Analyses of the proteins KpsM, KpsE and KpsD in the group 2 capsular polysaccharide export complex of *Escherichia coli*

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Life Sciences

2012

Eva Kathrina Haas
# List of Contents

- List of Contents ........................................................................................................... 2
- List of Figures .................................................................................................................. 7
- List of Tables .................................................................................................................... 9
- Abstract .......................................................................................................................... 10
- Lay Summary .................................................................................................................. 11
- Declaration ...................................................................................................................... 12
- Copyright Statement ..................................................................................................... 12
- Acknowledgements ....................................................................................................... 13
- Abbreviations .................................................................................................................. 15

## Chapter 1 Introduction ................................................................................................. 17

1.1 Bacterial polysaccharides ............................................................................................ 18
1.2 Capsular polysaccharides in pathogenic bacteria .......................................................... 18
   1.2.1 Pathogen-pathogen interactions: the role of capsules in biofilm formation .... 19
   1.2.2 Pathogen-host interactions: the role of capsules in host cell adhesion ............ 20
   1.2.3 Pathogen-host interactions: the role of capsules in resisting host immune responses ........................................................................................................... 22
     1.2.3.1 Intracellular survival ....................................................................................... 22
     1.2.3.2 Modulation of immune mediators and immune cell recruitment .............. 24
     1.2.3.3 Resistance to the complement pathway ....................................................... 25
1.3 Polysaccharide capsules of *Escherichia coli* .............................................................. 27
1.4 Group 1 and 4 capsules of *Escherichia coli* ................................................................ 27
   1.4.1 Group 1 synthesis and export .............................................................................. 31
   1.4.2 Inner membrane protein Wzc and outer membrane protein Wza ................. 33
1.5 Group 2 capsules of *Escherichia coli* ........................................................................ 39
   1.5.1 Group 2 synthesis ............................................................................................... 41
     1.5.1.1 Synthesis of K1 polysaccharides ................................................................. 41
     1.5.1.2 Synthesis of K5 polysaccharides ................................................................. 44
   1.5.2 Coupling of K5 synthesis and export .................................................................. 47
   1.5.3 Group 2 export – an ABC transporter-dependent process .............................. 48
     1.5.3.1 The ABC transporter KpsMT ....................................................................... 50
     1.5.3.2 The inner membrane protein KpsE .............................................................. 52
     1.5.3.3 The outer membrane protein KpsD .............................................................. 52
     1.5.3.4 Evidence for a group 2 translocon complex ............................................... 53
1.6 Group 3 capsules of *Escherichia coli* ...................................................................... 54
1.7 Neisseria meningitidis serogroup B capsule ...................................................... 55
1.8 Haemophilus influenzae serotype b capsule ...................................................... 58
1.9 Salmonella enterica serovar Typhi capsule ...................................................... 60
1.10 Streptococcus pyogenes capsule ...................................................................... 61
1.11 Type 1 secretion system: parallels to Escherichia coli group 2 capsular polysaccharide export ................................................................. 63
  1.11.1 α-Haemolysin export system .................................................................... 63
1.12 Aims and objectives of this study .................................................................... 68

