structure elements, so this could be a double domain, followed by a 20-residue linker, which precedes a second shorter domain of ββαββ elements. The second of these constructs omits the second shorter domain. More N-terminal regions of the *Thermus* PilQ protein (gel 2, lanes 19 and 20) did not
express as strongly, or did not yield soluble protein samples. This is unfortunate because the N-terminal regions of this protein differ from other PilQ secretins due to the predicted lack of beta domains, so structural studies on these parts of the protein may have raised some interesting points. Another noteworthy result is from a construct originally designed to produce a fusion of the B2 and KH-type domains from *Pseudomonas aeruginosa* (*gel 2, lane 8*). This construct produced a recombinant product, but the corresponding band appeared at a mass below 20 kDa, as opposed to the predicted molecular mass of 38 kDa for the full-length construct. A smaller band can be seen closer to the 36 kDa molecular weight marker, which may originate from the intact protein. The much more intense band at a lower apparent molecular weight may represent a stable degradation product of the protein, perhaps indicating that the tandem domain was sensitive to proteolysis.

Gels 3 and 4 in *figure 3.4.1* show the same constructs as gels 1 and 2, but as tested in the Rosetta cell line, which is used to enhance expression of genes containing rare codons. The results were similar to those from B834 cells, but there were a few differences worth noting. A few more constructs encoding PilQ regions from *Legionella* were expressed in Rosetta cells, namely two extended versions of B1 with alternative domain boundaries (*gel 3, lanes 8 and 11*). These produced multiple protein bands on the gel, so may have been susceptible to proteolysis. Another construct encoding the full N-terminal region also produced some purified products, but they also showed multiple bands on the gel (*gel 3, lane 10*). Two constructs from *Xanthomonas* encoding multiple N-terminal domains produced distinct bands at ~41 and ~32 kDa and did not appear to suffer degradation (*gel 3, lanes 18 and 19*). In addition, a double beta-domain (B1/2) from *Pseudomonas* (*gel 4, lane 10*), a combination B2 and KH-type domain from *Neisseria* (*gel 4, lane 11*) and a combination B2 and KH-type domain from *Xanthomonas* (*gel 3, lane 14*) produced good protein levels in the Rosetta cell line where they did not in the B834 cells. It therefore seems that some of the more complicated constructs containing multiple domains succeeded in the Rosetta but not in the B834 cell line. As the samples were analysed by SDS-PAGE after IMAC purification and were not whole cell extracts, it may be the case that these proteins formed inclusion bodies or were subject to proteolysis in the B834 cells and not in
Rosetta, rather than not being expressed at all. The two *Thermus* constructs encoding KH-type domains which were also successful in B834, expressed at even higher levels in the Rosetta cells producing much stronger protein bands (*gel 4, lanes 23 and 24*).

*Figure 3.4.1: Gel 1 and Gel 2: Soluble protein production using IPTG in E. coli B384 cells*

The gels above show soluble protein purified from IPTG-induced expression of the 47 constructs listed in Table 3.4, plus a GFP control (gel 2, lane 24). For clarity, lanes are divided into triplicate with gridlines. “M” represents molecular weight markers in kDa, which are indicated in this lane.
Figure 3.4.1: Gel 3 and Gel 4: Soluble protein production using IPTG in E. coli Rosetta cells

The gels above show soluble protein purified from IPTG-induced expression of the 47 constructs listed in Table 3.4, plus a GFP control (gel 4, lane 24). For clarity, lanes are divided into triplicate with gridlines. “M” represents molecular weight markers in kDa, which are indicated in this lane.
Soluble protein production with auto-induction medium

Figure 3.4.2 shows the results of construct expression using auto-induction medium. Most of the constructs tested at OPPF did not produce protein in either the B834 (gels 1 and 2) or Rosetta (gels 3 and 4) cell lines and it becomes clear that IPTG induction of these domains was the more successful option. There were, however, a couple of notable results. Firstly, a KH-domain from *Ralstonia* predicted to be ~17 kDa was overproduced in the B834 cell line (gel 2, lane 3) and in the Rosetta cell line (gel 4, lane 3). Secondly, a *Pseudomonas* double beta-domain was also overproduced in B834 cells (gel 2, lane 10) using auto-induction medium, whereas this was not produced using IPTG. Also, in the B834 cell line, a construct from *Thermus* encoding residues 310-547 (gel 2, lane 22) was successful, and this also expressed well using IPTG (figure 3.4.1, gels 2 and 4, lane 22). However in the Rosetta cell line, using auto-induction medium, this same protein runs at a much lower apparent molecular weight on the gel (figure 3.4.2, gel 4, lane 22) so may have been subject to proteolysis in these cells. This is unusual because when expressed in the same Rosetta cell line using IPTG, this was not the case (figure 3, gel 4, lane 22). Using the Rosetta cell line with auto-induction medium (figure 4, gels 3 and 4) produced several other unusual results like this from the constructs. Many run on the gels at a lower molecular weight than predicted, or seen for IPTG induction. It should also be noted that the GFP control (gels 2 and 4, lane 24), was weak or unsuccessful in auto-induction medium. Aside from the *Pseudomonas* construct noted earlier, there were no successes from the auto-induction methods that were not achieved using IPTG, therefore IPTG induction was used for constructs chosen for later, large-scale protein production. IPTG should also be used as the expression strategy for any future projects using these constructs, as recommended by these results.
Figure 3.4.2: Gel 1 and Gel 2: Soluble protein production using auto-induction medium in E. coli B384 cells

The gels above show soluble protein produced from the 47 constructs listed in Table 3.4 purified from E. coli B834 cells grown in auto-induction medium, plus a GFP control (gel 2, lane 24). For clarity, lanes are divided into triplicate with gridlines. “M” represents molecular weight markers in kDa, which are indicated in this lane.
Figure 3.4.2: Gel 3 and Gel 4: Soluble protein production using auto-induction medium in E. coli Rosetta cells

The gels above show soluble protein produced from the 47 constructs listed in Table 3.4 purified from E. coli Rosetta cells grown in auto-induction medium, plus a GFP control (gel 4, lane 24). The control was unsuccessful in this case. For clarity, lanes are divided into triplicate with gridlines. “M” represents molecular weight markers in kDa, which are indicated in this lane.
Table 3.4: High-throughput small-scale production of PilQ constructs

<table>
<thead>
<tr>
<th>Organism</th>
<th>Construct name</th>
<th>Predicted domain</th>
<th>Approx MW (kDa)</th>
<th>IPTG Induction*</th>
<th>Auto Induction*</th>
</tr>
</thead>
<tbody>
<tr>
<td>X. fastidiosa</td>
<td>XFPIQ216-219</td>
<td>B2</td>
<td>15798</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. fastidiosa</td>
<td>XFPIQ219-268</td>
<td>B2 plus B1</td>
<td>22278</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. fastidiosa</td>
<td>XFPIQ239-390</td>
<td>KH</td>
<td>25698</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. fastidiosa</td>
<td>XFPIQ240-250</td>
<td>KH</td>
<td>17728</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. fastidiosa</td>
<td>XFPIQ244-349</td>
<td>B2 plus KH</td>
<td>40108</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ23-301</td>
<td>B1</td>
<td>18238</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ23-305</td>
<td>B1</td>
<td>23378</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ23-309</td>
<td>KH</td>
<td>31859</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ23-340</td>
<td>Full NTD</td>
<td>52428</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ23-109</td>
<td>B1 extended</td>
<td>22938</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>L. pneumophila</td>
<td>LpPIQ22-138</td>
<td>KH extended</td>
<td>32846</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ49-149</td>
<td>B1 domain</td>
<td>16818</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ15-505</td>
<td>B2 plus KH</td>
<td>25678</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ20-106</td>
<td>KH</td>
<td>19978</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ31-197</td>
<td>KH</td>
<td>10738</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ35-501</td>
<td>Full NTD</td>
<td>41428</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ50-206</td>
<td>Full NTD</td>
<td>82408</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>X. campestris</td>
<td>XcPIQ115-191</td>
<td>Full NTD</td>
<td>34458</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>A. hydrophila</td>
<td>AhPIQ25-120</td>
<td>B1</td>
<td>15688</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>A. hydrophila</td>
<td>AhPIQ10-202</td>
<td>B2</td>
<td>18238</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>A. hydrophila</td>
<td>AhPIQ50-507</td>
<td>KH</td>
<td>25478</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>A. hydrophila</td>
<td>AhPIQ10-207</td>
<td>B2 plus KH</td>
<td>38128</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>R. solanacearum</td>
<td>RsPIQ15-112</td>
<td>B1</td>
<td>16018</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>R. solanacearum</td>
<td>RsPIQ15-178</td>
<td>B2</td>
<td>15798</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>R. solanacearum</td>
<td>RsPIQ25-202</td>
<td>B1 plus B2</td>
<td>30098</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>R. solanacearum</td>
<td>RsPIQ29-210</td>
<td>KH</td>
<td>12008</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPIQ22-27</td>
<td>B1</td>
<td>15798</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPIQ20-471</td>
<td>B2 plus KH</td>
<td>39118</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPIQ25-475</td>
<td>Full NTD</td>
<td>54008</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPIQ22-238</td>
<td>B1 plus B2</td>
<td>31068</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>M. marinigatis</td>
<td>MmPIQ112-416</td>
<td>B2 plus KH</td>
<td>39228</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>M. marinigatis</td>
<td>MmPIQ115-549</td>
<td>Full NTD</td>
<td>61558</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>M. marinigatis</td>
<td>MmPIQ115-540</td>
<td>B1</td>
<td>10558</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>M. marinigatis</td>
<td>MmPIQ121-132</td>
<td>KH</td>
<td>15798</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ12-140</td>
<td>KH domain(s)</td>
<td>53628</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ20-42</td>
<td>KH domain(s)</td>
<td>47288</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ10-10</td>
<td>KH domain(s)</td>
<td>63538</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ5-175</td>
<td>KH domain(s)</td>
<td>57498</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ9-100</td>
<td>KH domain(s)</td>
<td>56888</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ7-100</td>
<td>KH domain(s)</td>
<td>38018</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ7-220</td>
<td>KH domain(s)</td>
<td>18998</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ10-247</td>
<td>KH domain(s)</td>
<td>30318</td>
<td>Red</td>
<td>Red</td>
</tr>
<tr>
<td>T. thermophilus</td>
<td>TpPIQ10-480</td>
<td>KH domain(s)</td>
<td>22548</td>
<td>Red</td>
<td>Red</td>
</tr>
</tbody>
</table>

*Overproduction levels of soluble protein are indicated by colour scheme: red for no protein, amber for positive production in low levels or with apparent proteolysis and green for good levels of protein.
3.5 Other PilQ constructs from *Neisseria* and *Pseudomonas*

A smaller range of constructs was also produced at in the home laboratory, using the cloning method outlined in section 2.3. These included KH-type and beta-domains from *Neisseria*, along with beta-domains from *Pseudomonas*, and are detailed in Table 3.5. These constructs were tested for protein production levels in *E. coli* T7 Express cultures. All constructs were engineered to encode proteins with an N-terminal histidine purification tag, and levels of protein were analysed by western blot (*figure 3.5.1*). The *Neisseria* construct encoding the KH-type region (NmPilQ<sup>343-545</sup>) produced high levels of protein with either 0.1 mM or 1 mM IPTG, therefore 0.1 mM was deemed sufficient for large-scale overproduction. The B2-domains from both organisms were also successful (NmPilQ<sup>224-329</sup> and PaPilQ<sup>159-267</sup>), and expression levels were improved with higher levels of IPTG (1 mM) after 3 hours. There was no protein band detected for the *Neisseria* B2-domain after 1 hour induction with 0.1 mM IPTG, which seems unusual as there are bands in every other lane. This may be an anomaly or an error in the western blot. The B1-domains from both organisms (NmPilQ<sup>24-131</sup> and PaPilQ<sup>24-128</sup>) interestingly produced protein initially after 1 hour of induction, yet in both cases the bands became reduced after 3 hours. It is possible that the protein is unstable or susceptible to degradation during expression. In later experiments (3.9), it became evident that the B1-domains appear less stable than the B2-domains. Tandem B1/B2-domains, however, produce good levels of protein after 3 hours (NmPilQ<sup>24-329</sup> and PaPilQ<sup>24-267</sup>). The *Neisseria* construct appears to produce more protein than the *Pseudomonas* counterpart, and construct expression is improved with 1 mM IPTG after 3 hours of induction.
Figure 3.5.1: Overproduction of PilQ domains from Neisseria and Pseudomonas
Western blot analysis of PilQ constructs listed in Table 3.5. Cells were transformed with the appropriate construct, and whole cell samples were analysed by western blot either without addition of IPTG (U), or induced with 0.1 mM or 1 mM IPTG for 1 hour and 3 hours at 37°C. Molecular weight markers (M) are given in kDa.

Table 3.5: PilQ constructs from Neisseria and Pseudomonas

<table>
<thead>
<tr>
<th>Organism</th>
<th>Construct name</th>
<th>Predicted domain</th>
<th>Approx MW [Da]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. meningitidis</td>
<td>NmPilQ^{224-329}</td>
<td>Double KH domain</td>
<td>26468</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>NmPilQ^{24-121}</td>
<td>Beta-1</td>
<td>16018</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>NmPilQ^{224-329}</td>
<td>Beta-2</td>
<td>15978</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>NmPilQ^{24-329}</td>
<td>Beta-1 plus Beta-2</td>
<td>37798</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPilQ^{24-128}</td>
<td>Beta-1</td>
<td>15688</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPilQ^{283-267}</td>
<td>Beta-2</td>
<td>16128</td>
</tr>
<tr>
<td>P. aeruginosa</td>
<td>PaPilQ^{24-267}</td>
<td>Beta-1 plus Beta-2</td>
<td>30978</td>
</tr>
</tbody>
</table>
3.6 *Neisseria meningitidis* KH-type domain production and purification

A construct expressing the predicted KH-type region from *Neisseria*, NmPilQ^{343-545} was used for large-scale protein production and purification. This particular fragment is highlighted in blue on the alignment in *figure 3.3.1*. It includes a predicted loop region between residues 462 and 478, so based on the bioinformatic analysis could therefore be classified as a double KH-type domain, or an N0-domain and N1-domain. Small-scale trials showed that the full length protein was produced in T7 Express *E. coli* after induction with 0.1 mM IPTG for 3 hours at 37°C (*figure 3.5.1*). For isotopically labelled preparations however, 1 mM IPTG was used to maximise gene expression, which is often slower in minimal media. A 2 L cell culture gave a typical yield of ~5 mg of protein. Samples from the purification process were analysed by SDS-PAGE and are shown in *figure 3.6.1a*. Upon cell disruption by sonication, the protein was present in the soluble fraction of the cell lysate and could be purified by IMAC with Ni^{2+}-NTA resin. A low concentration of imidazole (20 mM) was used to wash non-specifically bound contaminants from the Ni^{2+}-NTA resin (*figure 3.6.1a, lanes 2-4*). The protein was then eluted with 200 mM imidazole (*figure 3.6.1a, lanes 6-9*), which was subsequently removed from the eluate by desalting. Typically, the protein would remain soluble for several days at 4°C in 50 mM NaH_{2}PO_{4}/Na_{2}HPO_{4} pH 7.4, 100 mM NaCl. Throughout the project, however, the protein was exchanged using a desalting column into several different buffers depending on the requirements of later experiments, and these are noted where necessary. For example, NMR experiments were conducted at pH 6.8, and sometimes the NaCl concentration was reduced to 50 mM. This did not seem to affect the stability of the sample.

For preparations where removal of the purification tag was necessary, NmPilQ^{343-545} samples were treated with ten units of thrombin per mg of protein for 16 hours at 4°C. Cleaved material was separated from any remaining uncleaved material by means of a reverse IMAC step (2.3.12) and the fractions were analysed by SDS-PAGE. These are presented in *figure 3.6.1b*. Thrombin treatment achieved near complete cleavage ~95% (*figure 3.6.1b, lane 1*). Thrombin was then removed from the sample before gel filtration. Removal of the His-tag did not appear to affect protein stability, as there was no visible insoluble material in the samples.
Figure 3.6.1: IMAC purification and thrombin protease treatment of NmPilQ$^{343-545}$

**a.** NmPilQ$^{343-545}$ was expressed in *E. coli* T7 Express cells and samples from the purification by IMAC are shown. Lane 1: cell lysate flow-through; Lanes 2-4: washing of Ni-NTA resin bound to NmPilQ$^{343-545}$ with 20 mM imidazole; Lanes 5-8: elution of NmPilQ$^{343-545}$ from Ni-NTA resin with 200 mM imidazole. Molecular weight markers (M) are given in kDa.

**b.** Thrombin protease treatment and reverse IMAC purification of NmPilQ$^{343-545}$. After treatment with thrombin, purified NmPilQ$^{343-545}$ was subjected to IMAC purification a second time. Lane 1: flow-through from nickel column, containing cleaved protein; Lanes 2-3: washing of Ni-NTA resin with 20 mM imidazole; Lane 4: elution of uncleaved protein with 200 mM imidazole. The low levels of uncleaved protein that elute indicate near complete cleavage (indicated with a red arrow). Molecular weight markers (M) are given in kDa.
3.7 Analysis of the multimeric state of NmPilQ$^{343-545}$

NmPilQ$^{343-545}$ was further purified by size exclusion chromatography. Figure 3.7.1a shows a typical size-exclusion chromatogram from a sample of NmPilQ$^{343-545}$. A single sharp peak elutes at 12.32 ml; in addition, there is a much smaller, broader peak at 21.86 ml. When the sample was analysed by SDS-PAGE, there was no indication of any lower-molecular weight contaminants or possible degradation products of NmPilQ$^{343-545}$ (not shown). Indeed, in many subsequent preparations of NmPilQ$^{343-545}$ the protein was not susceptible to degradation and remained stable for a considerable length of time in solution. This was the case, for example, during later NMR experiments in which the protein could last for a week or more in the magnet without deterioration of HSQC spectra. It was inferred that the smaller UV peak originates from a low molecular mass contaminant.

Size exclusion can be coupled to multi-angle laser light scattering (SEC-MALLS) for a more accurate estimation of the molecular mass and indication of the homogeneity of the protein in solution. The example in figure 3.7.1b shows the chromatogram produced by SEC-MALLS analysis of a sample from an NmPilQ$^{343-545}$ preparation with the purification tag intact. The material contained within the major peak fraction had a polydispersity of 1.003 and a calculated molecular mass of 29.04 kDa, indicating that the protein is a monomer (predicted molecular mass is 26.1 kDa).
Chapter 3

Figure 3.7.1: Analysis of the multimeric state of NmPilQ$^{343-545}$

a. Size-exclusion chromatography was used as an extra purification step after IMAC of NmPilQ$^{343-545}$. The protein elutes as a single sharp peak at 12.32ml. SEC was performed using a Superdex 75 10/300 GL column using 50mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 7.4, 100mM NaCl as the running buffer. Elution volumes of two molecular weight standards are shown by arrows.

b. Size-exclusion chromatography was coupled to multi-angle laser light scattering for an accurate representation of molecular weight and homogeneity. The running buffer used was PBS. The protein contained within the major peak, which is indicated by dashed lines, had a calculated molecular mass of 29.0kDa, indicating a monomeric form.
3.8 Characterization of NmPilQ\textsuperscript{343-545} by NMR

A one-dimensional \textsuperscript{1}H-NMR spectrum, taken from an unlabelled protein sample, provides a fast, reliable method to obtain valuable information about the biophysical state of recombinant proteins in solution \textit{in vitro}. Factors that can be analysed in this way include the folded or unfolded state of the protein, secondary structural content and relative concentration of samples. This is useful to assess samples before embarking on costly and time-consuming processes such as preparing labelled proteins for NMR, X-ray crystallography and even biochemical experiments which may be futile if using unfolded or inactive proteins.

For NMR screening, NmPilQ\textsuperscript{343-545} was exchanged into 50 mM NaH\textsubscript{2}PO\textsubscript{4}/Na\textsubscript{2}HPO\textsubscript{4} buffer at pH 6.8 with 50 mM NaCl and 10\% D\textsubscript{2}O. The 1D \textsuperscript{1}H-NMR spectrum is shown in figure 3.8.1a. There was a good dispersion of peaks in the region of amide protons at 6-10 ppm and methyl protons at 0.5-1.5 ppm. Therefore, multidimensional experiments were conducted requiring isotopically labelled protein. Cultures producing NmPilQ\textsuperscript{343-545} were grown in minimal media supplemented with \textsuperscript{15}N-ammonium chloride, and cells were induced overnight to achieve similar expression levels and a similar overall yield of protein to its unlabelled counterpart. The \textsuperscript{15}N-HSQC for a 500 \textmu{}M sample is presented in figure 3.8.1b. Crosspeaks are well dispersed across the spectrum indicative again of a folded protein sample. In a structured protein the presence of hydrogen bonding in helices, sheets and other structural features causes the backbone amide protons to have more varied chemical shifts than if, for example, they were all exposed to solvent. The standard random coil region for amide hydrogen is between 7.5 and 8.5 ppm. For helical regions the range is commonly 6.5-9 ppm and for beta sheet the range is usually 6.5-10 ppm. Noteworthy are backbone amide shifts above 10 ppm, which are unusual but do happen when a chelating centre is present. The structure and assignment of this protein is discussed in more detail in chapter 4.

Figure 3.8.1: Characterization of NmPilQ\textsuperscript{343-545} by NMR

a. One-dimensional $^1$H-NMR spectrum recorded for a 500\,$\mu$M sample of NmPilQ\textsuperscript{343-545} in 50 mM NaH\textsubscript{2}PO\textsubscript{4}/Na\textsubscript{2}HPO\textsubscript{4} pH 6.8, 50 mM NaCl, 10\% D\textsubscript{2}O.

b. Two-dimensional $^1$H-$^15$N-HSQC spectrum for a 500 \,$\mu$M sample of isotopically labelled $^{15}$N-NmPilQ\textsuperscript{343-545}.

Spectra were recorded at 298 K on a Bruker Avance III 800MHz spectrometer equipped with TCI triple resonance cryoprobes. Spectra were processed using Topspin2.1 (Bruker).
3.9 PilQ beta-domains

It became evident from the results of the protein production trials that constructs encoding beta-domains, in particular single beta-domains, from PilQ were most effective. Because these recombinant proteins are generally small, in the region of 14-16 kDa, and more likely to be monomeric in solution as the N-terminal regions of PilQ are not required for multimerization, it was decided that they would be most suited to NMR studies. Figure 3.9.1 shows an alignment of B1 and B2-domains from the genomes used in this study according to secondary structure prediction (figure 3.9.1a) and sequence identity (figure 3.9.1b). From figure 3.9.1a it becomes clear that predicted β-strand elements in B1-domains align very well with those in B2-domains, so it seems likely that the two would share a common fold. When the sequences are aligned according to sequence homology in figure 3.9.1b using ALIGN, there is higher degree of sequence identity within separately grouped B1-domains and B2-domains. Although it is very limited, some sequence conservation can be seen across the group of B1- and B2-domains as a whole. With the similarity in predicted secondary structure, it seems that two beta-domains may have arisen from an evolutionary internal gene duplication event, although we cannot be confident of this without structural data. It is important to note here that the first beta-domain within PilQ from Xanthomonas campestris and Xyella fastidiosa in both cases aligns better with the B2-domains of other organisms, and there is therefore no B1-domain apparent.
Chapter 3

Figure 3.9.1: Sequence alignment of PilQ beta domains expressed in this study

a. PROMALS alignment of B1 and B2-domains produced in this study according to secondary structure prediction. The general position of beta strands, which are depicted in blue text, is given schematically below the sequences. Predicted helical regions are given in red text.

b. Alignment of B1-domains with B2-domains according to sequence conservation. Conserved residues are highlighted in blue, with those most conserved shaded darker.

Image was produced using ALIGN. Although residues are most conserved within groups of beta-1 or beta-2 domains, there is still some conservation between the two.
- Characterization of beta-domains by NMR

Initial work on the PilQ beta-domains was carried out to characterise the folded state and relative stability of the recombinant proteins, and determine the best candidates for structure determination. For the purposes of a fast, high-throughput method, individual expression optimization trials were not conducted, as would normally be the case for large-scale protein production. Instead, successful constructs from those produced at both OPPF and the home laboratory were transformed into *E. coli* BL21 (DE3) cells, 0.5 L cultures were grown in 2YT media and induced using 1 mM IPTG for 3 hours at 37°C. For the purposes of characterization by NMR, cultures were grown in minimal media supplemented with $^{15}$N-NH$_4$Cl. The cultures were induced overnight to maximise protein production, which is generally lower from cells grown in this media. Proteins were purified from cell pellets using the usual method (2.3) and exchanged into 50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 6.8, 50 mM NaCl using a PD-10 desalting column (GE Healthcare). Purification tags were not removed.

All NMR spectra were recorded at 298 K using a Bruker Avance III 800MHz spectrometer. Spectra were processed using Topspin2.1 (Bruker) and the resulting $^1$H$^{15}$N HSQCs are presented in figure 3.9.2. The *Neisseria* B1-domain was at a low concentration after purification (panel A) and produced little signal above the noise in the HSQC. In contrast, the B2-domain (panel B) was at a much higher concentration (500 µM) and the spectrum was characteristic of a well-folded domain with a good dispersion of strong crosspeaks on the HSQC. The *Pseudomonas* B1 construct (panel C) produced a reasonable amount of protein (47 µM) but by the time this sample reached the magnet it appeared to be unfolded, as shown by poorly dispersed crosspeaks. The B2-domain was much better, showing a well dispersed spectrum indicating folded protein (panel D), but there was a much lower yield compared to that obtained from the equivalent *Neisseria* construct (40 µM compared with 500 µM). It became apparent that the B1-domains may be less stable or more difficult to purify as soluble entities than the B2-domains. Other B1-domains, from *Aeromonas* (panel E) and *Legionella* (panel F and G), did produce well dispersed spectra but the concentrations of protein in these samples were generally low. Therefore *Neisseria* B2 was the best candidate to attempt structure determination, and this is reported in chapter 4.
Figure 3.9.2: Characterization of PilQ beta-domains by NMR

All panels show $^1\text{H}^{15}\text{N}$-HSQC spectra for beta-domains expressed in this study. The concentration of protein is given on each panel along with the estimated molecular weight. All spectra were recorded at 298k on a Bruker Avance III 800MHz spectrometer equipped with TCI triple resonance cryoprobes. Spectra were processed using Topspin2.1 (Bruker).
- Characterization of a B1/2-domain (NmPilQ<sup>24-329</sup>)

The construct encoding the tandem beta-domain from *Neisseria* was also expressed very well so this was used for large-scale protein production. Isolating a single, well folded B1-domain had proven difficult; it was thought that the presence of a folded B2-domain might add some stability to the complex. The construct gene was expressed in *E. coli* T7 Express cells with 1 mM IPTG for 3 hours at 37°C and the protein was purified by the usual IMAC methods. For consistency with the remainder of the project, 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> was used at pH 7.8 with 100 mM NaCl. Fractions from the purification process are presented in *figure 3.9.3a*. Despite washing the column resin with 20mM imidazole (lanes 2-5), the protein eluted with several contaminants or degradation products (lanes 6-8). It was therefore purified further by gel filtration; the chromatogram is presented in *figure 3.9.3b*. When fractions eluted from the size exclusion column were analysed by SDS-PAGE, a prominent double banding pattern emerged with some material corresponding to the correct approximate size of the tandem domain at ~37 kDa, and about half of the material consisting of a ~25 kDa species, suggesting proteolytic cleavage. This may explain the unusual shape of the chromatogram, which forms a broad peak with a shoulder as the elution volume increases. It is tempting to speculate that as the B2-domain alone is relatively stable that proteolysis is due to the presence of the B1-domain, perhaps due to an intrinsically disordered domain as earlier results suggested or the linker region connecting the two. Either way, the tandem B1/2-domain does not seem suitable for structural studies, therefore it was not pursued further in this case.
**Figure 3.9.3: IMAC purification and SEC of NmPiiQ**

**a.** IMAC purification of NmPiiQ\textsuperscript{24-329}. Lane 1: cell lysate flowthrough; Lanes 2-5: wash steps with 20mM imidazole; Lanes 6-8: elution steps with 200mM imidazole. Molecular weight markers (M) are given in kDa.

**b.** SEC chromatogram for NmPiiQ\textsuperscript{24-329} and analysis of the elution fractions by SDS-PAGE. Molecular weight markers in kDa are shown on the chromatogram by arrows, and on the gel in lane ‘M.’ The corresponding SEC elution volume (ml) to the fractions analysed by SDS-PAGE is indicated in alternate gel lanes.
3.10 Discussion

A number of approaches have been taken in the field to try to dissect and deduce structural information for members of the secretin family. This project exploited the 'beads-on-a-string,' multi-domain organization of secretins; bioinformatic analysis of protein sequences was then used to make educated assumptions on protein domain boundaries. This, combined with a high-throughput screening and characterization process, enabled the production of well-folded, highly concentrated, soluble samples of a few of these targets with minimal impact on cost and resources.

It seems likely that unstable, partially folded or disordered elements of recombinant proteins would be the cause of insolubility and aggregation, particularly in the case of recombinant forms of proteins such as secretins which in their native forms have the propensity to oligomerize. However it should be noted that these recombinant forms are being produced in a different way, and not exported to the periplasm. Even so, by omitting what we now know to be the oligomerization domain (the C-domain) from the recombinant fragments, we still cannot be sure that there are no other regions that would not be exposed to solvent when in multimeric form, or in complex with a chaperone or pilotin, that become exposed in recombinant form. In the case of the secretin PulD from the T2SS, although the C-domain alone can retain its multimeric form after proteolysis to remove the N-terminal regions,\textsuperscript{33} when it is synthesised alone it fails to form SDS-resistant multimers even though it is successfully exported and forms a complex with its pilotin chaperone.\textsuperscript{30} Also, it has been proposed that part of the N-terminal region may stabilize or cause partial multimerization of the C-domain.\textsuperscript{30} Therefore, the N-terminal, periplasmic domains of secretins may indeed play a limited role in multimerization, although this has not been well characterized or mapped to any part of this region in particular. If this is the case, certain parts of protein sequence may be inherently prone to aggregation, which makes the multiple target selection approach used in this study all the more useful to avoid these areas in at least some of the constructs.

The constructs encoding alpha/ beta, or KH-type domains, produced in this study were expressed very well producing soluble proteins in several instances.
Examples can be seen in Table 3.4, notably AhPilQ$^{314-507}$, RsPilQ$^{292-408}$, and one construct produced in the home laboratory (Table 3.5; NmPilQ$^{343-545}$). These all consisted of predicted KH-type domains only, and not combination or tandem domains, such as a KH and B2. This perhaps highlights the insolubility of the (often long) linker regions between domains when produced as recombinant proteins. It should also be noted that some KH-type domains did not express well alone, or were expressed but proteins were not found in the soluble cell fraction. Two PilQ constructs from *X. fastidiosa* are such examples: XfPilQ$^{195-390}$, XfPilQ$^{195-318}$, and these are alternative versions of the same domain. This may simply be an example of different protein homologues behaving differently, or it may be an example of error when selecting domain boundaries from the bioinformatic predictions, which would demonstrate the importance of choosing a wide range of targets.

One target which was advanced to large-scale protein production was NmPilQ$^{343-545}$. This construct produced a high yield of soluble protein which remained relatively stable at high concentrations, sometimes being concentrated up to 1mM. There have been various attempts to overproduce recombinant forms of similar domains from other secretins using a range of approaches. Korotkov *et al* used a directed evolution approach with a C-terminal fusion with GFP as a reporter for correct folding. They found that the first 40 residues of their construct needed to be removed to improve solubility, and following this created shorter constructs from a range of organisms, which produced a sufficient yield of protein for crystal trials. However, we should note that only a few of these led to crystals, which then diffracted poorly until assistive nanobodies were used.$^{34}$ Indeed the construct used in this study also failed to crystallize upon a few initial attempts, suggesting an inherent flexibility of the recombinant protein which is a predicted two domains joined by a linker. This is elaborated upon in chapter 4, where a complete NMR chemical shift assignment is made for NmPilQ$^{343-545}$. This information was used to create a shorter construct containing less disordered regions, and the structure is presented in section 4.4.

Spreter *et al* published the crystal structure of the two N-terminal domains of EscC.$^{35}$ This group used a similar approach to obtain their protein as in this study; by subcloning the region residues 21-174 to omit the signal sequence and the C-
domain, they were able to overproduce in the *E. coli* cytoplasm and purify from cell lysates. The protein overproduced well and was soluble at concentrations of up to 18 mg ml\(^{-1}\)\textsuperscript{35}. This recombinant protein crystallized with no reported issues, which is in contrast to attempts with GspD\textsuperscript{34}. Similarly, the two recombinant N-terminal domains from HofQ were produced in the cytoplasm and produced a high yield of protein, which was soluble up to 17 mg ml\(^{-1}\) for crystal trials, and also crystallized successfully\textsuperscript{182}.

There are also examples of recombinant production of domains from other bacterial outer membrane proteins. A good example is the PapC usher protein of the chaperone/ usher pathway (1.2). The domain organization of PapC is such that the 24-stranded translocation pore folds out from the centre of the polypeptide to form a soluble plug domain, which occludes the lumen of the barrel. The N- and C-terminal domains of the polypeptide also form soluble domains, which sit in the periplasm. Ford *et al* overproduced three constructs in the periplasm to obtain the C-terminal domain in soluble, recombinant form. All of these were rapidly degraded to yield a stable smaller fragment. By this approach they achieved sufficient material, soluble at 15 mg ml\(^{-1}\), for structural studies by X-ray crystallography, NMR and subsequent structure determination\textsuperscript{197}. Also, the periplasmic POTRA domains of the *E. coli* OMP beta-barrel, YaeT, have been produced separately. Knowles *et al* produced constructs of domains 1-2 and 1-5, which purified to homogeneity with high yields of 11 and 7.5 mg ml\(^{-1}\)\textsuperscript{198}. Therefore, the work with secretins is typical of more general approaches adopted with other OM proteins which are often ‘dissected’ in these ways to obtain functional domains which are in high enough quantities for structural and biochemical studies by conventional approaches.

An important point to discuss with respect to KH-type domains from secretins is the oligomeric state of the recombinant forms of the proteins. NmPilQ\textsuperscript{343-545} produced in this study, was monomeric in solution when analysed by size-exclusion and multi-angle laser light scattering. This was also the case for the periplasmic domains of GspD\textsuperscript{34} and EscC,\textsuperscript{35} which invariably form monomers. This was consistent with the idea that it is the C-terminal domain, which drives the oligomerization of secretins. In the case of NmPilQ, this also extends to the beta-
domains, which appear to be monomeric. None form multimers, at least in recombinant form.

In contrast to this, however, the KH-type domains produced from HofQ formed a dimer in the crystal. The authors also believed this dimer to be the favourable form in solution upon both the analysis of the protein by size exclusion, and also from the PISA\textsuperscript{199} analysis of the crystallographic dimer. They do, however, note that elongated molecules can elute from size exclusion columns at an apparently higher molecular weight than their true size, so they cannot definitively conclude this from size exclusion. In addition, it may be possible that the dimer is a result of the construct being used which omits the transmembrane C-domain. If present, the C-domain would not be able to form an oligomeric ring structure when presented in the orientation seen in the dimer\textsuperscript{182}.

Oligomerization seems to be intrinsic to secretin function; it is a requirement in order to form a pore in the outer membrane for the passage of substrates. This may explain why there is a plethora of different proteins suggested to assist in secretin multimerization. It also suggests that the way secretin oligomers form is different from other OMPs, such as TolC. NmPilQ is very heat-stable, much more so than other OMPs\textsuperscript{12}. The monomeric form of recombinant N-terminal domains adds weight to the hypothesis that the C-domain drives oligomerization, yet it must maintain a degree of flexibility to allow passage of the pilus fibre. If the challenges can be overcome regarding the insolubility of this domain, it would provide an attractive area of investigation.

We must allow for the possibility that B1-domains are intrinsically only partially folded, although this was not definitively concluded throughout the course of this experimental work. Preliminary evidence, such as the NMR spectra presented here, and the difficulties experienced in obtaining sufficient yields of protein from the B1 constructs, are consistent with this hypothesis. The beta-domains appear to be unique to T4PB secretins, suggesting there is a specialized function with respect to type IV pilus assembly. It remains unclear what this specialized function may be. It is intriguing why T4PB secretins, with only one known substrate, would have different requirements from T2SS secretins, for example. The latter have
more diverse substrates and also assemble pseudopili. This may be related to a unique T4PB characteristic, such as pilus disassembly.
Chapter 4: Structural studies of domains from the type IV pilus-dependent secretin PilQ

4.1 Introduction

The previous chapter discussed in detail the domain organization of members of the secretin family. Bioinformatic analyses of secretin sequences predict a sequential organization of different domains as separate entities connected by unstructured loop regions of polypeptide. The results in chapter 3 showed that, at least in the case of PilQ from the T4PB system, these tools can be put to use in obtaining high quantities of folded, recombinant secretin domains, which are suitable candidates for structure determination by X-ray crystallography or NMR.

Secretins are multimeric, integral membrane proteins and therefore like many other membrane proteins will be notoriously difficult to overproduce and purify in large enough soluble quantities. Membrane proteins, due to their hydrophobic surfaces, must be extracted from the lipid bilayer with detergents which cover the hydrophobic surface allowing for solubilisation. The choice of detergent then also becomes an issue; with cost, protein stability, sample homogeneity and protein activity all being contributing factors. This is before the usual challenges faced in crystallographic studies with all recombinant proteins, in finding the right crystallization conditions and in structure determination.

The structures that are available for secretin multimers have therefore largely been elucidated by electron microscopy methods. Purified homo-multimers have been obtained by isolation from native sources, such as PilQ from the T4PB system in *Neisseria* and YscC from the T3SS in *Yersinia;* or alternatively by transformation and overproduction in *E. coli* outer membranes, for example the T2SS secretins PulD from *Klebsiella* and GspD from *Vibrio.* Interestingly many of these resulting structures reveal differences in the architecture of the multimers. All reveal large ring structures which form chambers in the periplasm. PilQ from *Neisseria meningitidis* forms a dodecameric structure which is effectively sealed at both the top and bottom ends. The pilus fibre is 65 Å in diameter, suggesting that the secretin structure must be flexible or gated in some manner to allow its passage. In addition, the 12 subunits form a multimer with overall 4-fold symmetry and quasi 12-fold symmetry, suggesting inherent flexibility
within the complex. In contrast to this, the structure of the intact T3SS needle complex from *Salmonella* showed the secretin component, InvG, to be in an open state allowing passage of the needle. The structure of the T2SS GspD is sealed at the outer membrane and open at the periplasm. There is a constriction of the central chamber formed by the N0-domain which is not seen in the PilQ reconstructions. The current model indicates that this domain interacts with the secreted cholera toxin substrate, and this may gate the pore allowing it to open. How any of these structures open at the C-terminal membrane-spanning region to allow substrates out of the bacterial cell, however, remains to be seen.

It is clear that atomic resolution structures of individual secretin domains from the T4PB system, such as those already obtained for the T2SS and T3SS secretins, would be informative and useful in determining their individual functions in mediating extrusion of the pilus fibre. As described in section 3.2, secretins from the T4PB system have a distinct feature due to the presence of one or two predicted beta-strand rich domains at the extreme N-terminus, which is unique to the T4PB system. These domains would reach into the periplasm, possibly forming links with the rest of the pilus machinery, and an understanding of their structure and function may therefore provide clues to the gating of the multimer or function of other secretin interacting partners, which are poorly understood. As discussed previously, this has already been shown to be the case for the T2SS secretin GspD and T3SS EscC.

This chapter reports how a combinatorial approach of methods based around nuclear magnetic resonance can be implemented to obtain high-resolution structures of recombinant PilQ secretin domains. This work follows on from the bioninformatics and high-throughput cloning and expression strategy implemented in chapter 3 to advance the best candidates for structure determination. Both NMR de novo structure determination, and a combination of calculated chemical shifts, CS-ROSETTA and homology modelling are used to produce high resolution structures of the second beta-domain and the N0-domain respectively, both from PilQ of *Neisseria meningitidis*. Also reported is a full chemical shift assignment for the first beta-domain from *Aeromonas hydrophilia*. 
The first attempts at structure determination were made with the NmPilQ\textsuperscript{343-545} construct, and a full chemical shift and secondary structure assignment was obtained. This told us that the construct consisted of an N0 and N1-domain, however unstructured regions within the protein hampered the collection of high-quality NOEs. Therefore a shorter construct, NmPilQ\textsuperscript{343-442}, was produced and purified and this led to a solution structure. Homology modelling and experimentally determined chemical shifts were then used to produce a model for the full N0/N1 region of NmPilQ. In addition, the structure for the B2-domain, NmPilQ\textsuperscript{224-329}, was solved by NMR. Figure 4.1.1 outlines the positions of these recombinant proteins within the full-length PilQ sequence, and compares this to the positions of domains in T2SS and T3SS secretins.

**Figure 4.1.1: Schematic illustration of secretin domain organization**

The relative positions of domains within the PilQ peptide are shown, compared with those from secretins of the T2SS and T3SS. The signal sequence (sig) and the pilotin-interacting S-domain (S) are also indicated. The extent of the *Neisseria meningitidis* PilQ constructs used in this chapter is indicated above the PilQ peptide.
The work in this chapter was carried out in collaboration with Professor Lu-Yun Lian at The University of Liverpool, with particular credit going to Dr. Marie Phelan for both advice and help with NMR principles, sample preparation, spectra recording and use of homology modelling software, whilst also carrying out the advanced structure determination and analysis.

4.2 Assignment of NmPilQ^343-545^343-545

As described in the previous chapter (3.6), a predicted two-domain fragment likely to adopt a variant of the alpha/beta type fold found in other secretins such as GspD^34^34 and EscC^35^35 was produced and purified from E. coli. The protein produced from the construct, NmPilQ^343-545^343-545, gave well-dispersed $^1$H$^{15}$N-HSQC spectra (section 3.8, figure 3.8.1). Based on secondary structure prediction and alignment with other secretins, it seems likely that this fragment would contain an N0-domain and a KH-type domain, often referred to as N1-domains, although this was uncertain at this stage without experimental evidence. A double-labelled $^{15}$N$^{13}$C preparation of this protein was purified from E. coli BL21 (DE3) cells cultured in M9 minimal media which was supplemented with $^{15}$N-NH$_4$Cl and $^{13}$C-glucose. The NMR spectra recorded for the purified protein were subsequently assigned using CCPN analysis^155^155 (see section 2.6.4-2.6.5 for experiments conducted). Triple resonance assignment was obtained, and CSI^205^205 was used to determine secondary structure. The full assignment of crosspeaks for the $^1$H$^{15}$N-HSQC spectra is presented across figure 4.2.1a, b and c. Methyl assignment is not shown.

Determination of secondary structure according to CSI^205^205 is presented graphically in figure 4.2.2a. From the assignment of secondary structure at this stage, it became likely that the two-domain species produced here comprised an N0 and N1 (KH-type) domain, as opposed to two KH-domains. The second domain contains three beta-strands and two alpha-helices, like other characterized KH-type domains. In contrast the first domain, which ends at around residue 440, contains five beta-strands and three alpha-helices. If we compare this to the output from a secondary structure prediction programme, which is shown in figure 4.2.2b for the residue range 250-540, the results are very similar. The secondary structure prediction, using the PHYRE2 web server^206^206 and PROMALS,^195^195 was used for the construct design as described in section 3.3. There is, however, an important difference. Between residues 515 and 540, there are two predicted beta-
strands (figure 4.2.2b) and for this reason those residues were included in NmPilQ^{343-545}. After this, a long disordered linker region is predicted, followed by the transmembrane C-domain (not shown in the figure). However, upon experimental assignment of secondary structure, the two strands are not present and the N1-domain ends with a long random coil extension of ~35 amino acids. This resulted in very intense peaks in the NMR spectra, which obscured many of the peaks from the folded N1 region. These peaks can be seen in figure 4.2.1c, as the poorly dispersed signals from residues ~510-545. This precluded the extraction of the high quality NOEs required for structure determination of the double-domain fragment. The presence of random coil in place of two beta-strands may be an artefact of the fragment being produced in recombinant form; for example the presence of the C-domain, or a portion of the linker, may be required for correct folding of this part of the polypeptide. Therefore, the construct was re-designed, to omit the N1-domain and contain only the larger first domain. Protein production and purification from this construct, NmPilQ^{343-442}, is described below.
Figure 4.2.1b
Figure 4.2.1: $^1H^{15}N$-HSQC spectra for NmPilQ$^{343-545}$

Panels A-C above show the $^1H^{15}N$-HSQC spectra for NmPilQ$^{343-545}$ buffer exchanged into 50 mM Na$_2$PO$_4$/Na$_2$HPO$_4$ pH 6.8; 100 mM NaCl; 10% D$_2$O at 298 K. Figures produced using CCPN Analysis.

a: Full $^1H^{15}N$-HSQC spectra

b: Spectra shown for the region between $^1H$: 9.4-6.6 ppm and $^{15}N$: 104-132 ppm

c: Closer view of the centre region, between $^1H$: 9.0-7.4 ppm and $^{15}N$: 115-127 ppm
Figure 4.2.2a

\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]

CSI

\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]
\[ \Delta \delta^{(13C)} \]

CSI
Figure 4.2.2b

**Figure 4.2.2**: Secondary structure content of NmPilQ<sub>343-545</sub>

- **a**: Chemical shift Index (CSI) calculated for deviations from random coil shifts of Ha, Ca and CO to determine the consensus secondary structure; graph adapted from CCPN analysis. Protein samples consisted of natural isotopic abundance or 98% <sup>15</sup>N and 99% <sup>13</sup>C labelled protein in 500 µl of 90% <sup>1</sup>H<sub>2</sub>O 10% <sup>2</sup>H<sub>2</sub>O.

- **b**: Predicted secondary structure for a region of NmPilQ including NmPilQ<sub>343-545</sub> for comparison to the experimentally determined structural content shown in (a). Algorithm and output graphic produced from the PHYRE web server.
4.3 Overproduction and purification of NmPilQ$^{343-442}$

The truncated construct, NmPilQ$^{343-442}$, containing the predicted N0-domain from PilQ was cloned into vector pET11b* to produce a recombinant protein with an N-terminal histidine purification tag. Because of the similarity of this construct to NmPilQ$^{343-545}$, and the need to quickly obtain a sample for NMR experiments, expression trials were not conducted and the protein production and purification procedure followed that of NmPilQ$^{343-545}$ (section 3.6). A trial 0.5 L culture of *E. coli* B834 (DE3) cells producing unlabelled NmPilQ$^{323-442}$ was produced and gave a comparable yield of protein to NmPilQ$^{343-545}$, and remained soluble after removal of imidazole in 50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$, 100 mM NaCl. Therefore a 1 L culture was produced in M9 minimal media supplemented with $^{15}$N-NH$_4$Cl and $^{13}$C-glucose to produce a double-labelled protein sample for triple resonance experiments. The protein was purified in the usual way by IMAC (not shown), and then analysed by size-exclusion chromatography. *Figure 4.3.1a* shows the elution profile of NmPilQ$^{343-442}$ from a size-exclusion column. The majority of material elutes as a single, sharp peak at 13.46 ml. If we refer back to size exclusion of NmPilQ$^{343-545}$, under the same conditions this elutes only slightly earlier, at 12.32 ml. NmPilQ$^{343-545}$ was found to be monomeric (section 3.7), so it is assumed from the elution volume that the truncated protein, NmPilQ$^{343-442}$, is also monomeric in solution and behaving in a similar manner. There was also a higher molecular weight peak, eluting at 8.82 ml, which was quite a large peak with an absorbance of 250 mAU. Therefore eluted fractions were analysed by SDS-PAGE to check whether this was caused by aggregates of NmPilQ$^{343-442}$ or another contaminant, and the gel is shown in *figure 4.3.1b*. The material contained within this peak, from 8.0 ml to 9.6 ml on the chromatogram, was close to 75 kDa (lanes 1-5), and therefore not NmPilQ$^{343-442}$. The smaller shoulder peak, from 9.6 ml to 11.2 ml, contained a species close to 25 kDa (lanes 6-8). The majority peak on the chromatogram eluting between 12.4 ml and 16.0 ml contains a species just below 15 kDa, corresponding with the predicted size of NmPilQ$^{343-442}$. The fractions between 13.2 ml and 16.0 ml containing the purest protein were pooled together and concentrated. The final concentration of the 500 µL sample used for NMR was

*Construct was kindly produced by Sravan Pandalaneni; technical assistant, The University of Liverpool*
0.76 mM, which means that from a 1 L culture a final yield of ~5.76 mg of protein was attained.

Figure 4.3.1: Purification of NmPilQ\textsuperscript{343-442} by size-exclusion chromatography

a: Superdex 75 10/300 SEC column chromatogram showing elution of NmPilQ\textsuperscript{343-442} as a major peak at 13.46 ml. Running buffer used was 50 mM NaH\textsubscript{2}PO\textsubscript{4}/ Na\textsubscript{2}HPO\textsubscript{4}, 100 mM NaCl. Molecular weight standards are shown by arrows.

b: NmPilQ\textsuperscript{343-442} was eluted from the SEC column in 0.4 ml fractions and these were analysed by SDS-PAGE. The corresponding elution volume on the chromatogram in (a) to each fraction is indicated on the gel.
4.4 Structure determination of NmPilQ\textsuperscript{343-442}

Although assignment of chemical shifts for NmPilQ\textsuperscript{343-545} had already been completed, a full assignment of NmPilQ\textsuperscript{343-442} was conducted in the same way using CCPN analysis,\textsuperscript{155} instead of by assignment transfer. This resulted in a confident conclusion that chemical shifts from common residues in NmPilQ\textsuperscript{343-545} and NmPilQ\textsuperscript{343-442} were very similar in both protein samples, and therefore that the overall fold was unaffected by truncating the construct. The assigned $^1$H$^{15}$N-HSQC is presented in figure 4.4.1.

Automated NOESY assignment and preliminary structure calculations were performed using CYANA 2.1 software,\textsuperscript{156,157} with input data of shift lists derived from $^{15}$N- and $^{13}$C-HSQC spectra. A total of 2337 NOESY peaks were assigned. CYANA 2.1 calculations ran with standard protocols using 7 cycles of automated NOE assignment and structural calculations, producing 100 structures per cycle. Water-refinement was conducted using CNS1.2\textsuperscript{158} with a total of 1309 unambiguous interproton distance restraints and 110 dihedral restraints. Dihedral $\varphi$ and $\psi$ torsion angles were produced by TALOS+\textsuperscript{159} and the final ensemble of the best 20 low-energy, water-refined structures, and validated with PROCHECK-NMR\textsuperscript{160} using the iCing interface (http://nmr.cmbi.ru.nl/icing/iCing.html). Structure calculation statistics are presented in Table 4.4.

Figure 4.4.2a shows the solution structure of NmPilQ\textsuperscript{343-442}, which adopts a fold similar to the N0-domain from other secretins such as GspD and EscC, as opposed to the N1-domain, as suspected from the secondary structure assignment. The structure was analysed by comparison with existing structures available in the PDB using the DALI server,\textsuperscript{207} and the closest structural relations are also shown in figure 4.4.2b-d.
Figure 4.4.1: Assigned $^1$H$^{15}$N-HSQC spectra for NmPilQ$^{343-442}$

Protein was buffer exchanged into 50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 6.8; 100 mM NaCl; 10% D$_2$O and spectra were recorded at 298 K. Figures produced using CCPN Analysis.

a: Full $^1$H$^{15}$N-HSQC spectra

b: Close-up view of the region between $^{15}$H: 7.5-8.6 ppm and $^{15}$N: 118-128 ppm
### Table 4.4: NMR Structure calculation statistics for NmPilQ$_{343-442}$

#### NMR constraints

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of distance constraints</td>
<td></td>
</tr>
<tr>
<td>Short range ($</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium range ($1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long range ($</td>
<td>i-j</td>
</tr>
</tbody>
</table>

#### Structure statistics (20 structures)

<table>
<thead>
<tr>
<th>Violation statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of NOE violations &gt; 0.3 Å</td>
<td>0</td>
</tr>
<tr>
<td>NOE violations &gt; 0.3 Å</td>
<td>0</td>
</tr>
<tr>
<td>Maximum NOE violation</td>
<td>0.29</td>
</tr>
</tbody>
</table>

#### Ramachandran statistics

<table>
<thead>
<tr>
<th>Residues class</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in most favoured regions</td>
<td>79.57</td>
</tr>
<tr>
<td>Residues in allowed regions</td>
<td>18.47</td>
</tr>
<tr>
<td>Residues in generously allowed regions</td>
<td>1.64</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td>0.36</td>
</tr>
</tbody>
</table>

#### RMS deviations from the mean structure

<table>
<thead>
<tr>
<th>Component</th>
<th>RMS deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein backbone</td>
<td>0.34</td>
</tr>
<tr>
<td>Protein heavy atoms</td>
<td>1.01</td>
</tr>
</tbody>
</table>
Interestingly, Dali analysis indicates that the fold of NmPilQ\textsubscript{343-442} is in fact more closely related to N0-domain structures from T3SS secretins, as opposed to those from T2SS secretins. This is surprising result because, as described in the main introduction, the T4PB biogenesis system shares many close homologues with the T2SS. Pairwise alignment with EscC from \textit{E. coli}\textsuperscript{55} and InvG from \textit{S. typhimurium}\textsuperscript{52} produced Dali Z-scores of 9.0 and 8.8 respectively. This is shown in \textit{figure 4.4.2b}, it should be noted that only EscC is shown here, as the InvG structure is a homology model based on EscC. The EscC/InvG structures also superimposed onto the NmPilQ\textsubscript{343-442} structure to give and RMSD of 1.52 Å over 78 residues and 1.46 Å over 73 residues respectively. Structural alignment with the T2SS GspD-N0 structure from \textit{E. coli}\textsuperscript{34} gave a lower Z-score of 7.1.

The DALI analysis also highlighted other close structural matches for the PilQ N0-domain from non-secretins. The first was a domain of the 43 kDa tail protein from the \textit{Shewanella oneidensis} prophage MuSO2 (\textit{figure 4.4.2c}), which aligns with NmPilQ\textsubscript{343-442} with a significant Z-score of 6.4 (Chang \textit{et al}, to be published). This protein in turn is an orthologue to gp27, one of two components of the bacteriophage T4 cell-puncturing device, at the baseplate of the needle structure.\textsuperscript{208,209} Studies have highlighted striking structural parallels and a common evolutionary origin between the bacteriophage cell-puncturing device and the Vgr proteins of the T6SS machine.\textsuperscript{209,210} NmPilQ\textsubscript{343-442} also has a similar backbone fold to the crystal structure of the \textit{E. coli} T6SS component VgrG, which is not a member of the secretin family, with a significant, albeit lower, Z-score of 3.1.

A second interesting result came from the outer membrane lipoprotein DotD, of the second subgroup of the T4SS, T4SSb, of \textit{Legionella pneumophila}, which is responsible for targeting of the outer membrane component of this secretion system, DotH.\textsuperscript{24} It is a small, 18 kDa protein consisting of a disordered N-terminus of about 80 residues and a conserved folded domain. DotD aligns with the NmPilQ\textsubscript{343-442} structure with a Dali Z-score of 5.9 (\textit{figure 4.4.2d}). This highlights an interesting parallel between the T4SSb and the T4PB/T2SS/T3SS, which is not obvious from amino acid conservation. It could be speculated that association of DotD/DotH may act as a secretin counterpart in a secretion system thought not to have this feature.
Other results highlighted include the 'plug' domain from the TonB-dependent receptor FpvA form *Pseudomonas aeruginosa* (Z-score 5.8), which has been discussed previously in the case of GspD-N0.\(^3\) In conclusion, all proteins with a similar fold to the NmPilQ N0-domain, as defined by a DALI Z-score of 2 or higher, belong to 3 main groups: Secretins, phage tail-associated proteins or T6SS components, and TonB-dependent ferripyoverdine receptors.
**Figure 4.4.2: Solution structure and structural homologues of NmPilQ**

**a:** Solution structure of NmPilQ$_{343-442}$, which is similar to the N0-domains from other secretins. In the ribbon plot, the unstructured cloning artefact is shown in grey and the helical linker to the N1-domain is shown in red. The full structural ensemble is shown on the right hand side, for the structured residues 343-431.

**b:** The crystal structure for domains N0 and N1 of the T3SS secretin EscC (PDB IS 3GR5) is shown in magenta superimposed onto NmPilQ$_{343-442}$, which is shown in green. The helical linker is not present in EscC.

**c:** The prophage MuSO2 tail protein (PDB ID 3CDD) is shown in magenta superimposed onto NmPilQ$_{343-442}$, which is shown in green. The tail protein adopts a trimeric ring assembly in vivo, to form the bacteriophage cell-puncturing device baseplate.

**d:** The T4SSb OM lipoprotein DotD is shown in magenta superimposed onto NmPilQ$_{343-442}$, which is shown in green. *Figures produced using CCP4MG.*
4.5 Model for NmPilQ\textsuperscript{343-545} by CS-ROSETTA and homology modelling

As discussed earlier, full structure determination for the double domain fragment NmPilQ\textsuperscript{343-545} was not possible, even after several experiments to extract high quality NOEs. Therefore, an attempt was made to produce a likely model for this fragment using chemical shift information and CS-ROSETTA,\textsuperscript{164} which would then be concatenated with the conventionally determined N0-domain structure using residual dipolar couplings to determine the relative orientation of the two domains which, as shown above, are joined by a helical linker. Chemical shifts were submitted to the CS-ROSETTA webserver for the two fragments, removing as much as possible of the unstructured regions outside the domain boundaries. The first fragment, spanning residues 343-437, included the N0-domain and the helical linker, and was largely intended for comparison with the \textit{de novo} determined structure for checking purposes. The second, spanning residues 419-514 would incorporate the helical linker and the N1-domain. It should be noted however there are some unusual backbone chemical shifts seen in NmPilQ\textsuperscript{343-545}, which were pointed out in the previous chapter (section 3.8). These are for residues 449K, 450Y and 495T, which have backbone amide shifts above 10 ppm, close to that of a tryptophan sidechain. This can occur when chelating a metal centre. Further experimentation, such as titration of EDTA into the sample, would shed further light on this. However these residues are present only in the tandem domain construct and not the shorter construct for which the structure was determined. Therefore they may affect the integrity of CS-ROSETTA modelling. In light of this, for a final model, several templates were used for the N1-domain. Using MODELLER\textsuperscript{165} in ‘multiple template mode,’ three template coordinate files for the C-terminal N1-region of NmPilQ\textsuperscript{343-545} were entered. The first two were N1-domain regions from the T3SS and T2SS secretins: EscC (PDB ID 3GR5) residues 105-173, GspD (PDB ID 3EZJ) residues 102-168. The third was the CS-ROSETTA model for N0/N1 NmPilQ. Then, for the helical linker, only the CS-ROSETTA model was used, spanning PilQ residues 419-437. For the remaining N-terminal residues, or the N0-domain, the \textit{de novo} NmPilQ\textsuperscript{343-442} structure was used as a template, up to residue 419. Use of separate templates for individual folded domains ensured no orientation bias in the model. MODELLER then
produces an output of 100 structures, which were subsequently aligned pair-wise and fed into the HADDOCK sever\textsuperscript{153} for clustering. This produced 7 clusters of structures related by an RMSD of 3 Å or less. Each cluster was treated as an ensemble and the closest to mean the representative was calculated. The seven representative models are presented in figure 4.5.1; each model differs with respect to the relative N0/N1-domain orientation.

An attempt was made to determine the correct domain orientation experimentally, by measurement of \textsuperscript{1}H\textsuperscript{15}N residual dipolar couplings. However this indicated no fixed orientation between N0 and N1 in solution, suggesting some degree of relative motion between the two domains. Different relative orientations between N0 and N1-domains were indeed reported in the case of EscC and GspD\textsuperscript{34,35}, although crystal-packing constraints could also be a factor in these cases. With the use of a cryo-EM generated density map of PilQ, we were able to identify one of the ensemble structures which best fits the EM data and other constraints: this analysis is presented in chapter 7 to form the main discussion.

It should be noted here, that to check the integrity of the CS-ROSETTA approach, the structure of the shorter construct, NmPilQ\textsuperscript{343-442}, was also determined this way and compared with the NOE-derived structure, with success, which installed confidence in the modelling for residues 419-545 (helical linker and N1). The helical linker region was included for both the NmPilQ\textsuperscript{343-442} and NmPilQ\textsuperscript{419-514} ROSETTA structure determination; in both cases the linker region consisted of a four-turn helix comprising P419-E432.
Figure 4.5.1: Models for the complete NmPilQ^{343-545} N0/N1 structure

The 100 resulting output structures from MODELLER were placed into seven clusters by HADDOCK under the criterion that structures within a cluster must give an RMSD of 3Å or less. Panels 1-7 show the closest-to-mean representative of each structure. Panel 8 shows the C-alpha trace of the seven representative structures superimposed. The main deviations lie within the linker, causing alternative orientations of N1 with respect to N0. Figures were produced using CCP4MG.
4.6 Structure determination of a B2 domain from PilQ by NMR (NmPilQ\textsuperscript{224-329})

As described in chapter 3, bioinformatic analysis of T4PB secretin sequences suggests the presence of one or more domains of high predicted beta-strand content at the N-terminus, before the N0 and N1-domains. Recombinant fragments from several gram-negative bacteria were shown to overproduce well (section 3.4) and some of these samples gave rise to good quality NMR spectra (section 3.9). The best sample, as determined by the concentration and quality of NMR spectra (section 3.9), came from a construct from \textit{N. meningitidis} PilQ spanning residues 225-329. This represents the second predicted beta-domain, which immediately precedes the N0-domain. The double-labelled $^{15}$N$^{13}$C protein sample prepared for the NMR screening in chapter 3 was used for these experiments. Spectra were processed using Topspin2.1 (Bruker) and the Azara processing package provided as part of the CCPN suite, and assignment was carried out using CCPN Analysis\textsuperscript{155} tools. Triple resonance experiments conducted on the sample are described in chapter 2.6. NMR spectra of the double-labelled $^{15}$N$^{13}$C sample were assigned to 92.3% completion and the labelled $^1$H$^{15}$N-HSQC crosspeaks are presented in \textit{figure 4.6.1}. The $^{13}$C-HSQC assignment is not shown.

Automated NOESY assignment and preliminary structure calculations were performed using CYANA 2.1 software,\textsuperscript{156,157} with input data of shift lists derived from $^{15}$N- and $^{13}$C-HSQC spectra. A total of 3881 NOESY peaks were assigned. CYANA 2.1 calculations ran with standard protocols using 7 cycles of automated NOE assignment and structural calculations, producing 100 structures per cycle. Water-refinement was conducted using CNS1.2\textsuperscript{158} with a total of 2607 unambiguous interproton distance restraints and 154 dihedral restraints. Dihedral $\phi$ and $\psi$ torsion angles were produced by TALOS-\textsuperscript{+}\textsuperscript{159} and the final ensemble of the best 20 low-energy, water-refined structures, and validated with PROCHECK-NMR\textsuperscript{160} using the iCing interface (http://nmr.cmbi.ru.nl/icing/iCing.html). Structure calculation statistics are presented in Table 4.6.
Figure 4.6.1: Assigned $^1H^15N$-HSQC for NmPilQ$^{224-329}$

Protein was buffer exchanged into 50 mM NaH$_2$PO$_4$/ Na$_2$HPO$_4$ pH 6.8; 100 mM NaCl; 10% D$_2$O and spectra were recorded at 298 K. Figure produced using CCPN Analysis.
**Table 4.6: NMR Structure calculation statistics for NmPilQ<sup>224-329</sup>**

<table>
<thead>
<tr>
<th>NMR constraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of distance constraints</td>
<td></td>
</tr>
<tr>
<td>Short range (</td>
<td>j-i</td>
</tr>
<tr>
<td>Medium range (1&lt;</td>
<td>j-i</td>
</tr>
<tr>
<td>Long range (</td>
<td>j-i</td>
</tr>
</tbody>
</table>

**Structure statistics (20 structures)**

<table>
<thead>
<tr>
<th>Violation statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of NOE violations &gt; 0.3Å</td>
<td>0</td>
</tr>
<tr>
<td>NOE violations &gt; 0.3Å</td>
<td>0</td>
</tr>
<tr>
<td>Maximum NOE violation</td>
<td>0.29</td>
</tr>
</tbody>
</table>

**Ramachandran statistics**

<table>
<thead>
<tr>
<th>Residues in most favoured regions</th>
<th>76.48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residues in allowed regions</td>
<td>19.75</td>
</tr>
<tr>
<td>Residues in generously allowed regions</td>
<td>3.09</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**RMS deviations from the mean structure**

<table>
<thead>
<tr>
<th>Protein backbone</th>
<th>0.64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein heavy atoms</td>
<td>1.40</td>
</tr>
</tbody>
</table>
The solution structure of NmPilQ revealed an eight-stranded beta-sandwich structure; this is presented in figure 4.6.2a. The structure was submitted to the DALI server for analysis against others in the PDB. The most significant matches belong to the group of eukaryotic heat shock proteins, with the closest result being the structure of wheat Hsp16.9. The beta-domain appears to be similar to the α-crystallin domain of this protein, which also forms a beta-sandwich. The structures align to give a Dali Z-score of 7.4, and superimpose to an RMSD of 3.02Å over 69 residues (figure 4.6.2b). The structures are also very similar in topology; the strand pairing is the same in each except that where B2 forms an 8-stranded beta sheet, Hsp16.9 forms a seven-stranded sheet, excluding β5. The X-ray structure of Hsp16.9 forms a dodecameric assembly, but this is unlike the PilQ dodecamer ring and instead formed by two trimers of dimers. However, the dimers within this structure form an elongated arrangement, which might be similar to the arrangement of B1 and B2 within PilQ assemblies.

Another similar fold identified from a different type of protein is that of the CS-domain from the human Sgt1 kinetochore complex with a Z-score of 6.5 (figure 4.6.2c). This protein domain functions to recruit Hsp90, an important chaperone for correct protein folding, to kinetochore and ubiquitin ligase complexes. However, the B2 domain fold is larger and includes an additional β-strand and unstructured loop, so that β4 pairs with β6, rather than β5. The two structures superimpose to give an RMSD of 2.11Å over 70 residues.

Sequence conservation on the surface of the B2 structure was analysed with ConSurf using sequences from 63 T4PB secretins, (PilQ homologues). The result is displayed in figure 4.6.3, where red regions correspond to highly conserved patches and blue regions to highly variable patches. Interestingly, the most highly conserved residues form a distinct patch on the surface of the structure, which brings together Lys 232 from the β1-strand with Asp 281 and Phe 282 from the loop between strands β4 and β5. This, together with the presence of highly conserved residues within an unstructured loop region, implies some possible functional significance, perhaps as a binding site for another protein.
Figure 4.6.2: NMR structure of NmPilQ$^{224-329}$ and structural homologues

a. NmPilQ$^{224-329}$ forms and eight-stranded anti-parallel beta-sandwich structure. Strands are coloured individually. A ribbon plot is shown on the left and a two-dimensional topology diagram is shown on the right.

b. The crystal structure for wheat Hsp16.9 (PDB ID 1GME) is shown in cyan superimposed onto NmPilQ$^{224-329}$, which is shown in magenta. The proteins are shown in the same orientation as (a) on the left, and rotated by 180° on the right, with the β5-β6-β4-β3 side facing forward. This face of Hsp16.9 differs in that β6-β4-β3 form a sheet and β5 is excluded.

c. The CS-domain from the human Sgt1 kinetochore complex (green) (PDB ID 1RL1) is shown superimposed onto NmPilQ$^{224-329}$ (magenta). The C-alpha trace is shown on the right.

Figures produced using CCP4MG.
Figure 4.6.3: CONSURF analysis of NmPilQ\textsuperscript{224-329}

Surface and ribbon plots of NmPilQ\textsuperscript{224-329} were generated using CHIMERA. Sequence conservation between 63 PilQ sequences was determined using CONSURF and mapped onto the surface of the structure. Red indicates a high level of conservation, white indicates medium and blue indicates highly variable areas of sequence.
4.7 Assignment of a B1 domain from *Aeromonas* (AhPilQ\textsuperscript{26-130})

The results presented in chapter 3 showed that attempts to overproduce the first beta-domain from PilQ in recombinant form were less successful than the B2-domain. Although some constructs were expressed reasonably well, protein products were produced in low-yield or were generally unstable, as judged by NMR screening. If we refer to chapter 3, *figure 3.9.2*, the most promising protein product came from a construct encoding the B1-domain from *Aeromonas hydrophilia*, AhPilQ\textsuperscript{26-130}. This gave the strongest and most dispersed \textsuperscript{1}H\textsuperscript{15}N-HSQC spectra, although this deteriorated over time. It was not therefore possible to extract the high quality NOEs required for structure determination in this case, although assignment of backbone chemical shifts was possible. Chemical shifts and secondary structure were assigned as described for NmPilQ\textsuperscript{343-545} (section 4.2). The assigned crosspeaks in the \textsuperscript{1}H\textsuperscript{15}N-HSQC spectra are shown in *figure 4.7.1a*. CSI information was used to deduce the secondary structure content, and this is presented in *figure 4.7.1b*. AhPilQ\textsuperscript{26-130}, or the B2 domain, contains eight β-strands very much like the B2 domain. This adds further evidence to suggest that B1 and B2 may share a common origin, and perhaps arise from an internal gene duplication event. It was mentioned in chapter 3, however, that not all T4PB secretins possess two beta-domains at their N-terminus, and also that some highly conserved residues found in B2 are not so in B1 (see chapter 3, *figure 3.9.1*).
Figure 4.7.1: Chemical shift and secondary structure assignment of AhPilQ<sub>26-130</sub>

a. Assigned crosspeaks on AhPilQ<sub>26-130</sub> <sup>15</sup>N-HSQC spectra. Protein was buffer exchanged into 50 mM NaH<sub>2</sub>PO<sub>4</sub>/ Na<sub>2</sub>HPO<sub>4</sub> pH 6.8; 100 mM NaCl; 10% D<sub>2</sub>O and spectra were recorded at 298 K. Figure produced using CCPN Analysis.

b. Chemical shift Index (CSI) calculated for deviations from random coil shifts of Ha, Ca and CO to determine the consensus secondary structure, graph adapted from CCPN analysis.
Discussion

The structural information presented in this chapter provides the most detailed model to date for the N-terminal periplasmic regions of the T4PB secretins. No atomic-resolution information about PilQ proteins is as yet available in the literature, with the majority of structural work concentrating on EM reconstructions of the PilQ dodecamer. Although similar work on the T2SS and T3SS secretins has been published, there has as yet been no definitive information to suggest the same structural features are conserved in the T4PB system. The results in this chapter demonstrate that the periplasmic N-terminus of PilQ from *Neisseria* consists of four distinct domains, namely two beta-domains, followed by an N0-domain and an N1 domain. In addition, the work presented here highlights some important differences from other secretins: the presence of a helical linker between domains N0 and N1 and the presence of N-terminal beta-domains.

The structured linker observed here between the N0 and N1-domains is not found in other secretins, such as those from the T2SS and T3SS, and it is also not found in structural homologues of N0, such as the T6SS VgrG proteins/ bacteriophage tail proteins and the T4SSb lipoprotein DotD. This adaptation of T4PB secretins may provide an added restraint on flexibility between these domains, which is not required in other systems. It should be noted that, although the residual dipolar coupling measurements indicated a degree of free rotation between these domains, the rotation correlation times obtained separately for each individual domain suggest that they do not tumble completely independently of one another.

The N0-domain structure also highlights previously unanticipated parallels with other secretion systems, which are not necessarily obvious from sequence alignment but come to light upon structural analysis. The similarity of N0 to part of the bacteriophage tail protein, the T6SS VgrG proteins, and the T4SSb lipoprotein DotD, indicate that this type of structure plays a vital role across the board in bacterial secretion, perhaps as important protein-protein interaction domain. The T6SS appears distinct from the T2SS, T3SS and T4PB systems upon first inspection, and yet evidence such as this is emerging to suggest a closer
relationship. This highlights the growing importance of structural homology between unrelated proteins.

The second important distinction between the T4PB and the T2SS/T3SS secretins is the presence of individual beta-domains at the N-terminus. These are unique and appear to be specific to the T4PB system secretins. The structure presented here is novel to the field and distinct from other published secretin domain structures. As yet, however, the function of beta-domains within the pilus system remains a mystery. It has previously been demonstrated that the pilin component of the extruded substrate is able to fill the PilQ chamber when the two proteins are combined. It is conceivable that the beta-domains, taking into account their position and specificity to the pilus system, may play a role in the regulation or gating of entry of the pilus fibre.

The challenges encountered in the isolation of homogenous, well folded preparations of the first beta-domain from several PilQ homologues indicate a degree of instability, at least in recombinant form, which may also extend to the in vivo state. The question remains as to the requirement of these domains for PilQ function. Some PilQ orthologues do not contain a B1 domain, and their suspected structural similarity to the B2-domain indicates that they may have arisen as a result of a gene duplication event. This could mean that the first beta-domain is surplus to requirements for a functioning secretin. The specific requirements of each beta-domain would form the basis for an interesting in vivo study, as deletion of the first domain and mutagenesis of the suspected binding site on the second may affect formation of pili on the surface of bacteria.

In summary, with regards to the N0/N1-domain structure, there is a general principle emerging where a common fold is shared between secretins of different systems and other proteins too. Structural similarity does not necessarily translate into functional similarity, but sometimes it can give a clue to it, where it may exist. Another example is the T3SS, where different components, namely EscC, PrgH and PrgJ, have similar folds and build into large periplasmic ring structures. Meanwhile, the sequence diversity maintains the specificity of the system by varying oligomerization number and providing unique surface properties.
Chapter 5: Structural analysis of the secretin-interacting lipoprotein PilP by X-ray crystallography

5.1 Introduction

The previous chapters discussed structural studies on secretin domains. An important area of investigation, therefore, is how these domains interact with other system components. The next part of the project focuses on structural studies of the essential secretin-interacting protein, PilP, in this chapter, followed by its interaction with PilQ in chapter 6. PilP is a ~20 kDa, inner-membrane anchored lipoprotein essential for the biogenesis of type IVa pili. The protein has a conserved lipid attachment site at the N-terminus preceding a folded beta-rich domain at the C-terminus. PilP is essential for piliation, and is co-transcribed with the secretin PilQ.\textsuperscript{141,185} Evidence suggests PilP and PilQ share a specific interaction: both the N- and C-terminal regions of PilP appear to recognize the central part of the PilQ monomer.\textsuperscript{141} If we refer back to the domain organization of secretins, those of the T4PB system have extra beta-rich domains at the extreme N-terminus which are not present in other secretins such as GspD, which start with an N0-domain. Therefore, the central part of a PilQ monomer might correspond approximately to the N0 or N1-domain.

The precise function of the PilP protein and the nature of its interaction with the secretin remains an enigma. Several studies have identified a number of partner lipoproteins known as ‘pilotins’ which are required for the functionality of secretin pores, specifically their correct assembly or insertion into the outer membrane. One such pilotin protein is partnered with the MxiD secretin, which forms part of the T3SS in \textit{Shigella}. MxiM is a small lipoprotein and forms a cracked β-barrel structure, disrupted by an alpha-helix.\textsuperscript{42} Within the cleft of this barrel, MxiM binds an 18-residue portion of its cognate secretin which adopts a defined structure upon binding, demonstrating a strong structure-function relationship. In addition, the hydrophobic binding groove of MxiM can bind lipid,\textsuperscript{42} so this may prompt the pilotin to release the secretin upon membrane targeting via an allosteric switch mechanism.\textsuperscript{43}
Another type of pilotin protein, belonging to the T2SS, is collectively referred to as the GspS pilotin. Each pilotin has been found to associate with the extreme C-terminus of the relevant secretin in some systems, such as *Klebsiella oxytoca* (PulS)\(^\text{215}\) and *Erwinia* species (OutS).\(^\text{50}\) However, a related component does not appear to be present in most other characterized examples of the T2SS, such as that of *Aeromonas* and *Vibrio* species.\(^\text{49}\) It is a small, ~12 kDa, peripheral outer membrane lipoprotein thought to associate with a disordered C-terminal region of its cognate secretin, generally known as the S-domain.\(^\text{129,215}\) It has been demonstrated previously that OutS is necessary for stabilization of the OutD secretin, and protects it from proteolytic degradation.\(^\text{27}\) In the absence of PulS, the PulD secretin can still form dodecameric complexes, but these mislocalize to the cytoplasmic membrane, inducing a phage shock-type response.\(^\text{216}\) It is perhaps possible that PulS is required to prevent premature multimerization of PulD monomers, and has also been shown to be required for correct targeting of PulD monomers to the outer membrane.\(^\text{217}\) It is also noteworthy that the PulS binding site is predominantly located within 30 residues of the unfolded PulD C-terminus,\(^\text{49,50,215}\) and therefore after the membrane-spanning C-domain; this region adopts helical structure upon its interaction with PulS.\(^\text{43,50}\) PulS adopts a novel, helical-bundle fold, with a groove along one side of the structure lined with charged and hydrophobic residues, which are believed to form the binding site for the PulD S-domain.\(^\text{49}\) Thus, PulS and MxiM both appear to play roles in secretin assembly and outer membrane localization, and both bind to regions within the secretin C-terminus. Their structures and the mechanisms by which they interact with their cognate secretins, however, have clear distinctions.

The assembly and localization of T4PB secretin, PilQ, seems to be promoted in a different way. PilW is fairly widespread across bacteria expressing type IVa pilins and in its absence no PilQ secretin multimers are detected.\(^\text{44,46,47}\) Similar to PilP, PilW possesses a canonical lipobox motif, and is also lipidated via an N-terminal cysteine residue. Lipidation is not crucial for PilW functionality *per se* although correct localization by the lipid anchor greatly enhances its stability.\(^\text{218}\) Crystals of PilW reveal a superhelix of six tetratricopeptide domains,\(^\text{44,218}\) which bears no similarity to the MxiM pilotin structure. The molecular basis for recognition of PilQ
by PilW is therefore likely to be different from MxiD/MxiM. Bioinformatic analysis shows that type IV pilus-dependent secretins also do not have the longer C-terminal extensions found in T2SS secretins; in other words, they are missing the parts which are recognized by T2SS pilotins.

PilP was initially thought to affect PilQ multimerization in Neisseria meningitidis in a similar fashion to PilW, but more recent work has shown that this is not the case and that the original observation was caused by a polar effect of the pilP mutation on pilQ transcription.\(^{141}\) PilP also does not stabilise PilQ multimers, as these can still be detected in pilP mutants where transcription of pilQ is not reduced.\(^{141}\) The precise function of PilP is therefore unclear and it cannot be classed as a pilotin, as has previously been suggested.\(^{219}\) Recent studies in Pseudomonas aeruginosa have shown that PilP in this bacterium forms part of an inner membrane complex, binding to the cytoplasmic protein PilM and transmembrane proteins PilN and PilO.\(^{112}\) The inner-membrane localization of the protein provides a clear distinction with the proposed mechanisms of GspS and MxiM in mediating secretin localization to the outer membrane and secretin assembly, and the previously proposed mechanism of PilP as a pilotin is now accepted not to be the case.\(^{45,219}\)

### 5.2 Similarity of PilP to GspC

Bioinformatic analysis of the PilP sequence suggests it bears similarity to the “homology region” (HR) domain of GspC from the T2SS. GspC is another protein that is now known to interact with the secretin.\(^{151,152}\) In the case of the Erwinia species, it is the only other gene in the Out cluster (besides OutD) within which mutations cannot be complemented by the corresponding gene of a closely related species, to enable the extrusion of species-specific type II substrates.\(^{220}\) This implies that a specific GspC homologue is required for either substrate recognition or secretin assembly. Unlike PilP, GspC consists of several domains (figure 5.2.1). A short N-terminal domain resides in the cytoplasm, and the protein then crosses the membrane via a single transmembrane helix. Within the periplasm there resides both the HR-domain and a second domain, which is usually a PDZ or coiled-coil domain.\(^{143,144,147,152}\) Gerard-Vincent and co-workers examined the domain functionality of XcpP and were able to narrow the non-exchangeable
domain down to a region of 35 C-terminal residues of the periplasmic domain, preceded by the transmembrane domain, which would correspond to the HR-domain. Therefore, like PilP, this domain is playing a vital role. More recent publication of the crystal structure of the GspC HR-domain has confirmed its structural analogy with PilP. This is discussed in more detail in section 5.7 of this chapter.

**Figure 5.2.1: Domain structure of PilP and GspC Proteins**

Schematic view of PilP from the T4PB system and GspC from the T2SS. PilP contains an N-terminal prokaryotic lipidation site, or lipobox, followed by a disordered proline rich N-terminal region which precedes the conserved folded domain. GspC possesses a helical transmembrane domain, a “homology region” (HR) domain and PDZ or coiled coil domain. The folded domain of PilP and the HR domain of GspC have a similar β-sandwich tertiary fold, although they are unrelated in sequence.
5.3 Domain organization of PilP proteins

The previously determined NMR structure of PilP from *Neisseria meningitidis*\textsuperscript{113} describes the C-terminal domain as a compact, globular beta-sandwich fold preceded by a flexible N-terminal region, containing a covalent lipid-attachment site at Cys16.

The PilP family, Pfam entry PF04351,\textsuperscript{221} comprises 466 sequences containing only the PilP domain and signal peptide sequence distributed across a wide range of proteobacteria. A clustalW2 ([http://www.ebi.ac.uk/Tools/msa/clustalw2/](http://www.ebi.ac.uk/Tools/msa/clustalw2/)) seed alignment of 54 non-redundant members of this phylum was generated, and a representation showing sequence alignments from eight genera is shown in *Figure 5.3.1*. The N-terminal region of the protein is proline-rich and secondary structure predictions, carried out using PSIPRED,\textsuperscript{222} suggest that the majority of the sequence between the conserved cysteine lipid attachment site and the beginning of the first short alpha helix in the structure, forms random coil. Sequence conservation is higher across the family at the C-terminus within the folded domain, as determined for *Neisseria meningitidis* PilP.\textsuperscript{113}

In order to analyse the interaction between PilP and PilQ in more detail the C-terminal domains of PilP from *Neisseria meningitidis* and *Pseudomonas aeruginosa* were expressed, as these were considered to be more likely to crystallize and/or be more suitable for NMR interaction studies with PilQ. Both C-terminal domains were produced in *E. coli*, purified and are abbreviated as PaPilPC and NmPilPC for the *Pseudomonas* and *Neisseria* proteins respectively.
Figure 5.3.1: Alignment of PiIP protein sequences

Alignment of PiIP sequences from eight genera. The extent of the conserved folded domain from *Pseudomonas* was expressed from the PaPiIP^C^ construct in this study is indicated above the sequences by an arrow. The positions of secondary elements are also indicated. Absolutely conserved residues are highlighted in bold.
Preparation and characterization of PilP proteins from *Neisseria meningitidis* and *Pseudomonas aeruginosa*

Based on secondary structure prediction algorithms and sequence alignments, constructs for the conserved, folded C-domains of PilP from *Neisseria meningitidis* MC54 (residues 77-164) and *Pseudomonas aeruginosa* PAO1 (residues 83-171) strains were produced. In addition, for later studies, a construct of the *Neisseria* homologue producing a longer protein, NmPilP\textsuperscript{20-181} was used. This construct produces a recombinant fragment, which includes some of the N-terminal flexible linker region as well as the folded domain, however it does not include the signal sequence or lipid-attachment site. NmPilP\textsuperscript{20-181} has been used for previous studies of PilP.\textsuperscript{113} All three proteins were engineered to contain an N-terminal histidine purification tag.

5.4 Production and purification of PilP proteins

- **Protein production trials**

  For protein production trials, the PaPilP\textsuperscript{C} and NmPilP\textsuperscript{C} constructs were transformed into *E. coli* T7 Express cells and grown in selective medium to an OD\textsubscript{600} of 0.6-0.8. The culture was separated into aliquots to induce protein production with 0.1 mM or 1 mM IPTG, and overnight induction at lower temperature was also tested. Relative levels of protein were compared by western blot, using an antibody against the His-tag to detect the protein of interest.

  *Figure 5.4.1* shows the western blot analysis of small-scale protein production trials. Western blotting, using a primary antibody against the His-tag, failed to detect protein without IPTG induction. After three hours of induction PaPilP\textsuperscript{C} and NmPilP\textsuperscript{C} were produced at comparable levels in samples induced with 1 mM IPTG compared with those induced with 0.1 mM IPTG. After 3 hours 1 mM IPTG gave a slightly larger amount of protein from both constructs. Although significant amounts of both proteins are produced at lower concentrations, for crystallization and NMR studies a higher yield of protein and additional purification steps will probably be required. Therefore, a larger scale production and purification was initiated in 2 L cell cultures using an induction step of 1 mM IPTG for 3 hours at 37°C for both constructs. It also was possible to obtain comparably high levels of
protein with an overnight induction at 20°C using 0.1 mM IPTG; this approach was sometimes used for minimal media preparations when cell cultures took considerably longer to reach mid-log phase (not shown). The bands on the western blot shown in the figure appear smeared, most likely because of the high levels of protein and the presence of cell debris. Therefore, the 3 hour induction period/1 mM IPTG samples were centrifuged at high speed for 20 minutes to remove cell debris, and analysed again. This is shown in the right hand panel in figure 5.4.1. In this blot, the bands are clearer, and there does not appear to be any multiple banding, which may have been indicative of proteolysis during expression.

Figure 5.4.1 Small-scale protein production trials of NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C}
Western blot analysis of whole cells transformed with NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C} constructs. Samples analysed were uninduced (U), and induced for 1 hour and 3 hours at 37°C with 0.1 mM IPTG and 1 mM IPTG. Molecular weight markers (M) are given in kDa. The right hand panel shows both proteins after induction for 3 hours at 37°C with 1 mM IPTG, after removal of cell debris by centrifugation.
- Purification of NmPilP<sup>C</sup>

Upon cell disruption, the protein was found to be present in the soluble fraction of the cell lysate; NmPilP<sup>C</sup> was subsequently purified from other *E. coli* proteins by immobilized metal-affinity chromatography. The initial preparation of this protein was intended for crystallography trials therefore a Tris, rather than phosphate-based, binding and lysis buffer was used. For the Tris buffer conditions, cells were lysed using Tris-HCl pH 8.0, 100 mM NaCl containing 1 mg ml<sup>-1</sup> lysozyme and 40 μg ml<sup>-1</sup> DNase I. The concentration of NaCl was generally kept at 100 mM to aid protein stability during storage but this was sometimes removed or reduced by desalting if necessary. For later preparations where the protein was required for NMR, the Tris buffer was replaced by 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 7.8. For the purposes of purification, the choice of lysis buffer did not seem to affect the yield of protein during the process, which was typically ~6 mg for a 2 L cell culture.

*Figure 5.4.2a* shows selected samples from the purification procedure. Prior to elution the Ni<sup>2+</sup>-NTA resin was washed in lysis buffer containing a low concentration of imidazole (20 mM), which removed some contaminants that form non-specific interactions with the nickel column (*figure 5.4.2a, lanes 2-4*). Due to the high level of protein produced the eluate contained ~80% NmPilP<sup>C</sup>, although some higher molecular weight contaminants are still present after purification (*figure 5.4.2a, lanes 5-8*). Therefore a further purification step by size exclusion chromatography was introduced. The imidazole was removed from the eluate by dialysis, and the purified NmPilP<sup>C</sup> samples were exchanged into the required buffer. Several downstream buffer conditions were used for the protein throughout the project for various experiments, and these are noted where appropriate.

For preparations requiring removal of the histidine purification tag, the protein was treated with thrombin for 16 hours at 4°C during dialysis. Thrombin cleavage should remove an estimated 1.9 kDa fragment, with the sequence MGSSHHHHHHSSGLVPR, which includes the purification tag. The remaining sequence at the N-terminus, essentially introduced as a cloning artefact, consists of four residues, GSHM, after cleavage. The thrombin was then removed from the sample (2.3.12), and a reverse IMAC purification step was conducted to remove
any remaining uncleaved NmPilPC. *Figure 5.4.2a lanes 9-11* show selected samples from this process. The thrombin-treated sample was passed down a nickel column and the flow-through collected (*lane 9*). After a wash step with 20 mM imidazole (*lane 10*), remaining His-tagged protein was eluted with 200 mM imidazole (*lane 11*). Although the recommended optimal temperature for thrombin activity is 16°C, the cleavage was >90% complete under these conditions (*lane 9*), leaving very little remaining uncleaved protein (*lane 11*). Some precipitate was observed in the protein preparation during the dialysis step, both in the thrombin-treated and untreated preparations. This was removed by centrifugation at 14000 x rpm for 30 minutes at 4°C. This may be attributed to removing the stabilising effects of high concentrations of imidazole, or in some cases the pH of the buffer may have been too close to the theoretical pI of uncleaved NmPilPC (6.39). The large level of expression could also be a contributing factor. The protein was often exchanged into buffers of pH 6.1 or 6.8 for NMR spectroscopy. The cleaved protein remained in solution without any apparent degradation, as judged by SDS-PAGE, at 4°C in most of the phosphate buffer conditions used for several days. Buffer pH values for various experiments were in the range of 6.1 to 7.4. This pH range may sufficiently deviate from the theoretical pI of NmPilPC (6.39 uncleaved/5.07 cleaved). The protein was stable up to concentrations of ~1 mM.

- **Purification of PaPilPC**

The C-domain from *Pseudomonas aeruginosa* PilP, PaPilPC, was purified in the same manner as the equivalent protein from *Neisseria*, and with similar results. As this protein was intended for crystallography trials, the protein was exchanged into 50 mM Tris-HCl pH 9, 200 mM NaCl after elution. PaPilPC was produced as a soluble protein in *E. coli* and samples from each step of the purification procedure are shown in *figure 5.4.2b*. As was the case for NmPilPC, the presence of *E. coli* contaminants in the eluate (*lanes 5-8*) meant further purification by size-exclusion was advisable. PaPilPC was also successfully cleaved with thrombin during the dialysis step at 4°C. *Figure 5.4.2b, lane 9* shows the flow-through from a nickel column after thrombin treatment. As with the *Neisseria* protein, there is little remaining uncleaved protein eluting from the column with 200 mM imidazole (*figure 5.4.2b, lane 11*). Both recombinant PilP constructs were engineered to
include part of the flexible N-terminus, to avoid truncating the folded domain. It is possible that, in both cases, the presence of the purification tag at the end of the predicted flexible N-terminus of PilP made it more accessible to the enzyme. This was useful as it allowed the protein to be kept at lower temperatures during treatment making it less susceptible to proteolysis or denaturation.

Figure 5.4.2: IMAC purification of NmPilP\(^C\) and PaPilP\(^C\)
NmPilP\(^C\) (a) and PaPilP\(^C\) (b) were purified by IMAC from \textit{E. coli} whole cell lysates and treated with thrombin protease to remove the purification tag. Samples from each step of the purification process were analysed by SDS-PAGE and the gels are shown above. For both gels the samples are as follows - Lane 1: cell lysate flowthrough from the nickel column; Lanes 2-4: wash step with 20 mM imidazole; Lanes 5-8: elution step with 200 mM imidazole. Lanes 9-11 represent stages after thrombin protease treatment. Lane 9: flowthrough from nickel column containing cleaved protein; Lane 10: wash step with 20 mM imidazole; Lane 11: elution of uncleaved protein with 200 mM imidazole.
5.5 Analysis of oligomeric state of NmPilP<sup>C</sup> and PaPilP<sup>C</sup>

The concentrated protein preparation was analysed further by size-exclusion chromatography (SEC). *Figure 5.5.1a* shows a typical elution chromatogram of a preparation of NmPilP<sup>C</sup> still carrying the purification tag. In this case, there is a single sharp peak at 14.22 ml elution volume (peak 2), along with a much smaller peak, of apparent higher molecular weight, eluting at 8.98 ml (peak 1). The elution volume for NmPilP<sup>C</sup> differed only marginally according to alternative running buffers used throughout the project. During this particular preparation, some precipitate was visible after dialysis which was removed by centrifugation. Therefore the elution fractions were analysed by SDS-PAGE, to check for NmPilP<sup>C</sup> aggregates in the high mass peak (peak 1). *Figure 5.5.1b* shows the results: upon analysis of the indicated fractions it was evident that the major peak on the chromatogram, peak 2, exclusively contained NmPilP<sup>C</sup>, which was in excess of 95% purity (*lanes 4-11*). The smaller, higher molecular weight peak, peak 1, did not appear to contain any NmPilP<sup>C</sup> (*lanes 1-3*). In *lanes 1-3*, a faint band can be seen at around 75 kDa, indicated by a black arrow on the gel. Therefore it is likely to be a contaminant, as opposed to NmPilP<sup>C</sup> multimers or aggregates. This band is difficult to see on the figure, but was more clearly visible on the actual gel.

*Figure 5.5.2* shows the typical elution chromatogram from a thrombin-treated sample of purified NmPilP<sup>C</sup> (*red trace*). Included in the same figure, is the equivalent chromatogram for thrombin-treated PaPilP<sup>C</sup> (*blue trace*), as the two proteins gave very similar elution profiles. Cleaved NmPilP<sup>C</sup>, as expected, eluted a little later than the untreated protein at 15.34 ml. Similar to the *Neisseria* protein, the SEC elution profile for cleaved PaPilP<sup>C</sup> showed a sharp single peak eluting at 15.22 ml, in the same running buffer. There were also minor peaks in these preparations accounting for some contaminant proteins. Fractions containing only PilP<sup>C</sup> were collected for all preparations and the protein was recovered with >95% purity. This was therefore deemed suitable for crystal trials.

For a more accurate estimate of the oligomeric state of NmPilP<sup>C</sup>, a sample complete with the His-tag was analysed by SEC coupled to multi angle laser light scattering (SEC-MALLS). The resulting chromatogram is shown in *figure 5.5.3*.

These combined measurements indicated that the protein in peak 1, highlighted...
on the figure, had an estimated molecular mass of 14.05 kDa with a polydispersity of 1.003; indicating a homogenous, monomeric sample in solution. This differs slightly from the predicted molecular weight calculated according to amino acid sequence, which is 12.05 kDa.

![Diagram](image)

**Figure 5.5.1: Size Exclusion Chromatography of NmPilPc**

a. Chromatogram for NmPilPc with the purification tag intact as it elutes from a superdex 75 10/300 GL size-exclusion column. The running buffer used was 50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 7.8, 100 mM NaCl. The protein elutes as a single sharp peak at 14.22 ml. Molecular weight standards are also shown.

b. Eluted fractions containing protein from peak 1 and peak 2 were analysed by SDS-PAGE. These are indicated by arrows on the chromatogram in (a). Lanes 1-3: Peak 1, elution fractions 8.5-10 ml. Lanes 4-11: Peak 2, elution fractions 13-17 ml.
Figure 5.5.2: SEC of NmPilP<sup>C</sup> and PaPilP<sup>C</sup> using a Superdex 75 column
Elution profiles for size exclusion of thrombin-treated NmPilP<sup>C</sup> and PaPilP<sup>C</sup> are shown in red and blue respectively. Both proteins elute as a single sharp peak at 15.34 ml (NmPilP<sup>C</sup>) and 15.22 ml (PaPilP<sup>C</sup>). The running buffer used was 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>pH 7.8, 100 mM NaCl for both proteins. Molecular weight standards are shown by arrows.

Figure 5.5.3: SEC-MALLS analysis of NmPilP<sup>C</sup> using a Superdex 75 column
Size-exclusion chromatography was coupled to multi-angle laser light scattering for an accurate representation of molecular weight and homogeneity. The running buffer used was PBS. The protein contained within the major peak, which is indicated as peak 1, had a calculated molecular mass of 14.05kDa, indicating a monomeric form of the protein.
5.6 Structural studies of NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C} by X-ray Crystallography

- Crystallization and data collection

This section reports attempts to crystallize both NmPilP\textsuperscript{C} and PaPilP\textsuperscript{C}, along with the determination of the crystal structure of PaPilP\textsuperscript{C}, the first structure of this type of protein determined by X-ray crystallography. Although an NMR structure of NmPilP\textsuperscript{C} had already been reported by our group previously,\textsuperscript{113} a crystal structure was also needed, as the starting point of a concerted structural programme to investigate the interactions between the proteins of T4PB. In the case of my project, trials were also conducted on possible PilP\textsuperscript{C}: secretin domain complexes; a crystal structure of the PilP\textsuperscript{C} component could aid in structure determination and understanding any structural changes that may occur upon complex formation. Shortly after this work was completed an NMR structure was also reported for \textit{Pseudomonas} PilP.\textsuperscript{142}

For crystallization trials, NmPilP\textsuperscript{C} was purified and concentrated in 50 mM Tris-HCl pH 9, 250 mM NaCl. Initial trials were conducted at both 4°C and 20°C and produced no suitable hits. Many wells contained multiple needle-like crystals and there was considerable precipitate in most of the drops. Therefore, although nucleation events occurred in many cases, large, well-ordered crystals failed to grow. Attempts at optimization by screening additives also failed, producing similar results. Re-seeding of nucleated protein into the initial screens was also attempted, but improved conditions supporting crystal growth did not become apparent.

Trials with PaPilP\textsuperscript{C}, however, were more successful. It may be the case that some homologues of PilP\textsuperscript{C} will crystallize more readily than others because of variations in the sequence. The crystal shown in \textit{figure 5.6.1a} was obtained from one of the initial trials. Only one crystal was present in the drop; when a seeding approach was used, several more grew under the same conditions, but these were smaller and more irregular in appearance (\textit{figure 5.6.1b}).

Data were collected from the crystal in \textit{figure 5.6.1a} to a resolution limit of 1.7Å using a lab X-ray source. It should be noted that 96.2% of the data for this crystal
form were collected to 2.69Å resolution, but higher resolution data ranges were slightly less complete. The data collection statistics are shown in Table 5.6. The final data set was from 360 images, and was 87% complete after discarding marginal data. The space group was indexed as $P2_1$, and data processed using MOSFLM$^{223}$ and SCALA$^{224}$.

An attempt was made to obtain experimental phases by production of PaPilP$^C$ from seleno-methionine labelled media, and re-screening using the initial hit crystallization conditions. Unfortunately it was not possible, even after several attempts, to reproduce the crystals. In addition, attempts to solve the structure by molecular replacement using the Neisseria PilP NMR structure$^{113}$ were unsuccessful. At this stage the more recent Pseudomonas PilP NMR structure had not been published$^{225}$. A search for an alternative crystal form, which could be used for phasing, was therefore undertaken.

The liquor composition of the second crystal form contained 0.6 M zinc acetate (figure 5.6.1c). This highlighted the possibility that Zn$^{2+}$ ions might be bound within the crystal lattice, which would provide an opportunity for phasing by anomalous diffraction. A fluorescence scan carried out on a single crystal, and processed using CHOOCH$^{226}$ showed a strong peak for zinc, as expected, and two datasets were collected at peak and inflection wavelengths. The statistics for this crystal are also shown in Table 5.6. Data processing using MOSFLM$^{223}$ and SCALA$^{224}$ confirmed the presence of a strong anomalous signal. This crystal was indexed into the $C2$ spacegroup. Phasing, density modification and model building was carried out using Crank,$^{227}$ part of the CCP4 suite of programs.$^{228}$ Site identification was carried out using SHELXD-2006$^{229}$ as implemented within CRUNCH2,$^{230}$ phasing was carried out using BP3,$^{231,232}$ hand selection using SOLOMON$^{233}$ and density modification using DM.$^{234}$ The mean FOM at this stage was 0.646, after identification of 6 zinc atoms. Automated model building, conducted using ARP/WARP,$^{235}$ built 283 residues at this stage. It was readily apparent at this stage that there are four PaPilP$^C$ molecules in the asymmetric unit in this crystal form. The model was completed by manual rebuilding using the graphics package COOT.$^{225}$ Maximum likelihood refinement was subsequently carried out using REFMAC5.$^{236}$ Non-crystallographic symmetry restraints were
applied during early stages of refinement but relaxed in the final stages; refinement statistics are summarized in Table 5.6. Electron density was generally weak or absent for the first 10 residues at the N-terminus, although some residues were built in this region where the density was sufficiently strong to unambiguously identify them. Density was also weak between Pro155-Trp161 for all four chains within the asymmetric unit, so these residues were also omitted from the final model. Chain A from the C2 crystal form was used to solve the structure of the P21 crystal form, using the program MOLREP.237 The structure comprised a dimer in the asymmetric unit and was refined, again using REFMAC5,236 without the application of non-crystallographic symmetry restraints. This data gave rise to a more complete model than that from the C2 crystal form. Electron density was better defined for the N-terminus, and also the Pro155-Trp161 loop could be built unambiguously where this was not possible for the C2 crystal (figure 5.6.2). The model was built, using COOT,225 for Lys84-Leu170 (chain A) and Ile83-Leu170 (chain B). Stereochemical parameters for both final structures, determined using PROCHECK,238 were within the expected ranges, or better, for a structure determined at 1.7 Å resolution.

Figure 5.6.1: PaPilPC crystals produced in this study

a. PaPilPC crystal grown in 0.1 M HEPES pH 7.5, 8% (v/v) ethylene glycol, 20% (w/v) PEG 10,000. This crystal indexed in the P21 spacegroup and a complete dataset was taken.

b. PaPilPC crystals were reproduced in the same crystallization liquor as in (a) (0.1 M HEPES pH 7.5, 8% (v/v) ethylene glycol, 20% (w/v) PEG 10,000) using a seeding method. Data was not collected.

c. PaPilPC crystal grown in 0.1 M Tris-HCl pH 8.5, 0.6 M zinc acetate. Data was collected and the crystal indexed in the C2 spacegroup.
Adjacent PaPilP\textsuperscript{C} monomers are linked in the \textit{C2} crystal form by the binding of zinc ions between residues Glu103 and His125 of adjacent molecules (\textit{figure 5.6.3a}). PaPilP\textsuperscript{C} thus forms a tetramer with quasi-four-fold symmetry (\textit{figure 5.6.3b}). Analysis of the structure by PISA\textsuperscript{199} shows that tetramer formation is driven by the zinc atoms in the crystallization liquor. The calculated free energy changes for interfaces formed by zinc atoms are considerably more favourable than those formed by polypeptide chain contacts. As shown earlier (section 5.5) PaPilP\textsuperscript{C} is monomeric in solution, therefore tetramer formation is an artefact of crystallization.

Both structures were analysed by MolProbity\textsuperscript{239} for structure validation. Ramachandran plots for each are presented in \textit{Figure 5.3.4}. For the \textit{P2\textsubscript{1}} crystal, 97.7\% of residues were within favoured regions and all were in allowed regions with no outliers. Similarly, for the \textit{C2} crystal, there were 97.0\% within favoured regions and all were within allowed regions, again with no outliers.
**Figure 5.6.2: Electron density for the PaPilP^C Pro155- Trp161 loop**

Fo-Fc map (σ1.51) for the Pro155-Trp161 loop was weak in the C2 crystal form (left), but the model could be built unambiguously into electron density from the P2_1 crystal (right). The green density indicates positive contours. Images produced using COOT.

**Figure 5.6.3: Zinc atoms bound within the PaPilP^C C2 crystal**

a. Electron density for Zinc atoms coordinated by glutamate and histidine residues of adjacent chains in the tetramer. Image produced using COOT.

b. PaPilP^C tetramer formation in the C2 crystal form. Chain A is shown in magenta, B in red, C in cyan and D in green. Six zinc atoms were bound within the asymmetric unit and these are shown in grey.

**Figure 5.6.4: Ramachandran plots for PaPilP^C**

General case Ramachandran plots for the P2_1 (left) and C2 (right) crystal forms are shown. In both cases, 100% of residues lay within allowed regions and there were no outliers. Geometry was analysed using MolProbity.
- Structure determination of PaPilP<sup>C</sup>

In both crystal forms, the domain comprises a seven-stranded beta-sandwich fold with two short $3_{10}$ helices at the N-terminus of the domain. The loop connecting the $\beta1$-$\beta2$-$\beta3$ sheet to the second $\beta5$-$\beta6$-$\beta7$ sheet packs against the N-terminal helices and incorporates the shorter strand $\beta4$. The presence of three, highly conserved, glycine residues within this loop - numbered 130, 134 and 138 in the *P. aeruginosa* PilP sequence - suggest it is important for the structural integrity of the domain (figure 5.6.5a). In the $P2_1$ crystal form, for each PaPilP<sup>C</sup> monomer, the $\beta1$ strand forms a $\beta$-$\beta$ alignment with a $\beta1$ strand from an adjacent symmetry-related copy in the crystalline lattice. PISA analysis indicated that this is the most energetically favourable interface found in the crystal (computed $\Delta G = -3.5$ kcal/mol). *Figure 5.6.5b* shows the dimer interface formed between symmetry mates which, as well as forming a $\beta$-$\beta$ mainchain interaction, form hydrogen bonds between the Thr107 and Ser109 side chains of respective polypeptide chains. This beta-strand pairing may provide a clue to the molecular basis of its recognition of the secretin PilQ, or the PilM,N,O inner-membrane platform. Evidence to date, however, suggests that PilP forms an interaction with the inner membrane platform via its disordered N-terminus, not its folded domain, so the latter scenario seems less likely.\(^{112,142}\) It is important to note here that the results presented later in chapter 6, show that this crystal contact is via the same beta-strand which is involved in forming an interaction with the PilQ N0-domain.

Another feature of the PaPilP<sup>C</sup> structure is the presence of a surface exposed tryptophan residue, Trp 161, in a loop region between strands $\beta6$ and $\beta7$, which appears to hold down the unstructured N-terminus. This is indicated on the structure in figure 5.6.6, and this residue appears to be well conserved, although it is not present in all PilP homologues (see figure 5.3.1). The results presented later in chapter 6 show that this residue is not involved in the PilP/PilQ interaction; however, as it is both conserved and surface exposed, it seems unlikely that it would not bear some functional significance. This residue does not form the same molecular interactions with the N-terminus in the solution-state NMR structure from *Neisseria* which is also shown in the figure. It would therefore seem that this interaction seen in PaPilP<sup>C</sup> is forced by crystal packing, rather than being of
biological relevance. However, this does not rule out the possible importance of this conserved residue for another unknown purpose.

A computed electrostatic surface of PaPilP\textsuperscript{C} is shown in \textit{figure 5.6.7}. The protein does not have any readily apparent localized concentration of charge to indicate a binding site or suggest functional significance. For comparison, the electrostatic surface for PilP from \textit{Neisseria} is also shown in \textit{figure 5.6.7}, as viewed from a similar orientation. This protein has a more acidic surface than its \textit{Pseudomonas} equivalent, indicating species-specific surface properties.

The PaPilP\textsuperscript{C} structure was also analysed using ConSurf\textsuperscript{213} to determine the localization of conserved residues on the protein surface and this is shown in \textit{figure 5.6.8}. There appeared to be no obvious concentration of conservation on the surface, which may have been indicative of an interaction site for another protein. Rather, conserved residues and variable regions appear to be distributed across the surface of the protein. Taken together with the divergent surface properties, the lack of localized surface conservation may go some way to explaining the observation by Tammam \textit{et al} of non-interchangeability of PilP proteins belonging to different organisms\textsuperscript{142}. 
Figure 5.6.5: Structural features of PaPiIpC

a. Three highly conserved glycine residues (green) are found within the loop connecting the two beta-sheets in the structure. These may play an important role in the structural integrity of the protein. Image produced using CCP4MG.

b. A dimer interface is formed between the first beta-strand adjacent, symmetry-related PaPiIpC chains in the C2 crystal form. This could provide a clue to the molecular basis of its interaction with the secretin PilQ. Image produced using CCP4MG.

Figure 5.6.6: Tryptophan residue exposed on the surface of PaPiIpC and NmPiIp

The surface exposed tryptophan in PaPiIpC, shown in magenta on the left, forms an interaction with the disordered N-terminus, pinning it to the β6-β7 hairpin. This tryptophan is also exposed in PilP from Neisseria, shown in green on the right, but it does not form the same interaction with the N-terminus in the solution state NMR structure. Image produced using CCP4MG.
Chapter 5

Figure 5.6.7: Electrostatic surface of PaPilP\textsuperscript{C} and NmPilP
The computed electrostatic surface of PaPilP\textsuperscript{C} has a fairly even distribution of charge with no obvious patches of localized acidic or basic residues. In comparison, the PilP homologue from \textit{Neisseria} has an overall more acidic surface. \textit{Calculated using CCP4MG}.

Figure 5.6.8: Consurf analysis of PaPilP\textsuperscript{C}
Surface conservation of PaPilP\textsuperscript{C} viewed from two orientations. The most highly conserved surface residues are coloured in purple, those with intermediate conservation in white, and those most variable across the family are shown in cyan. \textit{Figures produced using CHIMERA}.
5.7 Discussion

- Comparison to NmPilP NMR structure (2IVW)

A comparison of the PaPilP\textsuperscript{C} domain with the NmPilP C-domain NMR structure indicated some minor differences between the two, which may account for the problems in phase determination by molecular replacement. Figure 5.7.1 shows a representation of the two structures superimposed. The closest to mean structure from the Neisseria PilP ensemble (PDB ID 2IVW) was truncated to include only residues Asn 76 to Asn 164, which align directly with those which were successfully built into the electron density in the PaPilP\textsuperscript{C} crystal structure determined here. This superimposes onto the PaPilP\textsuperscript{C} crystal structure with an RMSD score of 2.4 Å over 70 residues. However, from the full Neisseria PilP NMR ensemble, the structure with least deviation from PaPilP\textsuperscript{C} superimposes onto PaPilP\textsuperscript{C} P2\textsubscript{1} chain A with an RMSD score of 2.04 Å over 73 residues, and this is shown as a C-alpha trace in figure 5.7.1a. The N-terminus of PaPilP\textsuperscript{C} (Lys84, Pro85) folds down to make contact with the β6-β7 hairpin loop, which is also more twisted than in the NMR structural ensemble. This can be seen in figure 5.7.1b, as a ribbon plot showing the closest to mean NmPilP (cyan) and PaPilP\textsuperscript{C} (P2\textsubscript{1} form, chain A; magenta). The most significant structural deviation, however, is the β1-β2 loop, which is much flatter in the crystal structure (figure 5.7.1c). This was the case for all four PaPilP\textsuperscript{C} copies in the C2 crystal asymmetric unit, and for both copies in the P2\textsubscript{1} crystal asymmetric unit. One reason for this observation could arise from crystal packing; it is often noted that slight conformational differences can be adopted in crystal structures compared to those determined from solution state NMR.

- Comparison to PaPilP NMR structure (2LC4)

Recently another group published an NMR structure for Pseudomonas PilP (PDB ID 2LC4), which superimposes onto the PaPilP\textsuperscript{C} crystal structure with an RMSD of 1.78 Å over 79 residues, with no major structural deviations figure 5.7.2. A similar Van der Waals' interaction to that noted in the recently published NMR structure\textsuperscript{142} of PilP from Pseudomonas\textsuperscript{142} can also be seen in the crystal structure between residues Pro85 and Trp161, and this was not noted in the previous structure from

185
Neisseria. This group also showed that N. meningitidis PilP cannot restore T4PB function in a P. aeruginosa pilP mutant,\textsuperscript{142} which is unsurprising given the difference in sequence between the two homologues. It appears that the while the PilP structure, and probably its function, has remained conserved, the individual properties of PilP homologues may have evolved to suit species-specific environments.

When Tammam et al reported the NMR structure determination for Pseudomonas PilP, they expressed that they had experienced difficulty in obtaining diffraction quality crystals. In a similar approach to that adopted by Golovanov et al,\textsuperscript{113} they pursued NMR structure determination with the smallest stable proteolytic fragment which, when sequenced, was found to be missing the first 71 residues of the mature protein. The construct used in this project was carefully designed to omit the first 82 residues, based on an analysis of the folded domain boundaries in Neisseria, combined with secondary structure prediction algorithms generated by the PHYRE server\textsuperscript{206} and sequence alignments with PilP orthologues. This recombinant PaPilPC crystallized relatively easily, which highlights how crucial construct design is for crystallography. Together with the results presented for PilQ in chapter 3, it reiterates that judicious use of bioinformatics tools can be valuable for streamlining structural projects.
Figure 5.7.1: Comparison of NmPilP and PaPilP<sup>C</sup>

a: C-alpha backbone trace showing the NmPilP NMR ensemble (2IVW; cyan) superimposed onto the PaPilP<sup>C</sup> crystal structure (P2<sub>1</sub> chain A; magenta). Arrows indicate the β1-β2 and β6-β7 hairpins.

b: Ribbon trace showing closest to mean NmPilP structure (2IVW; cyan) superimposed onto PaPilP<sup>C</sup> (P2<sub>1</sub> chain A; magenta). An arrow indicates the disordered N-terminus of PaPilP<sup>C</sup> which folds down to make contact with the β6-β7 hairpin.

c: Ribbon representation of the closest to mean NmPilP structure (2IVW; cyan) superimposed onto PaPilP<sup>C</sup> (P2<sub>1</sub> chain A; magenta). The β1-β2 hairpin is indicated by an arrow, and is much flatter in the Pseudomonas crystal structure. *Images produced using CCP4MG*

---

Figure 5.7.2: Comparison of Pseudomonas PilP NMR structure and PaPilP<sup>C</sup>

The closest to mean representative of the *Pseudomonas* PilP NMR structure (2LC4) (green) is shown superimposed onto the PaPilP<sup>C</sup> crystal structure (yellow). The structures are very similar with no obvious deviations. *Image produced using CCP4MG*
- Comparison to GspC of the T2SS

The PaPilP\textsuperscript{G} structure presented here, like the NMR structure, bears a striking similarity to the HR domain of GspC. Figure 5.7.3 shows the structure superimposed onto the co-crystal complex of the GspC HR-domain and GspD N0-domain\textsuperscript{152} (PDB ID 30SS). The two structures overlay with an RMSD score of 1.65 Å over 58 residues. This is closer than the difference between the Pseudomonas PilP NMR structure\textsuperscript{142} and GspC\textsuperscript{152} which can be superimposed with an RMSD of 2.36 Å. This perhaps highlights the differences between rigid crystal structures and proteins solved in solution by NMR. The main deviation again lies within the β1-β2 hairpin (figure 5.7.3, left hand panel), which in the GspC structure is seen in complex with GspD. It is possible that this structural alteration occurs upon complex formation, or may arise from crystal packing in the apo-PaPilP\textsuperscript{G} crystal. The only other real difference is the lack of the shorter strand β4 in the GspC structure, and there a few minor deviations in strand length (figure 5.7.3, right hand panel). Another GspC-type NMR structure, that of OutC from Erwinia species, does also not form the fourth beta-strand, yet it is present in other family members so may be dynamic in solution.\textsuperscript{151} It would therefore appear that, despite some subtle differences, PilP has a structurally analogous T2SS component in the form of a domain from a larger protein with which it lacks sequence identity. In particular it was noted earlier that the three highly-conserved glycines found in PilP are not found in the GspC proteins (figure 5.7.4). This is surprising if these glycines are playing a key role in maintaining the structural integrity of PilP, as it is so structurally similar to GspC-HR. It is possible that they serve another purpose, or their role is fulfilled in a different way in GspC and they are not required. It should also be noted here that chapter 6 presents a detailed description of the interaction between PilP and the PilQ secretin in Neisseria, which shows a similar interaction to that of GspC and GspD, indicating that the two proteins may also share functional parallels.
Figure 5.7.3: Comparison of PaPilP\(^C\) and GspC-HR

PaPilP\(^C\) (P2; chain A; magenta) is shown superimposed onto the crystal structure for GspC-HR (yellow) which is in complex with GspD N0-domain (white) (PDB ID 30SS). On the left hand panel, an arrow indicates the different positions of the β1-β2 hairpin which, in the case of GspC-HR, forms the interaction with the secretin GspD. The right hand panel indicates the short strand β4 of PiIP, which is not seen in the GspC-HR structure shown here.

Figure 5.7.4: Alignment of PiIP\(^C\) and GspC-HR sequences

The sequences for the folded domains of Neisseria and Pseudomonas PiIP are shown aligned with the sequences for the “homology region” (HR) domains of XcpP from Pseudomonas, GspC from ETEC and OutC from Dickeya. Absolutely conserved residues in PiIP are highlighted in yellow. Where these are also found in GspC proteins they are highlighted in cyan. Residues with conserved properties in both proteins are highlighted in red, and these mostly form the hydrophobic core. There is low sequence similarity between GspC-HR and PiIP despite similarities in structure.
- Conclusions and future direction

The availability of reproducible, diffraction quality crystals of PaPilP will also be useful for further studies. For example, some attempts were made to identify small molecules (fragments), by soaking, which could bind to a putative site for PilQ recognition. Using the VINA software, ligands were designed according to best theoretical fit, and a library of related compounds used to soak crystals of PaPilP, which were analysed at the synchrotron (data not shown). These preliminary attempts were not successful, as the resulting difference maps showed little density for most ligands, and this line of enquiry was abandoned. The availability of more detailed information on PilP-PilQ could suggest new avenues in this investigation, however.

An interesting feature of PilP, which has not been investigated here, is its disordered N-terminus, which was omitted from this structural study for the purposes of crystallization. As PilP has been previously shown to interact with the secretin PilQ and is also demonstrated in chapter 6 to form this interaction via its folded C-domain to the central portion of the secretin structure, it is likely that this N-terminus is anchored to the inner-membrane at one end, and reaches up into the periplasm to bind to the secretin at the other. However recent evidence has shown that it may do more than merely anchor the functional C-domain in place. Tammam et al demonstrated that the C-terminal domain alone is unable to co-purify with the inner membrane PilN/O complex, and also that this complex has stabilising effects on the proteolytically sensitive PilP N-terminus. Therefore the N-terminus may play an important role in the interaction of PilP with the inner-membrane components of the system.

In conclusion, this chapter has presented two high-resolution crystal structures for the C-domain of PilP from *Pseudomonas aeruginosa*. They indicate some interesting species-specific differences between PilP proteins from different organisms with regards to surface properties, although the structures remain very similar. It would be interesting to investigate how this extends to other organisms. The publication of the *Pseudomonas* PilP NMR structure during the course of this work is in agreement with these findings. Again in the case of the HR-domain of
GspC, a seemingly unrelated protein, there are striking parallels in tertiary structure. PilP is a good example of structural homology, and a crucial structure-function relationship, where seemingly different or unrelated proteins may provide a common function. This idea becomes more prominent in chapter 6 where the PilP/Q interaction is explicitly characterized.
Chapter 6: Establishing protein-protein interactions within the type IV pilus biogenesis system: PilP, PilQ and PilE

6.1 Introduction

Structural information for regions of PilQ presented in chapter 4 and the PaPilP<sup>C</sup> crystal structure presented in chapter 5, provided the foundation for a concerted structural programme to investigate the function of PilP, the PilQ secretin and their interactions with both each other and other components of the T4PB system. The production of manageable, folded domains from the PilQ secretin presented in chapter 3 has made feasible the possibility to explore these interactions biochemically in vitro.

Due to the complexity of multi-protein secretion systems, elucidating the location and roles of individual components has often relied on a reductionist approach to dissect the macromolecular machines. For example, early gene knockout studies<sup>82</sup> determined the vital components of the pilus biogenesis system, and since then a substantial effort has been made to establish protein-protein interactions within all secretion systems via both biochemical and structural approaches in order to build a comprehensive representation of how these systems operate as a whole. Various methods have been implemented to achieve this, perhaps the most recent example was the use of a battery of bacterial two-hybrid assays to identify binary interactions within the pilus system of *Neisseria*.<sup>241</sup> Also, we now have data emerging which details these important protein-protein interactions within multi-protein secretion systems at the atomic level. For example, as described in the main introduction, the inner membrane platform components of the T2SS have been characterized piece by piece. It was demonstrated that the EpsL and EpsM components from *Vibrio cholerae* would co-precipitate with antibodies against either component.<sup>99</sup> Following this, a co-crystal structure of the complex formed between EpsE and EpsL<sup>108</sup> emerged detailing the interaction interface, and subsequent work in *Pseudomonas* resulted in the isolation of a quaternary complex comprising the GspF,E,L,M and C proteins.<sup>105</sup>

GspC, which has a structurally similar domain to PilP of the T4PB, has been shown to interact with the secretin component, GspD,<sup>151,242</sup> which it requires for
correct localization into the T2SS. Very recently a detailed description of this interaction has emerged with the publication of the crystal structure of the binary GspC HR-domain and GspD N0-N1 domain complex.\textsuperscript{152}

More recently, work on the T4PB system of \textit{Pseudomonas aeruginosa} has similarly established the platform complex formed by PilM,N,O and P at the inner membrane.\textsuperscript{112} Within this complex, PilN and O were strongly dependent on each other for stability, and both also require the presence of PilM. The complex formed between PilM and PilN has also recently been characterized on the structural level, and this indicates that the two form a very tightly bound complex.\textsuperscript{115} Loss of PilP, however, did not affect the rest of the complex.\textsuperscript{112}

The findings thus far, therefore, seem to suggest that GspC and PilP may provide the link between the outer and inner membrane components of the system, and share some functional analogy. Interestingly, during the course of this work, Tammam \textit{et al} discovered that the disordered N-terminal region of PilP was likely to be responsible for binding to the inner membrane platform.\textsuperscript{142} Taken together with recent findings on the GspC/D complex, and the findings described within this chapter, we become closer to a more complete picture of an envelope-spanning complex in both systems.

The work in part A of this chapter explores at the molecular level the interaction formed between the PilP C-terminal domain and individual domains from the secretin PilQ. Furthermore, the segmental nature of PilQ raises intriguing questions as to the possibility of individual functions for these domains, possibly related to gating or substrate recognition. In part B of this chapter, the major subunit of the pilus appendage, PilE, was produced and purified with the aim of conducting NMR-CSP experiments with PilQ domains. Previous work has shown that the pilus subunit interacts with N- and C-terminal regions of PilQ, and that assembled pilin oligomers are able to fill the inside of the PilQ chamber.\textsuperscript{214} In light of this, successful identification of interacting residues on the surface of B2 and N0-domain solution structures (chapter 4) would help to orientate the inner and outer faces of these domains in an oligomeric PilQ model.
Part A: Analysis of interaction between NmPilQ\textsuperscript{343-545} and NmPilP\textsuperscript{C}

6.2 Nuclear Magnetic Resonance chemical shift perturbation (NMR-CSP)

Nuclear magnetic resonance is a powerful tool to study protein-protein interactions, particularly weak but biologically relevant interactions, at the atomic level. Chemical shifts for \textsuperscript{1}H-\textsuperscript{15}N amides within different residues are sensitive to the local environment and are therefore also sensitive to changes during protein-protein complex formation. If the respective \textsuperscript{1}H\textsuperscript{15}N-spectra can be assigned, this allows specific residues at the binding interface to be mapped. This is particularly useful for weak protein-protein interactions ($K_D$ values in the μM to mM range), where chemical shift changes are small and less likely to be due to conformational rearrangement of the protein backbone. Instead, these perturbations in the spectra are more likely to be attributed to direct participation of that residue in complex formation.

Solution state NMR-CSP was exploited to characterise the interaction between PilP and periplasmic domains from the secretin, PilQ. First, NmPilQ\textsuperscript{343-545}, a construct incorporating the N0-domain structure determined in chapter 4 and the second N1 or “KH-type” domain, was titrated into \textsuperscript{15}N-NmPilP\textsuperscript{C} and specific changes to the crosspeaks in the \textsuperscript{15}N-HSQC spectra were monitored. Figure 6.2.1 shows an overlay of the \textsuperscript{15}N-HSQC spectra obtained from each titration point. Each peak in the spectrum is coloured according to the ratio of PilP: PilQ used in the titration. Specific attenuation to chemical shifts was observed for particular resonances, which were identified from the previous assignment of NmPilP from Golovanov et al.\textsuperscript{113} The HSQC peaks were analysed to produce shift lists indicating the titration point at which the crosspeak broadened, given as a ratio of PilP: PilQ. This information is presented as a histogram in figure 6.2.2a. It is evident that interacting residues map largely to the first and second β-strands of the NmPilP structure,\textsuperscript{113} which form a hairpin extending from the globular fold. Severely affected residues in the \textsuperscript{15}N-HSQC form a well-defined patch when mapped upon the surface of the three-dimensional structure, which suggests a specific recognition site for NmPilQ\textsuperscript{343-545} on the surface of NmPilP\textsuperscript{C} (figure 6.2.2b).
Although this result verified that an interaction between NmPilQ<sub>343-545</sub> and NmPilP<sup>C</sup> was occurring; it was not possible to quantify in terms of a $K_D$ value because although specific signals broadened out from the apo-NmPilP<sup>C</sup> HSQC spectrum, these failed to reappear at alternative chemical shifts at the end of the titration.

**Figure 6.2.1: Titration of unlabelled NmPilQ<sub>343-545</sub> into <sup>15</sup>N-NmPilP<sup>C</sup>**

An overlay of <sup>15</sup>N-HSQC spectra is shown, and spectra are coloured individually according to the key on the left. Spectra were recorded from a sample containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>; 50 mM NaCl; 10% D<sub>2</sub>O; and protein components <sup>15</sup>N-NmPilP<sup>C</sup> and NmPilQ<sub>343-545</sub> according to the relative concentrations indicated in the key. Spectra were processed using Topspin2.1 (Bruker) and the Azara processing package provided as part of the CCPN suite using the previous assignment for *Neisseria meningitidis* PilP.<sup>113</sup> Figure was produced using CCPN Analysis.
Figure 6.2.2: Identification of the NmPilQ<sub>343-545</sub> binding site on NmPilP<sub>C</sub>

a. Graphical representation of peak attenuation in the <sup>15</sup>N-NmPilP<sub>C</sub> HSQC spectra upon titration of unlabelled NmPilQ<sub>343-545</sub> into the sample. The amino acid sequence is given below where the letter ‘s’ denotes an amide sidechain signal, and the position of secondary structure elements within the sequence is shown above the graph.

b. Surface representation of NmPilP<sub>C</sub> with residues corresponding to attenuated peaks coloured according to the first four titration points, shown in the key on the left. Residues involved in binding form a distinct patch on the surface of the structure, with Lys 102 (red) extending from the β1-β2 hairpin and integral to the binding site. *Figure produced using CCP4MG.*
The interaction was then explored in greater depth by conducting the reverse experiment with unlabelled NmPilPC titrated into a $^{15}\text{N}^{13}\text{C}$ double-labelled sample of NmPilQ$^{343-545}$. This titration experiment was also carried out using $^{15}\text{N}^{13}\text{C}$-PilQ$^{343-442}$, which incorporates the N0-domain only (the structure was presented in section 4.4). This time, changes to both $^{15}\text{N}$- and $^{13}\text{C}$-HSQCs were monitored for a more explicit analysis of the interaction. Like the reverse experiment, it became evident that while some crosspeaks remained unaffected by titration of unlabelled NmPilPC into the sample, specific line broadening could be seen in others. This information from both the $^{15}\text{N}$- and $^{13}\text{C}$-HSQCs was used, combined with the chemical shift assignments determined for NmPilQ$^{343-545}$ in section 4.2, to identify the interacting residues. Shift lists were produced indicating the titration point, given as a ratio of PilQ: PilP, at which the crosspeak broadened. Similar to the PilQ into PilP titration, specific peak broadening was observed and again, these crosspeaks failed to reappear at alternative chemical shifts, which would represent the bound form. It should also be noted here that in this experiment there were some residues for which the assigned crosspeak disappeared, and reappeared at the same chemical shift value. This was taken as a limitation of the experiment and only the highest titration point at which such crosspeaks disappeared permanently - i.e. they were not present at consecutive titration points - was used as a data point. Repetition of the experiment with the shorter NmPilQ$^{343-442}$ fragment was useful to verify which residues were truly affected by NmPilPC. From the shift lists, histograms summarising the data obtained are presented in figure 6.2.3, which shows a comparison between NmPilQ$^{343-545}$ and NmPilQ$^{343-442}$. Only residues for which an assignment was available are presented in the figure. The second experiment using the N0-domain alone, NmPilQ$^{343-442}$, gave similar results to those seen with the larger fragment, NmPilQ$^{343-545}$ (figure 6.2.3). It is readily apparent that it is the N0-domain, rather than the N1-domain, which is involved in recognition of NmPilPC. Figure 6.2.3b shows the severely affected residues mapped onto the surface of the N0-domain structure. The largest chemical shift changes mapped to one side of the protein, concentrated around the first $\alpha$-helix and $\beta$-strand in the fold, with residues also affected on the third $\beta$-strand and second $\alpha$-helix.
Figure 6.2.3: Identification of the NmPilP\textsuperscript{C} binding site on NmPilQ\textsuperscript{343-545} and NmPilQ\textsuperscript{343-442}

a: Graphical representation of peak attenuation in the NmPilQ\textsuperscript{343-545} and NmPilQ\textsuperscript{343-442} 15N- and 13C-HSQC spectra upon titration of unlabelled NmPilP\textsuperscript{C} into each sample. Bar height represents the titration point, given as a ratio of NmPilQ\textsuperscript{343-545} or NmPilQ\textsuperscript{343-442} to NmPilP\textsuperscript{C}, at which the 15N-crosspeaks attenuate. Negative bars represent unassigned residues. Bar colour represents the point at which respective 13C-crosspeaks attenuate. 13C-HSQCs were recorded every third titration point only.

b. Surface representation of NmPilQ\textsuperscript{343-442} with residues coloured according to the titration point at which they are affected, as shown in the key on the left. All ratios are given as NmPilQ\textsuperscript{343-545} to NmPilP\textsuperscript{C}. Therefore those most severely affected attenuate when there is a greater excess of NmPilQ\textsuperscript{343-545} in the sample.

Figure produced using CCP4MG.
6.3 Structural model for the complex formed by \( \text{NmPilP}^C \)/ \( \text{NmPilQ}^{343-545} \)

Attempts to co-crystallize \( \text{NmPilP}^C \) and \( \text{NmPilQ}^{343-545} \) were unsuccessful during the timeframe of this project. Therefore, using the previously determined \( \text{NmPilP} \) NMR structure by Golovanov et al. and the newly determined \( \text{NmPilQ} \) N0-domain structure, a docked complex model was produced. HADDOCK is a popular docking program which uses information from predicted or identified protein-protein interaction interfaces to drive the modelling of bimolecular complexes. It makes use of CNS as its structure calculation engine, and requires both a Unix platform and many CPU hours, as well as technical expertise. The HADDOCK server, however, is a web interface accessible to the broader community and provides the full set of parameters available on the program. Protein complex interfaces are determined by entering data as a set of active and passive residues; the former of which are directly involved in complex binding as determined by experimental evidence, and the latter are defined as those which may contribute to the interaction. Passive residues may be specified by the user or automatically defined based on those surrounding active residues. HADDOCK then generates a set of ambiguous restraints for the docking process, between each active residue and all active and passive residues on the interacting partner. These are satisfied during docking based on a distance of within 3 Å between interacting residues.

An advantage of HADDOCK docking is that it allows for some degree of movement in the structures upon complex formation. To save time, this is often restricted to designated semi-flexible segments, incorporating the residues involved in the interaction, but this can be defined manually by the user if other information is available. After an initial rigid body refinement, a semi-flexible refinement in torsion angle space is conducted based on analysis of the intermolecular contacts found in the rigid body stage.

The detailed information made available from NMR-CSP titration experiments in this project was used for this purpose. Active and passive residues were defined as those severely affected during the titration and those much lesser so respectively. This highlights a limitation to the HADDOCK process, as there is a considerable reliance on the user’s own interpretation of the data. For example, particularly with NMR-CSP data, one should be cautious that attenuation of a signal is not due to
global changes in the protein conformation or movement due to binding on another area of the protein. From the data presented in section 6.2, it was evident that the N0-domain of PilQ was facilitating the interaction with NmPilP\textsuperscript{C}. There was some attenuation of crosspeaks within the N1-domain region during the NmPilQ\textsuperscript{343-545} titration, but as mentioned these are more likely to be changing due to changes in the environment of these residues, caused by a remote conformational change away from the binding site rather than direct binding, as the majority of affected residues clustered to one face on the N0-domain. Therefore only the data from the N0-domain titration, NmPilQ\textsuperscript{343-442}, was used to define active and passive residues. Additionally, this data was used because it was of better quality, with a more complete assignment of the HSQCs possibly due to omission of unstructured regions within the NmPilQ\textsuperscript{3430545} construct, which somewhat precluded the spectra. Referring back to figure 6.2.3b, active residues were selected as those affected whilst still in a large excess of NmPilQ\textsuperscript{343-545}, i.e. those affected when little NmPilP\textsuperscript{C} was added, namely the early titration points of 10:1 to 3.33:1. Some residues where the amide peak attenuated at later titration points, but the methyl peak attenuated early were also included. Passive residues were selected as those affected at the 2:1 titration point. Those affected at 1:1 were not entered, to make the process more stringent. Entering data for the NmPilP\textsuperscript{C} interface was more straightforward as only one titration was conducted, so again those residues most affected were defined as active and those lesser so as passive. Semi-flexible segments were allowed to be defined automatically, and no distance or hydrogen bond restraints were customized.

This initial run produced a plethora of possible structure generations which were clustered by HADDOCK into five best solutions with a 7.5 Å cut-off. The best representative of each of each solution is presented in figure 6.3.1a. Each is given a HADDOCK score (HS), which is a weighted sum of electrostatic, Van der Waals’, desolvation and restraints violation energies,\textsuperscript{153} which may arise from a distance restraint not being satisfied, for example. The lowest energy structure is considered the best solution. In this case, the two best solutions came from a parallel β-strand pairing of strand β1 from both molecules. This seems unusual as a parallel arrangement of a small number of strands is generally less stable than
the antiparallel conformation. In addition, during the course of these experiments the co-crystal structure of the T2SS secretin GspD N0-domain in complex with the HR-domain from GspC was published, illustrating an antiparallel \( \beta 1-\beta 1 \) interaction between the two proteins. The extensive similarities drawn so far between GspC/D and PilP/Q led us to believe this interaction might occur in a similar way. The third and fourth best solutions did give a \( \beta 1-\beta 1 \) antiparallel arrangement at the binding interface, however neither were within hydrogen bonding distance so also seemed unlikely to be entirely correct. These clusters do, however, show an arrangement that is not wholly dissimilar from that seen in the GspC/D co-crystal complex. The fifth cluster showed an antiparallel arrangement of strand \( \beta 3 \) of the N0-domain with strand \( \beta 1 \) of NmPilPC. Interestingly a similar experiment was recently published also using NMR-CSP to define the T2SS OutC/D complex. This paper also found \( \beta 3 \), as well as part of \( \alpha 2 \), of the N0-domain to form a complex with strand \( \beta 1 \) of the HR-domain, contrary to the GspC/D co-crystal.

It became evident from the initial HADDOCK run, that there were five equally viable options all very similar in energy yet quite different in the arrangement of the complex. In addition, the side-chain residue Lys 102 of NmPilPC which is integral to the binding interface has lots of positions in the representative ensemble therefore was interfering with the protein-protein interface, and probably pushed out during the rigid body docking stage. CNS 1.2 on the other hand, uses intramolecular distance and hydrogen bond restraints as an input, rather than PDB co-ordinates. With this approach, Lys 102 has only a backbone hydrogen bonding restraint between its amide and residue 108. Therefore, attempts were made to improve the model also based on homology modelling with the GspC/D co-crystal structure, using CNS1.2. The crystallized complex (PDB ID 3OSS) was used to guide the identification of five additional intermolecular hydrogen bond restraints based on the argument that the homologous model structure does fit the chemical shift mapping. This produced a model based on the GspC/D crystal complex whereby the PilQ N0-domain and PilP recognize each other via an antiparallel beta-strand pairing of their first beta-strands, and this is shown in figure 6.3.1b. It must be reiterated as this is a homology model, and residues on the third \( \beta \)-strand
and helices of the N0-domain were affected, that a more complex or dynamic interaction may occur in vivo, particularly considering results from the OutD/C study by similar methods.\textsuperscript{151}

\textbf{a.}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{example.png}
\caption{Structural model for the complex formed by NmPilP\textsuperscript{C} / NmPilQ\textsuperscript{343-545}}
\end{figure}

\textbf{a.} Best representative models from the top five solutions deduced and clustered by HADDOCK. NmPilP\textsuperscript{C} is shown in bronze and NmPilQ\textsuperscript{343-442} is shown in blue.

\textbf{b.} Final model for the complex produced by CNS 1.2, coloured as in (a). \textit{Figures produced using CCP4MG.}
6.4 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was also used in an attempt to quantify the interaction between NmPilP$^{C}$ and NmPilQ$^{343-545}$. For these experiments the proteins were purified as for the NMR experiments, and transferred into 50 mM NaH$_2$PO$_4$/Na$_2$HPO$_4$ pH 7.0, 100 mM NaCl. A 1 mM stock solution of NmPilP$^{C}$ was sequentially injected in 1 µl increments into a 100 µM solution of NmPilQ$^{343-545}$ in the well chamber. Plots for the recorded enthalpy change versus time are shown in figure 6.4.1. The enthalpy changes decreased but failed to saturate after 20 injections, at which point precipitation was observed in the sample, causing an increasingly erroneous baseline and resulting in a partial curve (figure 6.4.1, left panel). The experiment was attempted a second time with double the volume of NmPilP$^{C}$ per injection and, again, enthalpy changes were observed and these failed to saturate before the solubility of one or both proteins was affected (figure 6.4.1, right panel). These enthalpy changes were, however, not observed in control experiments when NmPilP$^{C}$ was injected into buffer containing no protein, and also when buffer was injected into the NmPilQ$^{343-545}$ well solution (data not shown). This indicates that the observed peaks can be attributed to a protein-protein interaction, although it was not possible to quantify an accurate equilibrium constant, as a sufficient level of saturation was not reached in both titrations. Further attempts were made using the alternative PilQ construct, NmPilQ$^{343-442}$, in much lower concentrations and similar results were obtained (data not shown), and it was not possible to obtain a full data set and determine an accurate dissociation constant. Instability of the PilP/Q complexes formed throughout this project was a common observation. Precipitate was often visible in the Shigemi tubes used during NMR titrations when samples were mixed. The $K_D$ values obtained from fitting the partial curve are therefore subject to considerable error. The ITC results do, however, provide additional confirmation of an interaction between NmPilP$^{C}$ and the N0/N1-domains from PilQ.
Figure 6.4.1 ITC experiments with NmPilP\(^C\) and NmPilQ\(^{343-545}\)

The left hand panel shows the enthalpy change versus time plot obtained when 1 mM NmPilP\(^C\) was injected into 100 \(\mu\)M NmPilQ\(^{343-545}\). The right hand panel shows the repeat experiment when the injection volume of NmPilP\(^C\) was doubled to 2 \(\mu\)l. Data processed using Origin\(^{®}\) 7.0. 14.2
6.5 Surface Plasmon Resonance

Surface Plasmon resonance (SPR) was explored as an alternative biophysical method to detect and quantify the interaction between NmPilP\textsubscript{C} and NmPilQ\textsubscript{343-545}. The approach involves attaching one interacting partner, or ligand, to a sensor chip whilst the analyte is passed over the surface. The response generated upon binding of the analyte to the ligand is proportional to the bound mass. This method is highly sensitive and requires much lower concentrations of protein than ITC, which may circumvent the protein solubility problems encountered during ITC titrations and enable the dissociation constant to be determined more accurately for the interaction.

To verify an interaction between NmPilP\textsubscript{C} and NmPilQ\textsubscript{343-545} SPR experiments were conducted using the ProteOn\textsuperscript{TM} XPR system (BioRad). 2500 response units of NmPilP\textsubscript{C} were immobilized to a ProteOn\textsuperscript{TM} GLC sensor chip by amine coupling via lysine residues. Figure 6.5.1 shows the results of SPR with 10-60 nM NmPilQ\textsubscript{343-545} passed over the surface of the sensor chip. Interactions were determined in phosphate buffered saline at pH 7.4. There was a low overall response - reaching only 60 RU - considering the large amount of protein on the chip, indicating that most molecules are immobilized in an unfavourable orientation. This is comprehensible as NMR chemical shift perturbation studies determined that lysine residues 102 and 106 predominantly occupy the binding site of NmPilP\textsubscript{C}. There is only one other lysine at position 158 that is not involved in the interaction with NmPilQ\textsubscript{343-545}. Therefore most NmPilP\textsubscript{C} molecules may be coupled with the binding site facing the chip. Additionally, the data fits best to a bivalent analyte model, with a Chi\textsuperscript{2} value of 1.6; this was apparent upon repeated attempts with both NmPilQ\textsubscript{343-545} and NmPilQ\textsubscript{343-442} (not shown). The apparent dissociation constant of the interaction (\(K_D\)) is given as 268 nM, and this is a low value owing to the apparent slow off-rate of the reaction. Remaining bound material could only be removed using 50 mM NaOH as a regeneration condition.

In the light of issues encountered during lysine coupling, further attempts were made to improve the SPR data, with a number of changes made to the experiments. This time, a Biacore system (GE Healthcare) was used. Firstly, the
amount of NmPilP\textsuperscript{C} coupled to the chip was reduced according to Biacore’s recommendations. The binding capacity of material immobilized on the surface of the chip is related to the mass of material on the surface and therefore the molecular weights of the ligand and analyte by the following equation: analyte binding capacity = (analyteMW/ ligandMW) x immobilized ligand level. Assuming activity, or percentage of successful orientations, is 50-75\%, around only 60 RU of immobilized PilP would be required - considerably less than the amount used in the first attempt.

To address the issue of the unfavourable orientation of NmPilP\textsuperscript{C} molecules by amine coupling, an NTA-coated chip was tried instead. For these experiments a longer recombinant PilP fragment, NmPilP\textsuperscript{20-181}, complete with its histidine purification tag was used for attachment to the NTA-chip through nickel chelation. This PilP construct was previously used in another study from our group, and the protein was produced and purified in the same manner.\textsuperscript{113} The rationale behind this was that the presence of the longer, unstructured N-terminus may act as a tether allowing more of the surface of the folded domain to be exposed to the analyte. Also, the intention was to orientate the NmPilP\textsuperscript{20-181} molecules onto the chip via the His-tag, rather than lysine residues, to avoid masking of the binding site. Thrombin-cleaved NmPilQ\textsuperscript{343-545} was again the analyte constituent which was passed over the surface of the chip. However this presented another issue, as during the experiment the NmPilP\textsuperscript{20-181} molecules appeared to gradually dissociate from the sensor chip surface, which made binding difficult to characterize. In addition, there was considerable interaction between NmPilQ\textsuperscript{343-545} and the reference cell; this was likely to be via the dextran chip surface as the purification tag had been fully cleaved. To address the dissociation of NmPilP\textsuperscript{343-545}, covalent amine coupling was used after molecules were bound via the His-tag. This should orientate the molecules onto the surface via the His-tag, and permanently fix them into place. This method eradicated ligand dissociation from the surface. The issue of an interaction between PilQ\textsuperscript{343-545} and the reference cell was, however, still problematic. Several approaches were taken to try to reduce the level of interaction. These included a range of additives, such as low levels of detergents, dextran and increased salt, in both the running buffer and the protein sample. In
addition, prior to ligand capture, the surface of the chip was treated with several rounds of EDC/NHS and ethanolamine, anticipating that fewer active groups on the surface would reduce the reference cell interaction, but this also failed to reduce the non-specific interaction. The best data that could be obtained from these experiments therefore, was that using the ProteOn™ system, shown in figure 6.5.1. This provides confirmation of an interaction between the proteins, but raises issues regarding the stoichiometry and quantification of the interaction that could not be resolved after several attempts.

![Figure 6.5.1 Analysis of the interaction between NmPiiPC and NmPiiQ343-545 by SPR](image)

**Figure 6.5.1 Analysis of the interaction between NmPiiPC and NmPiiQ343-545 by SPR**

2500 response units of NmPiiPC were immobilized to a PRoteon™ GLC sensor chip by amine coupling via lysine residues. 10-60 nM NmPiiQ343-545 was then passed over the surface. Running buffer used was PBS. Data was analysed and fit to a bivalent model by the ProteOn Manager™ software which was also used to produce the figure.
6.6 Probing an interaction between NmPilP\textsuperscript{C} and the PilQ B2-domain

The next stage of the project aimed to determine whether NmPilP\textsuperscript{C} also formed an interaction with any other parts of the secretin. The work presented in chapter 3 showed that the PilQ B2-domain, (NmPilQ\textsuperscript{224-329}), could be produced as a soluble folded fragment in high enough concentrations suitable for NMR. After structure determination of this fragment (chapter 4), it was logical to proceed to investigate its function. The first stage was to see if an interaction between this portion of the secretin and NmPilP\textsuperscript{C} could be established: this was investigated by NMR-CSP. The results are given in figure 6.6.1, which shows the $^{15}$N-$^1$H-HSQC for 100 μM $^{15}$N-labelled NmPilQ\textsuperscript{224-329} before and after addition of a 2-fold excess of unlabelled NmPilP\textsuperscript{C}. There was no apparent change to the spectra, and therefore no evidence of binding at this concentration in the absence of other PilQ domains. This may reflect either a lack of interaction or requirement for other domains. As NMR is a very sensitive tool to investigate protein-protein interactions, this line of enquiry was not followed any further.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.6.1.png}
\caption{Analysis of NmPilP\textsuperscript{C} and NmPilQ\textsuperscript{224-329} by NMR-CSP}
\end{figure}

Natural isotopic abundance NmPilP\textsuperscript{C} was titrated into a $^{15}$N-NmPilQ\textsuperscript{224-329} and changes to the crosspeaks were monitored. The $^{15}$N-HSQC for NmPilQ\textsuperscript{224-329} after addition of a two-fold molar excess of NmPilP\textsuperscript{C} is shown (red) overlaid onto that for apo-$^{15}$N-NmPilQ\textsuperscript{224-329} (teal). 

\textit{Figure produced using CCPN Analysis.}
Part B: Investigation of an interaction between PilQ and the pilin subunit, PilE by NMR

During the course of this project, evidence emerged of an interaction between the Vibrio cholerae T2SS secretin GspD and both its cholera toxin substrate, and the T2SS pseudopilus tip complex. This interaction was observed using a fragment of GspD incorporating the N0, N1, N2, and N3 domains. In addition, N-terminal regions of PilQ have been shown to bind its secreted substrate, the pilus subunit. Therefore, it was necessary to investigate whether the NmPilQ (N0/N1) fragment produced in this study formed a similar interaction with PilE, the major pilin subunit.

6.7 Production and purification of the pilus subunit, PilE

For this experiment a soluble, monomeric sample of NmPilE was required for titration into NmPilQ samples. As shown in previous studies a construct incorporating only the globular domain of PilE, minus the initial α-helix which is required for multimerization, can be produced as soluble protein. Bioinformatics tools were used to align PilE sequences and predict secondary structure content. It was likely that the prediction of the globular domain boundary would be straightforward and accurate, due to previous publication of crystal structures of pilin proteins of Neisseria gonorrhoeae and Pseudomonas aeruginosa. The DNA encoding the required fragment (residues 36-170) was purchased from GeneArt (Invitrogen) and manufactured to incorporate BamHI and NdeI restriction sites for cloning into the pET15b vector to incorporate an N-terminal histidine purification tag. The fragment was digested and ligated into the vector as described in materials and methods (2.3). The construct NmPilE was then transformed into E. coli Origami B (DE3) cells which contain mutations promoting disulphide bond formation in the cytoplasm.

Figure 6.7.1a shows a western blot produced from small-scale protein production trials of NmPilE with 0.1, 0.5 and 1 mM IPTG for 3 hours at 37°C. Overall there is not a great variation using any concentration, but there appears to be marginally less protein produced using 0.1 mM IPTG than the higher
concentrations. There is no visible difference between 0.5 and 1 mM IPTG from the western blot, so 0.5 mM IPTG was used for induction.

During large-scale protein production, the growth of Origami (DE3) cells was very slow, taking double the usual required time, compared with BL21 (DE3) cells, to reach mid-log phase. Therefore cultures were induced at 20°C overnight for 14-16 hours. The protein was purified using the usual methods (2.3) by IMAC with 50 mM NaH₂PO₄/Na₂HPO₄ pH 7.8, 100 mM NaCl plus lysozyme and DNase I as lysis buffer, and elution with the same buffer plus 200 mM imidazole. Fractions from the purification process were analysed by SDS-PAGE and are shown in figure 6.7.1b. Fractions containing the eluted protein were pooled and exchanged into 50 mM NaH₂PO₄/Na₂HPO₄ pH 6.8, 100 mM NaCl for NMR.

Due to the presence of considerable impurities in the IMAC eluate, the protein was purified further by gel filtration. The resulting chromatogram is given in figure 6.7.2. PilE20-170 eluted as a major peak at 14.39 ml, close to that of the lysozyme molecular weight marker (14.3 kDa). This recombinant PilE fragment has a predicted molecular mass of 16.59 kDa. In comparison, NmPilP had a calculated mass from light scattering of 14.05 kDa and eluted from the SEC column at 14.22 ml (5.5). It can therefore be inferred that recombinant NmPilE36-170 is also monomeric under these conditions. Only those elution fractions comprising the major peak were collected and pooled, which omitted the higher and lower molecular weight contaminants from the sample, improving its purity.
Figure 6.7.1 Expression and purification of NmPilE\textsuperscript{36-170}

a. Western blot analysis of NmPilE\textsuperscript{36-170} expression in whole cells. An uninduced (U) sample is shown, alongside those induced with 0.1 mM, 0.5 mM and 1mM IPTG for 3 hours at 37°C.

b. Fractions from IMAC purification of NmPilE\textsuperscript{36-170} analysed by SDS-PAGE. Samples shown are cell lysate flowthrough (FT), wash steps with 20 mM imidazole (W1-W3) and elution steps with 200 mM imidazole (E1-E4).

Figure 6.7.2: SEC Analysis of NmPilE\textsuperscript{36-170}

NmPilE\textsuperscript{36-170} was further purified by SEC, using a superdex 75 10/300 GL column. Molecular weight markers are indicated by arrows. Running buffer was 50 mM NaH\textsubscript{2}PO\textsubscript{4}/Na\textsubscript{2}HPO\textsubscript{4} pH 7.8, 100 mM NaCl. The protein eluted as a sharp peak at 14.39 ml elution volume.
6.8 NMR titration experiment: PilE and $^{15}\text{N-NmPilQ}^{343-545}$

Similar to previous experiments outlined in section 6.2 and section 6.6, an interaction between the pilin subunit and the periplasmic portion of PilQ incorporating the N0/N1 domains was analysed by titrating unlabelled NmPilE$^{36-170}$ into $^{15}\text{N-labelled NmPilQ}^{343-545}$. The resulting $^1\text{H}^{15}\text{N}-\text{HSQCs}$ are shown overlaid in figure 6.8.1. Up to a two-fold molar excess of NmPilE$^{36-170}$, no extensive changes to crosspeaks were observed (intermediate titration points are not shown). There are very few instances, indicated by green boxes on figure 6.8.1, where peaks appear to lose intensity or alter. These do not map to a concentrated area on the surface of the structure, furthermore residue 495T resides in the N1-domain and 320S within the cloning artefact. Also, upon a two-fold excess of one component to the other one might expect a more pronounced effect on chemical shifts if a specific interaction was taking place in solution. Therefore an interaction between the two proteins was not demonstrated in this case.

![Figure 6.8.1: Analysis of NmPilQ$^{343-545}$ and NmPilE$^{36-170}$ by NMR-CSP](image)

Natural isotopic abundance NmPilE$^{36-170}$ was titrated into an isotopically labelled $^{15}\text{N-NmPilQ}^{343-545}$ sample to a two-fold molar excess. $^{15}\text{N}-\text{HSQC}$ spectra before and after the titration are overlaid. Possible affected residues are highlighted in green boxes. Figure produced using CCPN ANALYSIS.
6.9 Discussion

The findings presented in this chapter provide evidence from structural and biochemical studies, which establishes that PilP interacts with the central N0-domain of PilQ through a beta hairpin structure within the PilP beta fold. Taken together with evidence for a PilM/N/O/P complex, this observation suggests that PilP binds to PilQ in the periplasm, and plays a role in bridging the outer-membrane components to the inner-membrane platform of the T4PB machinery, forming a macromolecular complex of proteins which span both membranes. This provides a model for the interaction, whereby the secretin is linked to the inner-membrane platform via PilP.

The findings presented here are in agreement with those from NMR-CSP experiments published by Gu et al., who studied the suspected equivalent proteins in the T2SS. They observed that strand β1 of the HR domain of OutC (GspC) formed an interaction with the N0-domain from the secretin, although the exact point of interaction on the N0-domain is not entirely in agreement with this data. Nevertheless, they also observed no effects on peak intensity in the 15N-HSQC spectra for OutC when unlabelled, recombinant N1/N2-domains from OutD were titrated into the sample. It is evident from these results that a similar situation exists in the T4PB system. Specific residues forming a well-defined patch on the surface of the conserved C-domain of PilP are shown to be susceptible to peak broadening in the 1H15N-HSQC spectra. These residues are largely concentrated to the first and second beta-strands, similar to GspC. As residues affected by PilP also reside largely within the N0-domain of PilQ, and not the N1-domain, we can infer that there may be parallels in function between PilP and the HR domain of GspC, as well as structure. These recent findings alter the previous impression that PilP has no T2SS equivalent.

The findings by Gu et al. also differ slightly from those by Korotkov et al., who produced a crystallized complex of GspC/D. NMR-CSP studies on OutC and OutD showed that the affected residues localized to strand β3 and helix α2 of the N0-domain. In the crystal structure of GspC/D the interface was formed by the antiparallel arrangement of strand β1 from both molecules. In the initial HADDOCK
runs for this project, one of the best solutions did indeed use strand $\beta 3$ of the N0-domain as the main contributor to the interface, so residues in this region were affected. However the $\beta 1$-$\beta 1$ pairing also satisfied the parameters. The final model produced in this project was based on homology to the GspC/D co-crystal, therefore using a $\beta 1$-$\beta 1$ antiparallel arrangement, based on the argument that this is in agreement with the chemical shift mapping and where residues were most affected. However there may be subtle differences between the ways in which these proteins interact within different secretion systems. This already seems apparent, as the OutC/D interaction appears to be weaker and more transient than the GspC/D interaction, based on the evidence in each paper.$^{151,152}$ Or alternatively, we may not be seeing the full picture. The nature of the interaction in vivo may be more complex, dynamic or even switch between binding sites. The precise reason and function for both of these interactions - that between GspC and GspD, and that between PilQ and PilP - remains unknown.

NMR can be a little more ambiguous than isolating a crystallized complex, because residues can be affected and therefore attenuate even if not directly involved in the interaction but nearby and affected by possible changes in the conformation of the protein. In addition, HADDOCK does not differentiate between titration points, however it does expect the user to interpret the data and only to specify severely affected residues as active, although this still remains a matter of opinion. Intermolecular NOEs provide the ideal information for NMR complex structures, but often these are elusive due to unfavourable chemical exchange timescales, low solubility of the complex and weak interactions. It was especially clear in this case that obtaining NOEs was going to be unsuccessful, as extensive line-broadening was observed and a set of peaks for the protein-protein complex could not be identified. This was most likely due to unfavourable exchange and increasingly low solubility of the complex during the titration.

The interaction between NmPilPC and NmPilQ$^{343-545}$ demonstrated by NMR was confirmed biochemically by ITC and SPR. The results from ITC experiments were not ideal, producing only partial curves and therefore only approximate $K_D$ values. The $K_D$ values obtained were weak, in the 0.1-1.0 mM range. The problems encountered with regards to stability of the complex could not be overcome by
reducing the concentrations used in the experiment, and a lack of sample homogeneity in the cell may have contributed to the increasingly erroneous baseline, further skewing the data. It may be the case that complex formation promotes aggregation of one or both components. There was, however, evidence that a weakly exothermic reaction was taking place, which was not observed when either component was titrated into buffer, and the possibility of buffer mismatch was ruled out also.

The SPR experiments produced a more positive outcome, albeit with limitations. For example, amine coupling was not an ideal approach to immobilization due to the presence of lysine 102 within the identified binding site of NmPilPC. Therefore a large amount of protein was coated onto the chip, making it difficult to ascertain the amount of 'active' protein that was present. This may have affected the stoichiometry result. The results from the NMR titrations and molecular modelling suggest it is unlikely that the N0-domain from PilQ is bivalent, i.e. interacting with two PilP molecules. Attempts were made to circumvent this by metal affinity immobilization via the His-tag, and immobilising a much lower amount of protein, but this too was unsuccessful. In each case, PilP was successfully immobilized but no net binding when the reference response was subtracted was observed, i.e. the signal to noise ratio was insufficient. This could be due to a non-specific interaction between the analyte and the chip, or again unfavourable orientation of the ligand preventing any complex formation.
Chapter 7: General Discussion

The T4PB system is a complex macromolecular machine consisting of many gene products, which work together in a concerted manner to control the assembly and retraction of the type IV pilus appendage on the bacterial cell surface. Type IV pili are versatile fibres and govern many vital functions in pathogens such as motility, host cell adhesion and biofilm formation. Therefore pili are important virulence factors for the organism, and components of the T4PB system have the potential to highlight attractive therapeutic targets. An understanding of protein structure and function relationships, and of crucial protein-protein interactions within this system is an important first step required to achieve this, and has been the focus of most studies in the field in recent years. The work presented in this project explores one of the lesser characterized interactions; that between the secretin PilQ and the essential inner membrane lipoprotein PilP. The results highlight some important parallels between PilP and an unrelated protein, GspC, from the closely related T2SS, demonstrating the importance of structure determination in understanding protein function.

The secretins are an intriguing group of integral outer-membrane proteins, which form dodecameric channels required for the passage of a range of substrates including protomers and assembled oligomers. They differ from other, less complex, outer membrane proteins in the nature of their overall architecture and requirement for a number of auxiliary factors for correct targeting and assembly. As members of the secretin family play a crucial role in the export of substrates in the T2SS and T3SS, as well as the transit of the pilus in the T4PB system, a detailed understanding of these differences and of the role of secretins is important for a proper understanding of the operation of these systems. Given their structural complexity and diversity it seems naïve to assume that secretins serve only to provide a regulated “hole” in the membrane, playing a crucial yet passive role in the secretion process. More recently their modular composition, which comprises a number of distinct periplasmic domains has become of particular interest. The combination and number of these domains differs from one secretion system to the next, and seems likely to bear functional significance. With the emergence of high resolution structural information and biochemical functional
studies focusing on the periplasmic domains, we are beginning to see a role for secretins which stretches beyond that of a passive channel to a requirement for substrate recognition and coordination of the system. It has already been shown that the Type II cholera-toxin secretion channel specifically recognises its substrate via its N0/N1/N2 region,\textsuperscript{37} and the recent descriptions of a GspD/C complex in the T2SS have demonstrated a link between a secretin and its inner membrane components.\textsuperscript{151,152}

The first stages of this project focus on the T4PB system secretins, or PilQ proteins as they are generally known. The aim was to produce recombinant forms of the periplasmic domains of PilQ homologues, in order to obtain high-resolution structural detail, infer subsequent functional information and therefore to better understand the role of PilQ within the T4PB system. During the course of the project studies have been published with similar intentions, for both the T2SS and T3SS secretins.\textsuperscript{34,35} Although the periplasmic region of PilQ was predicted by bioinformatic analysis to consist of similar domains, due to a lack of sequence conservation in these regions the findings from the T2SS and T3SS studies could not necessarily be extrapolated to the T4PB system. There may even be subtle differences in the way secretins within the same systems operate, for example regarding their specificity and recognition of different substrates. In addition, the presence of predicted $\beta$-strand domains in PilQ proteins, which do not appear in other secretins, may be a vital functional requirement for T4PB. It was therefore still relevant to pursue these studies with PilQ proteins.

The multiple target selection approach taken in this project involved generating a large number of constructs for single and combinations of domains from a variety of PilQ homologues. Methods employed in the past for studying secretins often involved isolating the protein from native sources,\textsuperscript{28,33,202,245} however this approach gave poor yields. There have also been attempts to synthesize secretins \textit{in vitro}, and this was met with some success in the case of PulD, which was successfully inserted into liposomes - therefore behaving as the native form. However this method was not suited to secretins such as NmPilQ, which require accessory factors such as PilW for correct membrane insertion.\textsuperscript{246} As the main area of focus for this project was the structure and function of the periplasmic domains, as
opposed to the secretin multimer as a whole, the best approach seemed to be to attempt cytoplasmic production of recombinant domains in an E. coli expression host, for purification as soluble proteins. Of course without a clear idea of domain boundaries within PilQ, or the stability of these domains individually, production of multiple targets increased the chances of success. In addition, it was not clear whether recombinant PilQ domains from some organisms could generally be produced more successfully than others, which may simply be a result of differences in their genetic sequence. Obtaining concentrated samples of soluble, pure protein is often the bottle-neck step of structure determination by crystallography and NMR. The high-throughput approach essentially provided a survey of a range of PilQ homologues, so that the best candidates could be used for subsequent experiments. As described in chapter 3, several of these targets did produce soluble, homogenous samples of recombinant protein suitable for structural studies. This led to two solution structures: the second beta-domain, and the N0-domain from Neisseria meningitidis. Seven models for the N0/N1 region were then generated using homology modelling, each with a slightly different relative orientation of the two domains (4.5). It should be stressed that only the N0-domain structure was solved and the rest of the model built largely by threading the sequence onto the EscC and GspD N1-domain structures, with some added information from chemical shift indices using ROSETTA. Therefore, although some experimental information was used, a solution structure for the full N0/N1 region was not determined here.

The seven closest-to-mean model structures were obtained by clustering of the 100 output structures from Modeller using HADDOCK, and this was the furthest point the data from this project alone could be stretched to. However, the work presented in this project was published alongside an improved 3D-reconstruction of the Neisseria meningitidis PilQ dodecamer, obtained from cryoelectron microscopy data collected by colleagues at The University of Manchester. This was used to select the structure that would allow docking of the N0- and N1-domains into the density map. It is relevant to include this information as part of this discussion in order to help put the N0/N1 and B2-domain models/structures that were obtained in this project into the wider context of the PilQ dodecamer.\textsuperscript{154}
The cryoelectron microscopy 3D-reconstruction of NmPilQ presents an enclosed chamber-like structure, which is sealed at both ends (figure 7.1a). The flatter end of the structure is most likely to form the membrane-spanning C-domain, according to previous work\textsuperscript{39,203,214} and comparison with the C-domain of other secretins such as PulD.\textsuperscript{33} This would mean the four modular periplasmic domains continue down towards the tapered end of the structure, from the C-terminus in the order of N1, N0, B2, B1, forming the walls of the chamber. The positions of these domains are marked on the structure in figure 7.1b, based on reasonable approximations of their length from the structures we have. Using MultiFit\textsuperscript{166} each of the N0/N1-domain models were docked into the density map, by confining the fit to the relevant section of the density map as indicated in figure 7.1b, and also favouring the orientations where the N1-domain is closest to the C-domain, as it would be in the polypeptide. The N0/N1 model with the best arrangement of the two domains with respect to each other, which initially fit best into the cryo-EM density map, was cluster 4 (figure 4.5.1). Therefore the closest-to-mean representative of this model was used for docking. Further constraints were then applied to differentiate between multiple solutions, taking into account minimization of steric clashes, and exposure of the PilP binding site determined in chapter 6 on the outer face of the chamber.\textsuperscript{141} The highest-scoring solution was selected, which did not cause a steric clash between PilP and other PilQ monomers.
Although a *de novo* structure for the N0/N1-domain was not obtained, the model provides some important information. A specific point of interest remains the relative orientation of the N0 and N1-domains with respect to each other. It has already been noted in other publications that this differs significantly between the determined N0/N1 structures of secretins from the T2SS and T3SS.\cite{34,35} When these structures are superimposed via their N1-domains, along with the PilQ N0/N1 model which best fits into the cryo-EM density, the N0-domains are rotated by a considerably varied degree with respect to N1 (figure 7.2). Although the PilQ N0/N1 structure is based on homology modelling, the cryo-EM volume represents the full PilQ dodecamer and the model fits well. In addition, the GspD and EscC crystal structures have been docked very reasonably into the contours of their respective EM reconstructions,\cite{62,167} so it seems unlikely that the alternative domain arrangements are attributed solely to crystal packing, although this could still be a contributing factor. Overall, the picture emerges that the N0- and N1-domains are flexible in their relative orientations, and this seems to be true for the
other secretins as well. It is feasible that the secretins from different systems differ in the arrangement of their domains in the multimer, and also that their periplasmic domains are flexible \textit{in vivo}. This is supported with findings by Gu \textit{et al}, who realised that in addition to an N0-β3/N1-β6 domain contact observed in the GspD crystal\textsuperscript{,34} N0-β3 was also capable of forming contacts with a neighbouring N0-β3 in solution\textsuperscript{,151}. The presence of long linkers between folded domains, differences in domain orientations of solved structures, and evidence for different conformational states of secretin multimers\textsuperscript{167,214} suggest that a flexible arrangement of domains may be the case. Presumably this flexibility is important for secretin function, allowing for conformational changes to accommodate substrates.

The B2 domain was docked into the cryo-EM density in a similar manner to the N0/N1 region. Following the sequence of domains in the polypeptide, this had to fit into the density map with its C-terminus within a certain distance from the N-terminus of the N0-domain, so that it could be reached by the length of the omitted linker between the two domains. This restricted the vertical orientation of the domain, or rotation around the \textit{x}-axis. It also restricted the domain orientation around the \textit{y}-axis, as the C-terminus had to be on the outside of PilQ and facing upwards. This was otherwise ambiguous, because no binding site for another protein was determined experimentally and mapped onto B2 during this project, and the flat, sandwich shape of the domain meant it could fit easily into the density map a number of ways. A solution was then also identified for the B2-domain, producing complete model for a large portion of the periplasmic region of PilQ. The final model is shown in \textit{Figure 7.3}.

It becomes apparent from the model shown in \textit{figure 7.3}, that there is insufficient volume at the tapered end of the PilQ structure to fit twelve copies of the B1-domain, for which we were unable to obtain a structure. The chemical shift assignment of B1 from \textit{Aeromonas} suggests it would adopt a fold similar to that of B2. Attempts to produce several B1 targets in this project resulted in low protein yield and only partially folded species, even when they were produced in tandem with the very stable B2 domain. The B1 domain may, therefore, adopt only a partially folded structure in the PilQ oligomer, therefore contributing little density to the map. The significance of this for the function of the secretin remains
unexplained; the B1 domains may be redundant or may require another component of the pilus machinery for their stability. Of course, we should remember that not all T4PB secretins include a B1 domain, although they have a B2, and the Thermus PilQ protein appears to have no Beta-domains at all. The key issue here, however, is that the presence of one or more beta-domains is an adaptation of most T4PB secretins, and they are not found in other members of the family. At the time of completing this project there is, as yet, no other published literature, which explores the structure or function of these domains. So the question remains, why do T4PB secretins have an extra domain at their N-terminus, with a completely distinct fold to the other characterized modules in the periplasmic region? Their position within the overall PilQ structure suggests they may be involved in closing the chamber at the periplasmic end, and the distinction between the T4PB system and other secretion systems is the extrusion and retraction of an assembled oligomeric structure, as opposed to secretion of proteins. As discussed in the general introduction, many secretion systems produce fimbrial structures, but it is the rapid retractile and force-withstanding property of type IV pili, facilitating the movement of the entire organism across surfaces, which makes this system stand out. Previous observations by Collins et al have shown that PilQ monomers interact directly with PilE, the main component of the pilus fibre and therefore the substrate of the secretin. This interaction was pinpointed to an N-terminal region between residues 25 and 354, and no interaction was observed in the central portion between residues 218 and 478. This indicates an interaction with what we now know to be the beta-domains, and not the N0/N1 region. This is in contrast to the cholera toxin secretin of the T2SS, which recognises its substrates via the N0/N1/N2 domains. Collins et al also describe a conformational change in the PilQ oligomer as it forms a complex with the pilus, which is able to fill the chamber. Therefore a sensible proposition may be that the beta-domains are an adaptation that has been acquired to regulate the entry of pilin, or so that the secretin may open to accommodate a dynamic assembling and disassembling pilus fibre, perhaps allowing more flexibility to T4PB secretins that is not required in other systems. It would seem that diversity in secretins from different systems has arisen by modification of their periplasmic domains, thereby resulting in specificity for their secreted substrates. This may
explain the presence of N0, N1, N2 and N3 domains in T2SS secretins, which recognise different substrates, and the contrasting B1, B2, N0, N1 organization of T4PB secretins, although this hypothesis requires further investigation.

**Figure 7.2 Relative domain orientation in secretin N-terminal domains**
The figure shows a comparison of the PilQ, GspD and EscC N-terminal domains when they are superimposed via the N1-domain (indicated by dashed lines). All adopt a different relative domain orientation. *Figures produced using CCP4MG*

**Figure 7.3: Model for the periplasmic domains of NmPilQ**
The B2 (purple) and N0/N1 (blue/green) domains were modelled into the NmPilQ cryo-EM contour map (2.9σ) to produce a model for the periplasmic domains of NmPilQ. Published in reference 154 (appendix 1).
It would be interesting to see how mutagenesis of important residues on the periplasmic domains of PilQ affects substrate recognition or pilus biogenesis in vivo, whilst maintaining the overall fold of the domains. There appears to be a recurring theme of common protein folds, which bear little sequence similarity and differ in their mode of operation, and this has been highlighted in this project. The similarity of the PilQ N0-domain to not only the T2SS and T3SS N0-domains, but also to folds in proteins involved in other secretion apparatus such as that of T4SSb and the bacteriophage cell-puncturing device, strengthens the evolutionary relationship between these diverse systems. In a similar manner, the striking structural parallels between the PilP structure and that of the HR-domain of GspC highlight a possible T2SS functional equivalent for PilP, adding to the many similarities between the two systems. Both proteins interact with the inner membrane complex and contact the outer membrane secretin in a similar manner so their roles may overlap; however, it is important that GspC has an additional coiled-coil or PDZ domain so it is unlikely that PilP and GspC are complete functional equivalents. We should also not assume that just because the two share a common fold and mode of interaction that they are performing the same function. A common fold to the N0-domain is seen, for example, across various components of the T3SS basal body including the secretin, and the inner ring components and there is nothing to suggest these domains are all performing the same role. There are often differences in the way in which structurally similar domains that reappear from system to system function, and this is not surprising given the variations in amino acid sequences. For example, the T2SS secretin GspD N0-N3 specifically recognises its secreted substrate and the pseudopilus tip complex. In contrast, it would seem that the B1/B2 domains of PilQ of the T4PB system are responsible for recognition of its primary substrate, pilin, as it was previously noted that the region corresponding to N0/N1 did not form an interaction. This is supported by preliminary findings in this study; there were no significant changes to chemical shifts in the $^{15}$N-HSQC spectra for NmPilQ N0/N1 when truncated PilE was titrated into the sample. This does not, however, rule out a role for these domains perhaps in recognising minor pilins or the pilus tip adhesin, which would require further investigation. There are even subtle differences within the same secretion systems of different organisms. For example,
in the T2SS, GspD and GspC from ETEC form a high affinity complex which co-crystallizes,\textsuperscript{152} whereas the equivalent \textit{D. Dadantii} OutC and OutD interaction caused specific attenuation of NMR chemical shifts in solution, but the weak binding meant a complex could not be isolated by either co-crystallization or size exclusion chromatography.\textsuperscript{151}

This project highlights a specific interaction between PilP and the N0-domain of PilQ which is similar to that seen in the T2SS between GspC and GspD. Similarly to that of OutC/D, a complex was not isolated by size exclusion and would not crystallize in these attempts. With hindsight, however, the inability to crystallize may be attributed to the construct design of N0/N1, which upon chemical shift assignment was shown to include a substantial disordered region. When the construct was revised to produce only the N0-domain, crystal trials were not attempted with the protein product during the timeframe of this work. Judging by the ITC data, which indicated a weak exothermic reaction, it was suspected that the interaction between PilP and PilQ was not one of high affinity. The SPR data, also confirmed binding but this could not be quantified due to the extensive limitations of amine coupling to these particular proteins. NMR chemical shift perturbation analysis however indicated a specific interaction, which could be mapped onto the surface of both protein structures. The PilP binding site was mapped to an area around the first and second $\beta$-strands, which included the hairpin loop. Small shifts were seen on other parts of the protein, as was the case for the OutC HR-domain.\textsuperscript{151} Although the GspC/D crystal indicated a $\beta$1-$\beta$1 strand complementation in the complex,$\textsuperscript{152}$ the solution state experiments on OutC/D indicated some effects of OutD on $\beta$3 of the OutC HR-domain also, perhaps hinting at a dynamic nature of interaction.$\textsuperscript{151}$ The results produced in this project are largely in agreement with both studies. With solution-state NMR-CSP, there is limitation as binding interface cannot be visualized such that orientations of interacting residues and identities of specific interacting partners can be determined unless sufficient intermolecular distance restraints can be collected. In addition, it bears the limitation that global changes to the protein upon binding could cause attenuation of backbone shifts in areas of the protein not directly involved in the interaction. However, it has the advantage of highlighting the
specificity of perhaps weak, yet biologically relevant interactions between proteins operating in systems such as these where, for example, multi-valency effects may be occurring in vivo between a multimer and several monomers.

It was evident from the crystal structure of Pseudomonas PilP determined in chapter 5, the previously determined Neisseria PilP NMR structure, and the available GspC HR-domain structures that hydrophobic residues lining the core crevice of the beta-sandwich are conserved in both systems. Therefore, it would seem, these residues are playing a vital role in maintaining the protein fold. The residues mediating interaction with the secretin, however, differ but this is not surprising given the difference in secretin partners. The β-strand complementation seen in the PaPilP dimer dimers formed in the P2, crystal, between β1 and the same strand of a symmetry related copy (5.6), hinted that this part of the PilP structure may be involved in binding another protein in the same way. After analysis of the PilP/Q interaction it seems that this was indeed the case, and a similar mode of binding is seen in the T2SS.

Using the HADDOCK and CNS 1.2 model for the complex formed between PilP and PilQ (6.3), the NmPilP structure can be placed onto the model for the PilQ dodecamer, and this is presented in figure 7.4. In this model, PilP forms a wedge between the N0 and B2-domains. This differs from the positioning of GspC to GspD in the model by Korotkov et al, in which GspC points inwards towards the bottom opening of the secretin chamber. This is perhaps unsurprising given the dynamic nature of secretins, the differences observed between the orientation of periplasmic secretin domains between the systems, and the presence of N-terminal beta-domains in the T4PB system, which would prevent this arrangement. In the model presented in figure 7.4, movement of the N0, B1 and B2-domains would be required to accommodate an assembled pilus fibre into the chamber. PilP may be a requirement for stabilizing either conformation of PilQ. The closed chamber of NmPilQ also differs from the EM-reconstruction of GspD, which is open at the base. Instead, the T2SS secretin chamber is constricted at the centre by the presence of an N3-domain, similar in structure to N1, which would prevent passage of the substrate. The lack of beta-domains in T2SS secretins, and the lack of an N3-domain in T4PB system secretins, may be a critical
adaptation to the difference in substrates. Although the substrates require a similar sized pore for extrusion, the requirement for pilus retraction may be the key difference here.

In the model presented in figure 7.4, twelve PilP monomers have been placed onto the PilQ dodecamer in a 1:1 stoichiometry, which seems sensible given the experimental and structural constraints. However, it is difficult to speculate on the stoichiometry in vivo. Korotkov et al point out that, in the case of the T2SS, the secretion ATPase GspE is hexameric and forms a 1:1 complex with the inner membrane protein GspL. Also, the GspL cytoplasmic domain is structurally similar to PilM, whilst the periplasmic portion is similar to PilN of the T4PB system. A PilM/N/O/P complex has been reported, and it is believed to have equimolar stoichiometry which would suggest that six GspC or PilP molecules may interact with the secretin dodecamer. However, the GspE/L interaction cannot necessarily be extrapolated to the T4PB system, as an interaction between the T4PB ATPases and PilM has not been observed. Therefore the stoichiometry of the T4PB inner membrane components remains an intriguing area for investigation.

The work presented in this study describes an important link between the inner and outer membrane components of the T4PB machinery, from which we can begin to see a picture of the system as a whole. The results show that the parallels between GspC and PilP extend beyond a similar fold to a similar mode of action towards the secretin. There are important differences between the two proteins, however, as GspC is a transmembrane protein which also contains a PDZ interaction domain. PilP is anchored to the inner membrane by a lipid moiety, then extends through the periplasm via an unstructured N-terminus and interacts with the secretin via its folded domain. As PilP also interacts with the inner membrane components, probably via its extended N-terminus, it may provide the crucial link and align or anchor the components of the system, which may explain why it is essential in order to form pili. Also, as suggested earlier, the protein may provide some stability to the conformational changes required of PilQ for the passage of the pilus.
As structural and functional information for secretion system components becomes more readily available, it is becoming evident that the evolutionary relationship between the six designated secretion systems and the T4PB system is stronger than previously thought. Systems are often discussed separately, as they share few sequence homologues and it was assumed based on this that they also share few functional equivalents. Recent advances have indicated that this may not be the case, and there may be a basic requirement for a set of conserved elements to constitute a functional secretion system. If so, it could be speculated that the various secretion systems have diverged from a common ancestral complex. There must of course be marked differences, and the secretion systems have diverged considerably in order to maintain specificity and the integrity of the cell membrane. There are subtle differences even within the same systems, as some organisms have multiple systems for different purposes. One example is *Pseudomonas aeruginosa*, which appears to utilise two T2SSs, the Xcp and Hxc systems, which it must distinguish between. However there are strong, underlying parallels between the diverse systems. One aspect of this is the conservation of protein structural motifs. The fold of the N0-domain from PilQ provides one example. Based on their amino acid sequences, secretins were believed to be dissimilar in their N-terminal regions and only share conservation at the C-terminus. The crystal structures for EscC, GspD and HofQ N-terminal domains showed that secretins of different systems shared a common fold. Some of the work presented in this thesis also shows that similarity this extends to the T6SS and a lipoprotein of the T4SSb. Another aspect is the overall similarity between the systems as a whole and their further similarity to other cellular systems. All of the secretion systems discussed here require an outer membrane pore, an energy source and a macromolecular complex of other factors in order to function. Also, many have adapted from other processes; for example, the T4SS is an adaptation of the conjugation system, the T3SS shares similarity with the flagellar apparatus, the emerging T6SS may have derived from the bacteriophage cell-puncturing device and most notably for the purposes of this project the T2SS shares extensive parallels with the T4PB system. The extraordinary diversity and complex evolution of secretion and fimbrial systems represents an arsenal of bacterial weaponry enabling pathogens to evade immune responses and infect
their host; however, the secretion systems may also become a weakness to these organisms because as our growing understanding of the common structural and functional themes increases, so might our ability to exploit them for antimicrobial purposes.

![Figure 7.4: Model for the trans-periplasmic assembly formed by PilQ and PilP](image)

**Figure 7.4: Model for the trans-periplasmic assembly formed by PilQ and PilP**

a. Side and top projections of the PilQ/PilP complex model. Based on the CNS 1.2 model for the NmPilQ<sub>343-442</sub> and NmPilP<sub>C</sub> the position of NmPilP<sub>C</sub> was docked onto the surface of the modelled N0/N1-domain dodecamer.

b. Cut-away projection of the PilQ/PilP complex model, showing one complex on either side of the chamber. PilP fits as a wedge between the N0 and B2 domains. Published in reference 154 (appendix 1).
Appendix

Appendix 1: The publication arising from the data presented in this thesis is appended in pages 231-245.

Appendix 2:

Assigned chemical shifts of AhPilQ$^{224-329}$, NmPilQ$^{343-442}$ and NmPilQ$^{343-545}$ were deposited in the BioMagResBank and assigned accession numbers 18419, 18459 and 18428 respectively.

Atomic coordinates and NMR restraints of B2PilQ$^{224-329}$ and N0PilQ$^{343-442}$ have been deposited in the Protein Data Bank under the accession codes 4AQZ and 4AR0 respectively.

Atomic coordinates and structure factors for the $P2_1$ and $C_2$ PaPilP$^C$ structures have been deposited in the Protein Data Bank under the accession codes 2Y4X and 2Y4Y respectively.
Structure and Assembly of a Trans-Periplasmic Channel for Type IV Pili in *Neisseria meningitidis*  

Jamie-Lee Berry1,*, Marie M. Phelan1,*, Richard F. Collins1, Tomas Adomatavicius1, Tone Tenjum1, Stefan A. Fyfe5, Louise Bird1, Ray Owens1, Robert C. Ford1, Lu-Yun Lian2,* and Jeremy P. Derrick1,*  

1 Faculty of Life Sciences, Michael Smith Building, University of Manchester, Manchester, United Kingdom; Institute of Invertebrate Biology, University of Liverpool, Liverpool, United Kingdom; 3 Centre for Molecular Biology and Neuroscience, University of Oslo, Oslo, Norway; 4 Oxford Protein Production Facility, Research Complex at Harwell, Oxford, United Kingdom  

Abstract  
Type IV pili are polymeric fibers which protrude from the cell surface and play a critical role in adhesion and invasion by pathogenic bacteria. The secretion of pili across the periplasm and outer membrane is mediated by a specialized secretion protein, PilQ. However, the way in which this large channel is formed is unknown. Using NMR, we derived the structures of the periplasmic domains from *N. meningitidis* PilQ. The N-terminus is shown to consist of two β-domains, which are unique to the type IV pilus-dependent secretions. The structure of the second β-domain reveals an eight-stranded β-sandwich structure which is a novel variant of the HSP20-like fold. The central part of PilQ consists of two α/β fold domains; the structure of these is similar to domains from other secretins, but with an additional β-bridge which links it to the second α/β domain. We also determined the structure of the entire PilQ dodecamer by cryoelectron microscopy; it forms a cage-like structure, enclosing a cavity which is approximately 55 Å in internal diameter at its largest extent. Specific regions were identified in the density map which corresponded to the individual PilQ domains; this allowed us to dock them into the cryoelectron microscopy density map, and hence reconstruct the entire PilQ assembly which spans the periplasm. We also show that the C-terminal domain from the lipoprotein PilP, which is essential for pili assembly, binds specifically to the first α/β fold domain in PilQ and use NMR chemical shift mapping to generate a model for the PilP–PilQ complex. We conclude that passage of the pilus fiber requires disassembly of both the membrane-spanning and the β-domain regions in PilQ, and that PilP plays an important role in stabilising the PilQ assembly during secretion, through its anchorage in the inner membrane.

Introduction  
Type IV pilus are long (1–5 µm), mechanically strong polymers which extend from the surface of many Gram-negative bacteria, including *Neisseria meningitidis*, *Pseudomonas aeruginosa* and *Escherichia coli* [1,2]. They are known to mediate a variety of functions, including attachment to host cell surfaces for invasion during infection [3], natural DNA competence [4] and a phenotype termed twitching motility, a flagellum-independent process which enables a bacterium to move rapidly (1 mm min−1) across surfaces [5]. The pilus core contains principally of subunits of pilin (PilE in *N. meningitidis*), a small protein which adopts an α/β fold and assembles into a helical structure which confers mechanical strength on the assembly [6,7,8]. Twitching motility is associated with a notable feature of type IV pilin: an ability to retract rapidly at a rate of approximately 1,000 pilin subunits per second, generating a powerful mechanical force which has been measured to be up to 100 pN per fiber [9,10].  

The secretins are a large and diverse family of integral outer membrane (OM) proteins which comprise key components of the type II and type III secretion systems, as well as the biogenesis systems for type IV pili and filamentous bacteriophage [11]. Three-dimensional reconstructions of secretin structure by electron microscopy have revealed that they adopt multimeric structures, characterized by the formation of large chambers which lie within the periplasm. Our previous work on PilQ from *N. meningitidis* showed a dodecameric structure, with a channel sealed at both ends [12]. Studies on the type II secretion system (T2SS) secretin PvdD [13] and, more recently VagopD which is responsible for the secretion of E. chaffeensis, revealed a cylindrical-shaped structure with 12-fold symmetry enclosing a large chamber which is open at the periplasmic end but closed at the OM [14]. The structure of a type III secretion system (T3SS) secretin can also be extracted from the 10 Å resolution cryoelectron microscopy density map of the *Salmonella* needle complex: this shows the secretin in an open state, with the needle passing through both ends of the chamber [15].  

Figure 1A shows a schematic illustration of the domain structure of *N. meningitidis* PilQ and two prototypical T2SS and T3SS secretins. All share a well-conserved C-terminal region which
Appendix

Author Summary

Many bacteria which cause infectious disease in humans use large fibers, called pili, to attach to the surfaces of the cells of their host. PI P is also involved in a particular type of movement of bacteria, termed twitching motility, and the uptake of DNA into the bacterial cell. They are made up of thousands of copies of a specific pilin protein. The process of assembly of pili is complicated: it requires the cooperative action of a group of proteins which span both the inner and outer membranes of bacteria. Here we have determined the structure of one of the proteins which span both the inner and outer membranes of bacteria. One of the proteins, PilQ, is organized in a segmented way, being divided into separate domains which are joined, hence allowing the proteins to move relative to each other. We infer that this movement is critical to the functioning of the channel, which must open up to allow passage of the pilus fiber. We suggest that the function of the other proteins we have studied, PilR, is to maintain the PilQ assembly during pilus secretion.

Here we report the structural determinations of the PilQ pili protein domains by using a combination of NMR and homology modeling. The original reconstruction of the PilQ oligomer, which we reported was generated using cryoelectron microscopy [12], which is similar overall to the domain structure of the complex, cannot reliably be used for automated docking of constituent domains into the density map. We therefore also report a new 3D reconstruction of the PilQ oligomer, generated by single particle averaging from cryoelectron microscopy data of uninfected specimens, and use this to dock the domain structure and generate the docking assembly. Finally, we take a combination of NMR, chemical shift perturbations and modeling to generate the complex formed between the first z/β domain in PilQ and the C-terminal domain of PilP. We propose that the segmented organization of the domain structure within PilQ is intrinsic to its ability to open up and form a channel to allow entry of the pilus fiber into the chamber, and its subsequent passage across the periplasm and OMP.

Results

NMR structure of the PilQ z/β-domains

Biometric studies suggested that the N-terminal regions of PilP-dependent secretins generally contained one or two putative domains, predicted to be rich in β-sheet and characteristically different from the z/β-domain observed in T2SS and T3SS secretins [32]. We therefore adopted a cloning and expression strategy which over-produced these β-domains from T3SS-dependent secretins originating from a number of different Gram-negative bacteria, including X. monogatrus, E. aerogenes, E. coli, hydrophilus, N. cambria and Y. pestis. We generally found the β2 domain more amenable to over-production and purification of the pilin protein than the pilin protein itself. The solution structure of a sequence from X. monogatrus PilQ (PDB accession 24723) is shown in Figure 1A. NMR spectra of the Q/β domain of PilQ show an α-helix of 16 residues in the N-terminal region of the domain (Figure S1). The most similar fold identified within the SCOP database [33] is the C-domain from the human Sg1 kinesin-like complex [34]. The C-domain fold is larger, however, and includes α-helical elements, such as two additional α-strands, two additional α-strands, such that B is paired with α6, instead of α4, as in the case of the domain (Figure S1). A comparison of the sequences of the second β-domain from PilQ, in different Gram-negative bacteria revealed a high degree of conservation within the region between B4 and B5, including the β5 strand (Figure S2). This observation was highlighted by application of the program CONSerf [35], which maps sequence conservation to protein structure; in this case an alignment of 63 different T3SS-dependent secretins are mapped to the surface of the B2 domain (Figure 2B). Strikingly, the most highly conserved residues map to a single patch on the domain surface, incorporating Lys252 from B4 and Asp281 and Phe282 from the B4/B5 loop. The implication is that this patch forms a binding site, possibly to another unidentified T3SS biogenesis protein.

In contrast to the B2 domain, attempts to over-produce the B1 domain from several sources generally met with limited success: protein products were either produced in low yield and/or exhibited poor stability. The best progress was made with the B1 domain from X. monogatrus: assignment of the NMR spectra

PLOS Pathogens | www.plospathogens.org

September 2011, Volume 8, Issue 9, e1002932

232
and use of chemical shift indices show that the A. hydrophila B1 domain contains nine β-strands (Figure S3). The poor stability of this single domain precluded the collection of the high-quality NOE data required for structural determination. Nevertheless, the similarity in secondary structure between the B1 and B2 domains determined by the NMR chemical shift indices suggest that they share a common origin, as seems to be the case with the repeated N0/N1/N2 domains within the N-terminal sections of the T2SS- and T3SS-secretories [11]. Most TFP-dependent secretories contain two β-domain, although the first β-domain is missing from some (e.g. S. flexneri 2a). It is noteworthy that residues which are highly conserved in the B2 domain (Figure 2B) are not found to be so in the B1 domain and vice versa. In addition, an interesting variation in prosthetic PIQ is the presence of low complexity repeat sequences, termed small basic repeats (SBRs), which lie between the B1 and B2 domains and have been shown to influence the efficiency of TFP formation [30]. The presence of such polymorphic repeat elements is unprecedented within the secretin family. As we show below, electron density within the cryo-electron microscopy map for the whole PIQ oligomer cannot accommodate 12 copies of the B1 domain if it folds into a compact, globular structure similar to the B2 domain, so it may be the case that the B1 domains adopt a partially unfolded state in the assembled oligomer.

NMR structure of the N0/N1-domains

Secondary structure predictions and sequence alignments suggested the existence of two domains which are likely to adopt...
Figure 2. Structure of the H. meningitidis FHO B2 domain. A) Two views of a ribbon plot (above) and structural ensemble (below) of the B2 domain (1B2G, 1B2H). The ribbon plot and fold topology diagram (right), generated using PyMOL [52], are shown with a monochromatic gradient from the N- to C-terminus. B) Surface and ribbon plots of the [1-domain, generated using CHIMERA [65], showing sequence conservation determined using CONURY [75]. High sequence conservation is shown in purple, medium in white and low in light blue.

doi:10.1371/journal.ppat.1002923.g002

a variant of the α/β-type fold identified in other secretins [20,21].

In a similar approach to that employed for the β-domains, single and multiple domain fragments from different bacteria were overproduced, purified and analysed by NMR. A two domain fragment from H. meningitidis, NOS:FHO[36-405] (Figure 1A), exhibited well dispersed NMR spectra; it was subsequently assigned and its secondary structure determined (Figure 5A). Both the N0 and N1 domains are folded, but N1 contains a long random coil extension of over 30 amino acids at its C-terminal end. The very intense peaks from this region obscured many of the
peaks from the folded domain of N1 and precluded extraction of the high quality NOEs required for a complete structure determination of the N0/N1 tandem domains. Using the Chemical Shift Index (CSI) information as a marker for the domain boundaries, a smaller fragment was produced which encompassed only the first domain [N0RIQ(N165-235)] and its NMR structure determined by conventional methods using NOE restraints. The high quality structure adopts a fold similar to the N0 domain identified from GdpD and EscC [20,21] (Figure 3A; Table 1). Comparison of the spectra from the single and double-domain protein samples verified that the chemical shifts from common residues in the first domain are very similar in both samples (not shown). A striking and novel feature of the domain structure is the presence of an helix at the C-terminus of this domain (circled in Figure 3A); from sequence alignments, this appears to be a general feature of the TFP-dependent secretion and is absent from other secretion types. The structure of the N1 domain was constructed using the CSI data, CS-ROSETTA and homology modeling, based on the crystal structure of the same domain from EscC [21] (Figure 3B).

Analysis of the 31N-H residual dipolar couplings (RDCs) indicated that the N0 and N1 domains have no fixed orientation relative to each other in solution; it was therefore not possible to obtain a common orientation in the alignment frame for the N0 and N1 domains from the RDC measurements. However, the rotation correlation times, calculated from the 1H N T2 and T1 values obtained separately for the single N0 domain (T2 ~ 9.6 ns) and the N0/N1 double domain (T2 ~ 14 ns), suggest that the N0 and N1 domains do not tumble completely independently. It is likely that the helical part of the linker between the two domains reduces the flexibility in this region. We therefore generated 100 structures of the N0/N1 double domain using CS-ROSETTA [37], with varying inter-domain orientations. The relevant sections of the cryo-electron microscopy density map was then used to identify the cluster of structures which gave the best fit, as well as satisfying other constraints (see below). Interestingly, the relative orientation of the PIPQ N0 and N1 domains bears a closer similarity to that observed in the TSSS secretin EscC [21], rather than the TSSS secretin GdpD [39]. Clearly, crystal packing constraints and other factors can also influence relative domain orientations. Nevertheless, our observations do lend weight to the idea that the flexibility of the N0/N1 secretin domains could be an integral part of their function.

**Binding of the PIP C-terminal domain (PIP72-154) to the PIPQ N0 domain**

PIP72-154 is a recombinant fragment which corresponds to the C-terminal domain of the PIP (protein Figure 1B). Titration of
unlabelled NUNPIQ"b) at "c" of the structure, from the chemical shift changes [25,38], a patch of residues on the FP domain surface involved in binding. These were constrained mainly into an area around the B1-B2 hairpin in the PIPb structure (Figure 4A). The reverse experiment, where unlabelled PIPb was titrated into NUNPIQ"b) at "c" of the structure, demonstrated that it is the N1 domain, rather than the N0 domain, which is involved in recognition of PIPb (Figure 5). The binding of PIPb to the N1 domain was confirmed with a broad range of binding experiments (Figure 5A), in which particles of the complex were blocked in the box region of the FP domain surface involved in binding, with the exception of the FP domain surface involved in binding (not shown).

The identified residues involved in binding on the surface of each protein were used as input into the restraint-based docking programme HADDOCK [39,40] to generate a structural model for the N0-N1 complex (Figure 5B). The selected HADDOCK-generated clusters were ranked based on the distance between the structures and their RMSD from the 2D68 structure (Figure 5C). The binding site is centered around an edge on interaction between the two N0 domains in each domain. Residue conservation was mapped on to the NNPIQ"b) structure using CONSURF [33], in a similar manner to its implementation for the B2 domain (above), and provided evidence that the proposed binding site for PIP is moderately or well conserved within type IV pilus-dependent secretion (Figure 4D). We conclude that the C-terminal domain of PIQ (Figure 1B) recognizes the N0 domain from N. meningitidis in a similar manner to that for GspG and its cognate GspD secretin. This is therefore a further example of the congruence between the type II secretion and type IV pilus biosynthesis systems.

**Cryo-electron microscopy structure of the PIQ dodecamer:**

We have previously reported on the structure of the intact N. meningitidis PIQ oligomer, using negative stain-based methods [12,43]. This work established that PIQ forms a dodecamer, in common with the T2SS secretion apparatus [35]. In order to generate an accurate structural model of the complex, we determined the 2D68 structure, which would allow docking of the domain structures presented above, we determined a 3D reconstruction of the complete PIQ dodecamer by cryo-electron microscopy. PIQ particles were well dispersed and clearly identifiable (Figure 5A). Single-particle analysis of 25,000 particles generated a good range of top, side and intermediate views (Figure 5B). The final structure, measured 15 Å in height and 110 Å in its widest external extent, forms a shell around a large central cavity (Figure 5C). The structure of the PIQ dodecamer is illustrated in Figure 6A. From our previous work [12,43], and comparisons with the structures of other secretion systems, we ascertain the flattened circular density at the base of the structure to the monomer-spanning C-terminal domain, which is tightly conserved within the secretion family. Our work above has established that PIQ structures in common with the other secretion systems, adopts a "string of beads" type domain organisation. Combining the evidence, we conclude that PIQ is structurally well conserved in the secretion family. The N-terminal region, encompassing the "beads" domain, therefore, form the part of the oligomer which closes the chamber at the bottom (outlined in orange in Figure 6A).

Alignment of the PIQ density map with the T2SS secretin VcGspD [14] shows some key structural differences between the two. PIQ is more compact and, critically, closed at the base, where VcGspD has a more open funnel to the secretin chamber (Figure 6B). We attribute this difference to the presence of the B1 and B2 domain in PIQ, which are absent from VcGspD (Figure 1A). The periplasmic cage structure found in VcGspD, which is known to be much more flexible than its N0/N1 domains at the base of the cage (Figure 6A). A superposition of PIQ on to the T2SS secretion apparatus structure of the T2SS needle complex from Yersinia pestis [35] enabled a comparison with the structure of the secretin in the open form. The 3-helix component from the needle complex forms a cylindrical structure in the open form, with the base of the needle forming a symmetrical structure (Figure 6C). A comparison suggests that both the top and bottom parts of PIQ might open up to allow passage of the type IV pilus fiber, in keeping with our previous observation that TFP can bind into the PIQ chamber when it is open [41]. Direct comparisons of domain organisation to respective density maps were complicated by possible differences in detergent mass associated with the transmembrane regions, and the large amount of predicted coiled-coil structured polypeptide in the PIQ sequence, with associated uncertainty about the degree to which these regions might contribute to observed density. Nevertheless, it is clear that significant structural differences exist between different secretion types, and also that such structures must be dynamic to allow passage of secreted pilus fibers and coiled-coil substrates.
Figure 4. Structural model for the PIP C-domain bound to the RHQ Nδ domain. A) Peak attenuation mapped on to the PIP C-domain (PDB accession 2WVJ) following titration with N-terminally truncated RHQ. Ratios of PIP/RHQ were colored as follows: 1:0, red; 1:0.2, orange; 1:0.4, yellow; 1:0.6, pale-yellow. Left, ribbon plot with (1) strand marked; right, surface plot. B) Peak attenuation mapped on to N-terminally truncated RHQ. Ratios of PIP/RHQ were colored as follows: 1:0, dark-blue; 1:0.2, blue; 1:0.4, cyan; 1:0.6, pale-blue. C) Model of the PIP-RHQ complex generated from CNS12.54, with PIP in red and RHQ in blue. Residues at the N- and C-termini have been removed for clarity. D) Surface plot of the RHQ C-domain generated using CONSurf [35] and Chimera [36], with the same color scheme as used in Figure 2B. A ribbon plot of the PIP C-domain structure (2WVJ) is shown in green. The same sequence set was used for CONSurf as employed in Figure 2B. doi: 10.1371/journal.ppat.1002031.g004
Figure 5. Determination of PIQ oligomer structure by single particle averaging. Raw data image and class averages. A frozen-hydrated specimen of PIQ complex was resuspended in a 1% glutaraldehyde-containing buffer at 1 mg/ml. The particles in the 100 nm diameter hole in the carbon support film were selected interactively for image processing and single particle alignment, and then a preliminary model was generated by selecting projections with bilayer symmetry as well as 12-fold rotational symmetry. This model was then used as an alignment reference to which all particles were aligned, after which a refined model was generated. After several rounds of iterative refinement, no further improvement in the model was detected (as judged by Fourier Shell Correlation between the Nth and (N+1)th iteration). B. Comparison between different orientations of the final model (even numbered projections) and averages of all the particles best corresponding to those projections (subsequent odd-numbered projection averages). Deviations between the odd and subsequent even numbered projections reflect the errors in the processing procedure, especially with respect to classification and rotational and translational alignment, and are a useful separate measure of the resolution and reliability of the structural data.

Docking of PIQ domain structures into the cryo-electron microscopy density map and assembly of the PIQ/PFP dodecameric complex

Structures of the BC domain [2PIQ] (129-310) and N0/N1 double domain [1PIQ](350-335) were docked into the cryo-electron density map using MULTIFIT, a program which has been shown to work well for structures with multiple components, even with low resolution maps [14]. In addition to optimal fit to the density and minimization of steric clashes, further constraints were applied to differentiate between multiple potential solutions. First, fitting was confined to the relevant sections of the map for each domain, as shown in Figure 6 A. Second, orientations of the N0/N1 structure which placed the N1 domain closer to the membrane-spanning region were favored. Third, the PFP binding site

Figure 6. Cryo-electron microscopy structure of the PIQ dodecamer and comparison with Salmonella type III secretion system needle complex. A) Left panel: surface-orientation map, scale bar = 100 A; right panel, as left, but with the first half of the volume removed to reveal major domain boundaries. B) Superposition of the PIQ density map (dark gray) onto Vibrio cholerae GspC (light gray, PDB-1763). The periplasmic gate structure is outlined in blue. C) Superposition of the PIQ density map (dark gray) onto Salmonella T3SS needle complex map (light gray, PDB-1767). The approximate locations of the outer membrane (OM) and inner membrane (IM) are shown and the density attributed to the IMG secretin is highlighted in blue.
needed to be exposed on the outer surface, in keeping with our previous demonstration that this is the case [27]. Some orientations were also precluded because they create steric clashes between PIP and adjacent PIPQ molecules. Fourth, the distance between the C-termini of the second β-domain and the N-terminus of the N0/N1 double domain needed to be lower than the maximum span which could be plausible bridged by the missing residues. This latter criterion ruled out an "inverse" orientation of the second β-domain, in which the direction of the last β-strand is towards the base of the PIPQ oligomer (i.e. the N-terminal end). These constraints were applied to the highest-scoring solutions obtained from MULTIFIT [44], and succeeded in identifying a unique solution for the locations of both R1PQ[234-236] and N0X1PQ[234-236] which satisfied all the criteria (Figure 7A). A striking feature of the resulting assembly is the location of the C-terminal helix in the N1 domain, which is oriented vertically, lining the sides of the top of the chamber and presumably forming a link to the transmembrane domain at the C-terminus. Although the B2 domain fitted extremely well into the relevant part of the map, there was insufficient volume remaining to accommodate a further 12 copies of the B1 domain, if it is assumed that it adopts a similar folded, globular structure. As
discussed above, however, our structural work on several such domains from different bacteria did not identify any that were completely folded. We therefore propose that the first β-domain adopts a partially folded structure in the PIP oligomer, sufficient to contribute some density to the map, but that once it is formed in our model it remains poorly defined or purposeless and inert.

Using the model for the PIP structural domain complex determined above, the PIP-C-terminal domain can be placed onto the PIP assembly (Figure 2B). PIP protein omitted from the assembled PIP complex, in an orientation which is different from the T2SS GspD-GspC complex: in that case, GspC was placed closer to the interior of the secretin channel [83]. It is also readily apparent that the PIP-C-terminal domain lies close to the B2 domain, essentially洗礼ed as a ‘bridge’ between the N0 and B2 domains (Figure 7C). An assembled TFP fiber measures 60 Å in diameter [45]; passage of the pilus fiber would therefore require movement of the PIP-C-domain (Figure 6A), as well as the B1-B2 domains and possibly also the linker between the N0 and N1 domains (Figure 7C). One obvious function of PIP, therefore, is to stabilize the PIP oligomer during secretion, preventing disassociation and consequent disruption of the channel.

Discussion

Recent structural work has started to shed light on secretins and the way in which they mediate the translocation of exoproteins across the OM. A question of particular importance is how secretins are able to function in different secretion systems but here we highlight a critical adaptation of TFP-dependent secretins which is not found in members of the family elsewhere: the presence of separate β-domains which are involved in closing the channel at its periplasmic end. The β-domain pair appears to be involved in this purpose and, by inference from our previous observations on the filling of the PIP domain with TFP [43], be involved in guiding the entry of pilis or an assembled pilus fiber. A prevailing theme in structural studies on secretins is the modular organization of their domains. Here we provide evidence that, even in the central part of the chamber where the gripper of the pilus fiber is in closest contact, there must be some mechanism to accommodate the pilus fiber during secretion and retraction (Figure 7C). We do note, however, that the type IV pilus N-terminus can undergo a force-induced narrowing to a form with a diameter reduced by 40% [46]. We cannot exclude the possibility, therefore, that the PIP domain could house the pilus fiber in an intermediate and narrower state. Flexibility of movement between adjacent domains, which we have demonstrated experimentally for the N0 and N1 domains, is likely to be a critical part of secretin function. There is also evidence that secretins somehow recognize their secreted substrates [14,47]. These observations suggest a model in which the secretins associated with different secretion systems have diversified by modification of their periplasmic domains, and it seems likely that this is where the specificity for recognition of their secreted substrates resides. Such specificity may be necessary in organisms such as P. aeruginosa which have the capacity to express more than one secretin and may therefore require mechanisms to distinguish between them.

The B2 domain sequence is well conserved in PIP sequences from other bacteria (Figure S2), suggesting that our observations can be generalized, at least to type IVa pilus-dependent secretins [2]. It is less clear, however, whether type IVa pilus-dependent secretins adopt the same domain organization as shown in Figure 1A. Sequences of BfpB from E. coli, and TepC from F. novicida, did not align well with the respective B2 domain sequence, leaving this as an open question at present. The type IVb pilus-dependent secretins differ in other respects: they have lipid attachment sites at the N-terminus, for example, and make no readily apparent equivalent of the PIP equivalent.

Our previously reported structural studies on V. vulnificus PIP, by electron microscopy and cryo-electron tomography, were carried out using negatively stained specimens [12,42], whereas the current structure has been determined in the absence of stain in vitreous buffer. To date, the best structure available for the PIP oligomer was obtained using cryo-negative stain, a procedure which involves addition of a negative stain agent (ammonium molybdate) to the sample before freezing. The additional contrast obtained using negative staining led to a higher quoted resolution value (12 Å) than that cited here for a low contrast, unstained sample (19 Å). However, the fitting of domains into a low resolution structure requires a good representation of the true distribution of protein density across 3D space. The resulting map records the molecular envelope well, but not the internal hydrophobic features of a protein which exclude the stain. The structure reported by Collins et al. [12] is adequate to delineate the gross structural features of the PIP oligomer but could not be reliably used for automated docking using MULTIDOCK [44], or similar programs, which make no allowance for the contribution of negative stain. Additionally, negative staining of hydrophobic regions of protein may sometimes occur, resulting in an incorrect envelope and a protein deficit where protein density should actually be observed. Finally, the staining pH and ionic strength are usually under nonphysiological conditions, resulting in structural changes in the protein that may be artefactual. We therefore argue that the current structure, although it is at lower resolution than that reported by Collins et al., is nevertheless a much better map into which domains can be fitted. A second difference between the two structures concerns the symmetry applied: C12 symmetry was apparent in the structure studied by Collins et al., but C4 symmetry was applied as a more conservative option, given that the C4 signal was stronger and the apparent partial duplication of particles valid for the data. There, much stronger evidence has emerged for C12 symmetry of secretins [14]. Application of C4 symmetry in the refinement of either C6- or C12-symmetry models invalidates the representation of the structure during refinement, validating the imposition of C12 symmetry on the structure presented here.

There are a number of well-reasoned similarities between the proteins involved in TFP biogenesis and the T3SSs. These include not just the secretins and cytoplasmic ATTDoms, but also structural components such as the cytoplasmic protein PIM, which has a similar fold to the T3SS protein EplA [48]. Here we have shown that PIP binds to the N0 domain of PIP, in a similar manner to the recognition of the GspD secretin by GspC [51]: the analogy therefore extends from similarity in fold between the two pairs of proteins, to a similarity in their mode of recognition. This provides further weight to the view that the two secretion systems are evolutionarily related. There are also important differences between the two systems, however. GspC is a multidomain protein, with a transmembrane helix and a C-terminal PDZ domain, as well as the HB domain which is similar in fold to PIP. PIP is also membrane-associated, but through a lipid anchor which is covalently attached to its N-terminus. Between the lipid anchorage site and the beginning of the globular domain in PIP at the C-terminus, there is a proline-rich sequence comprising some 60–70 residues which is unstructured, at least in the V. vulnificus protein [29]. Sequence alignments and secondary structure predictions suggest that this is also the case in other Gram-negative pathogens (not shown). Work on PIP from F. novicida.
Appendix

A Trans-Periplasmic Channel from N. meningitidis

has established that it also binds to the inner membrane proteins PIP and PS2, probably through the unstructured N-terminal region [38]. This view has also been confirmed recently in N. meningitidis [49], and through pull-down experiments with N. meningitidis PIP in solution (unpublished data). Why is it the case that expression of PIP is critical to TPP assembly in N. meningitidis? [50] Our structure-based model of the PIP-PIP complex, combined with these other recent observations, suggests that it could play a key role in maintaining assembly of the PIP oligomer during plasmid secretion. There would be much reduced contact between adjacent PIP monomers in the oligomer, once the C-domain and BI/IB domains have opened up (Figure 6A). We note that some of the secreted periplasmic domains studied do not form dodecamers when expressed separately in recombinant form, suggesting that the interactions between adjacent monomers in this part of the oligomer are generally weak. PIP, on the other hand, is linked to the inner membrane through a lipid moiety and interacts with PIB and PS2; through its flexible N-terminus [48,49]. Our current hypothesis is that PIP's need to maintain the integrity of the PIP oligomer during secretion, and that it does this by effectively forming a bridge between the PIP periplasmic domains and the inner membrane.

The large periplasmic channel formed by PIP is reminiscent of similar structures found in other OM protein secretory complexes. The WaaL translocase for capsular polysaccharides, for example, forms a more elongated channel but it is also sealed at the periplasmic end [50]. The type IV secretion system complex spans the entire periplasmic space; in this case, it is a double walled structure with an opening on the cytoplasmic side [51]. Similar studies on the TFP biogenesis system have been complicated by difficulties in isolating correctly folded and assembled full length PIP in recombinant form, and in reconstituting the core secretion platform from purified inner and OM components. Our deconstruction of the PIP-PIP binding site and ability to reassemble the PIP-PIP complex therefore represents a first, but crucially important, step on the pathway to reassembling this complex molecular machine.

Materials and Methods

Protocols for expression and purification of all proteins used in this study described in Text S1.

NMR Spectroscopy

Data collection. 2D and 3D PIP samples prepared for NMR analysis consisted of natural isotopic abundance or 98% 15N and 99% 13C labelled protein (1 mM-550 mM) in 500 μl of 90% H2O, 10% D2O (or 90% 2H2O for specific experiments on NONIPQ) [54-56] in a solution containing 30 mM NaCl and 50 mM sodium phosphate at pH 6.8. All NMR experiments were carried out at 298K on Bruker Avance III 600 MHz and 800 MHz spectrometers equipped with TCI triple resonance cryoprobe. Spectra were processed using Topspin 2.1 (Bruker) and the Astra processing package provided as part of the CCPN suite with assignment carried out using CCPN Analysis [52]. Triple resonance assignment was obtained utilizing two-dimensional 1H and 15N HSQC in conjunction 3D HNCA, HNCACB, HNCACOA, HNCO, ODCA/CNOH, HBCNH, HBCACONH, HBCCGD/HD and HCC-HOCY experiments. Distance restraints were obtained from 3-dimensional 1H-15N NOESY and 1H-13C NOE using a mixing time of 100 ms. In a set of experiments were collected on NONIPQ343-353 in 100% H2O; 1H-15N HSQC, 15N-13C NOESY. Chemical shifts were referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Assigned chemical shifts of the first B-domain (B1) from Anabaena variabilis, B2PQG239-257, NIPQ343-353 and NONIPQ343-353 were deposited in the BioMagRedbank and assigned accession numbers 18149, 18159 and 18428 respectively. The Chemical Shift Index [53] of each protein was used to determine degree of secondary structure.

NMR solution structure calculation & validation. Automated NOE assignment and preliminary structure calculations of B2PQG239-257 and NONIPQ were performed using CYANA 2.1 software [54,55], with input data of shift lists derived from 1H- and 15N-HSQC spectra. For B2PQG239-257 a total of 4550 unassigned NOE peaks were picked of which 3981 were assigned. For NONIPQ343-353 a total of 2077 unassigned NOE peaks were picked of which 2377 were assigned. CYANA 2.1 calculations were ran with shocked protocols using 7 cycles of automated NOE assignment and structural calculations, producing 46 structures per cycle. Structures were subsequently water-refined using CNS1.2 [56] with a total of 2607 unambiguous interproton distance restraints and 151 predicted restraints for B2PQG239-257 and a total of 1399 unambiguous interproton distance restraints and 116 predicted distances for NONIPQ343-353. Dihedral φ and χ torsion angles were produced by TALOS+ [57] and the final ensembles of the best 20 water-refined structures were selected on the basis of low total and NOE energies, and validated with PROCHECK-NMR [58] using the iCrime interface (http://mrzcmr.dtu.dk/CRIME/CRIME.html). Atomic coordinates and NMR restraints of B2PQG239-257 and NONIPQ343-353 have been deposited in the Protein Data Bank under the accession codes 4AOZ and 4A3O respectively. The secondary structure of B2PQG239-257 was calculated using the STRIDE webserver (90.00). Surface analysis employed NACCESS (Holmabad, S.J. & Thornton, J.M. [95]). ‘NACCESS’, Computer Program, Department of Biochemistry and Molecular Biology, University College London) for identification of exposed hydrophobic residues. CPMDMG (61) was used for calculation and display of electronic surface potentials, and Pymol (The PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC) for secondary structure and side chain analysis. Schematic representations of the secondary structure layout were built using TDPRAW [62]. Characterization of the domain field was carried out using SDOP [23], alignment of the C-terminal family employed PROMALS3D [63] for secondary structure-driven sequence alignment and MULTIPROT [64] for homologous structure alignment. Random coil index (RCI) analysis was carried out using the RCI webserver [57].

NONIPQ343-353 model. The structure of the N1 domain (NONIPQ343-353) could not be determined using conventional methods due to the low number of unambiguous NOE's. Therefore the H1, H2, C1, C2, N1, and H3 assigned chemical shifts were submitted to the CS-ROSETTA webserver [65] available on the eNMR grid. The C terminus domain boundary of NONIPQ343-353 was identified using RCI [57]. To check the suitability and reactivity of the CS-ROSETTA approach, the structure of NONIPQ343-353 was also determined using CS-ROSETTA and compared with the NMR-derived structure. The helical linker region was included for both the NONIPQ343-353 and NONIPQ343-353 ROSETTA structure determinations; in both cases the linker region consisted of a four turn helix comprising P419-E432.

A complete model of the NONIPQ343-353 double domain was assembled using MODELLER [66] in multiple template mode. For the N1 domain the NMR structure was used as a template and
for the linker region the ROSETTA model structures spanning PQQ residues 419–437 were used. For the N1 domain, homology
geneous domains from the Euc2 and GypD structures (SCOP residues 185–173 and 96–120 respectively) together with the
ROSETTA model were used. To eliminate orientation bias
between both the N0 and N1 domains and the linker, each
template consisted of sub-structures 64-64-64 domains. As no restrictions
were placed on interdomain orientation, 160 structures
were calculated and clustered accordingly. These 160 structures were
divided into 7 clusters based on the criterion that the RMSD of the
cluster be no more than 3 Å and each cluster must comprise four
or more structures. For each cluster a representative model was selected as the closest to residue structure. The best approximate
orientation was then selected from these structures based on
the quality of its fit to the electron microscopy density map using
MULTIBYTES [44], as described in the main text.

**Titrations.** For the PIP-PiQ titration, a low concentration
(50 μM) sample of “N-linked PIP75–185” was prepared in 50 mM
NaCl and 50 mM sodium phosphate at pH 6.8. The titration was
carried out to yield the following ratios of [PIP]:[PDQ]:18.1, 10.2, 14.5, 10.0, and 1:1 with a final concentration of 60 μM PIP75–185
2D 1H 15N HSQC experiments were carried out with 2 scans, 256
increments. To ensure peaks were not lost due to dilution effects at higher
concentration points two datasets were acquired, one with low and
the other with higher numbers of scans; in all cases; it was possible to
confirm that the loss of peaks was due to binding rather than
dilution. The reverse experiments using natural isotopic abundance
PIP75–185 and “N–15 PiQ (NIPF75–185) or N1/PiQ75–185”,
again identified a set of peaks that remained upon binding. Spectra
were collected at titration points of [PIP]:[PiQ]:1:9, 1:9, 2:10, 1:3, 1:3,
1:40 and 1:1. All spectra were collected and processed using
TOPSPIN 2.1 (Bruker, Billerica)

**Pip75–185 assignment and binding site mapping.** Pip75–185
1H 15N HSQC was assigned by transfer from BMRB using the
SCP program. The 15N-1H HSQC was then identified and processed using
TOPSPIN 2.1 (Bruker, Billerica). The spectral assignment
showed that five backbone NH assignments were obtained for residues 79–163 based on
the closest single matched peaks between BMRB reference assignments
and experimental data. Peaks affected by the binding of Pip75–185 domains
were mapped onto the PIP structure and color-coded according to the
to the backbone NH peak

**Generation of a structural model for the N1/PiQ75–185.
** Pip75–185 complex. Initial protein-protein docking utilized the
restraints-driven docking program HADDOCK [Highly disputable
biomolecular Docking] [39, 40] with the NMR structures
of NIPF75–185 and the C-domain of PIP-PEF DB2 V1.29
AIR restrains generated from peak attenuation measured during
titrations were used as constraints for the rigid body docking.
Five clusters were produced with a 7.5 Å cutoff, one of which, closely
resembled the equivalent GepGoA complex [PDB: 3OPB.pdb].
This cluster structure was used to guide the identification of
five intramolecular hydrogen bond restraints in the PIP-PiQ
complex: 80(810) Nε1 Nε2 85(84) P 85(84) Nε1 85(84) Nε2 85(84)
80(810) Nε1 Nε2 85(84) P 85(84) Nε1 85(84) Nε2 85(84)
Using CNS1.2 [56], 100 structures were then calculated using these additional restrains,
together with the intramolecular NOE restrains obtained from
the NIPF75–185 and PIP-PiQ75–185 structures. An ensemble
comprising the 20 lowest energy structures was obtained.
The advantage of this dual HADDOCK-CNS approach was that it
overcame problems associated with rigid body docking.

**Cryoelectron microscopy and image analysis.** 3 αl samples were applied to a grid in Quantifoil R1.3/2 hole
carbon-coated EM grids and blotted using Whatman No-1 filter
paper (2×1 see Box) at 90% humidity, and then frozen in liquid
ethane using a Vitrofit plunge freezing system (FEL, Hillsboro, OR).
Cryo-EM was performed using a Tecnai F20 200 kV EM operating in
low dose mode at 200 kV. Micrographs were recorded using a Gatan 4 k Lk CD at underfocus in the range
1.5–5 μm and with a calibrated magnification corresponding to
4.5Å/pixel at the specimen level. Images were recorded under
low-dose mode with an overall electron dose of 30–25 electrons/
Å². Particles were selected into 64-64 pixel boxes (equivalent to
296–296 Å²) from the digital micrographs using the EMAN
software package [69] and masked with a circular mask of radius
135 Å. After correction of the microscope contrast transfer
function (CTF) and removal of outlier particles (based on size),
a final dataset of 25,303 particles were used to calculate the low resolution
3D structure of PIP. An initial model was generated by
sidechain (<5.5 Å) subsets of particles with the strongest
K-means symmetry and strongest bilateral symmetry, and then
calculating a noisy 3D structure assuming an orthogonal
relationship between the two sets of particles. (EMAN command
kmeans). Based on prior work [12] we generated preliminary
models for both C4 and C12 symmetry. Iterative refinement of
the initial structures was subsequently carried out using the
entire dataset, and using both C4 and C12 symmetry for refinement of
each model. Comparison of projections of the 3D structures with the
corresponding particle class averages, showed a good
agreement with the C12 symmetric structure (Figure S5). Moreover,
call using C12 symmetry in the refinement of either C4 or
C12-symmetric preliminary models led to convergence. Estimation of
the resolution of the final structure using the same method applied
by Collins et al. [12], measuring the value at which a comparison of
the Fourier shell correlation (FSC) of one half of the dataset with the
other reached 0.5, gave a value of 1.8 Å. Application of the
more recently introduced, and more conservative, measure software [68], gave a value of 1.75 Å resolution. Maps derived
by electron microscopy were displayed with the CHEMERA
software package [69]. The Pip75–185 density map was deposited in
the EMDBank with accession code EMD-2105 and coordinates for the
modelled Pip75–185 complex are available as PDB deposition
6AV2.

**Supporting Information**

**Figure S1** Comparison of the folds of the R2 domain from N. meningitidis PIP (2PIQ75–185) with the CS
domain from human SCG1. The second fold is shown on the
left, in green, and the CS domain on the right, in light blue
(PDB accession code 1RL1). (TIF)

**Figure S2** Structure-based sequence alignment of R2
domains. The locations of β-strands in the N. meningitidis
structure are shown. Numering is for the N. meningitidis
sequence. Residues which are well conserved are highlighted in
bold. Example sequences shown are: Pseudomonas aeruginosa (Pa), Xantho-
monas campestris (Xc), Aeromonas hydrophila (Ah), Legionella pneumophila
(Lp) and Synechocystis sp. (Sp). Uniprot codes: A5LE4A, B0RCP3, A0K930, Q8Y32
and B5B732 respectively. (TIF)

**Figure S3** Chemical shift and deduced secondary structure
assignments for the R1 domain from Aeromonas hydrophila.
Top CSI calculated for deviations from random coil
shifts of H2, C2 and CO to determine the consensus secondary
structure, graph adapted from CUPIN analysis. Robust
alignment of RI domains in A. hydrophila (Ah), N. meningitidis (Nm), P. aeruginosa
(Pa).
Appendix

A Trans-Periplasmic Channel from N. meningitidis

Test S1: Protein expression and purification. Methods for the expression and purification of all proteins used in this study. (PDF)

Acknowledgments

We thank Sraya Pandurangan (University of Liverpool) for assistance with the NFPiQG-42-77 (Gene and Dr. R. Harris (University of Georgia) for his contribution to electron microscopy data collection.

Author Contributions

Conceived and designed the experiments: JPDL YLY RCF TT. Performed the experiments: JLY YMP YAL YRF RCF ECY SAF. Analyzed the data: MMP YLY JRL RCF RCF ECY. Contributed reagents/materials/analysis tools: SAF TT. Wrote the paper: JFL YMP JRL ECY RCF RCF TT.

References

Reference List


References


Durand, É. *et al.* Type II protein secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *Journal of Bacteriology* 185, 2749-2758 (2003).


Py, B. a., Loiseau, L. & Barras, F. An inner membrane platform in the type II secretion machinery of gram-negative bacteria. *EMBO Reports* 2, 244-248 (2001).

REFERENCES


199 Krissinel, E. H., K. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute, http://www.ebi.ac.uk/pdbe/prot_int/pistart.html