Chapter 2 Materials and Methods ................................................................. 69
2.1 Chemicals and reagents .................................................................................... 70
2.2 Bacterial strains, growth conditions and plasmids ............................................. 70
2.3 Preparation of competent cells ......................................................................... 70
  2.3.1 Chemically competent cells ....................................................................... 70
  2.3.2 Electrocompetent cells ............................................................................... 75
2.4 Transformation of competent cells .................................................................... 75
  2.4.1 Transformation of chemically competent cells ............................................ 75
  2.4.2 Transformation of electrocompetent cells .................................................. 76
2.5 DNA analysis and manipulation ...................................................................... 76
  2.5.1 Polymerase chain reaction ......................................................................... 76
  2.5.2 PCR-based manipulation of DNA fragments .............................................. 76
  2.5.3 Site-directed mutagenesis .......................................................................... 77
  2.5.4 Colony PCR .................................................................................................. 77
  2.5.5 Plasmid DNA extraction and purification .................................................... 78
  2.5.6 Restriction endonuclease cleavage .............................................................. 78
  2.5.7 Ligation ........................................................................................................ 78
  2.5.8 DNA sequencing .......................................................................................... 78
  2.5.9 Agarose gel electrophoresis ........................................................................ 79
  2.5.10 Agarose gel extraction of DNA fragments .................................................. 79
2.6 K5-specific bacteriophage sensitivity assay .................................................... 80
2.7 Protein expression, membrane extraction, protein purification and analysis..... 80
  2.7.1 Standard induction of protein expression .................................................... 80
  2.7.2 High cell density growth and induction of bacterial cells ............................. 81
  2.7.3 Autoinduction of protein expression ............................................................ 81
  2.7.4 Preparation of periplasmic and cytoplasmic fractions .................................. 82
  2.7.5 Cell lysis by French Pressing and preparation of total membrane fractions.. 82
  2.7.6 Preparation of inner and outer membrane fractions .................................... 82
  2.7.7 Cysteine-scanning labelling of proteins ...................................................... 83
  2.7.8 Nickel-affinity chromatography of histidine-tagged proteins ....................... 83
  2.7.9 Nickel-affinity chromatography combined with TEV protease cleavage ...... 84
  2.7.10 Size exclusion chromatography ................................................................. 85
Chapter 3  Analysis of the interaction of KpsE with KpsM and KpsD  .................................................. 90

3.1 Introduction ............................................................................................................................................. 91
3.1.1 Use of the pBAD28 vector system ................................................................................................. 91

3.2 Mutagenesis in the APQ motif of KpsE ............................................................................................... 93
3.2.1 Clustal alignment of APQ motif in KpsE with related proteins .................................................... 93
3.2.2 Site-directed mutagenesis of glutamine 289 in KpsE ................................................................. 95

3.3 Deletion of the C-terminus of KpsE .................................................................................................... 98
3.3.1 Clustal alignment of the C-terminus of KpsE with homologues ................................................ 98
3.3.2 Deletion mutagenesis of the KpsE C-terminus ........................................................................... 98

3.4 Interaction of KpsD with KpsE lacking the C-terminus .................................................................. 100
3.4.1 Localisation and dimerisation study of KpsD in the presence of KpsE lacking the C-terminus .... 100

3.5 Interaction of KpsM with KpsE lacking the C-terminus .................................................................. 104
3.5.1 Generation of construct carrying hisKpsM and kpsE lacking the C-terminus ............................ 105
3.5.2 Expression and location study of hisKpsM and KpsE lacking the C-terminus ............................ 105
3.5.3 Co-purification study of hisKpsM and KpsE lacking the C-terminus ........................................... 106

3.6 Discussion ............................................................................................................................................ 108
3.6.1 Glutamine residue 289 of KpsE is not essential for native KpsE function and K5 export ........... 108
3.6.2 C-terminal domain of KpsE is crucial for native KpsE function and K5 export ....................... 109
3.6.3 C-terminal domain of KpsE is not required for interaction of KpsE with KpsD ..................... 111
3.6.4 The impaired function of KpsEΔCterm is due to disruption of the KpsE-KpsM interaction ......... 113

Chapter 4  Strategies to purify KpsE and KpsM as higher order complex .............................................. 116

4.1 Introduction ........................................................................................................................................... 117
4.2 Nickel-affinity chromatography and proteolytic cleavage of KpsE-hisKpsM ......................... 117
4.2.1 Generation of pBAD28 construct carrying kpsE, hiskpsM and TEV cleavage site ................. 118
4.2.2 Expression analysis of KpsE and hisKpsM from DH5α (pBADEthisTEVMT) ......................... 122
4.2.3 Co-purification of KpsE and KpsM using TEV protease cleavage ........................................... 122
Chapter 5 Strategies for the topology analysis of KpsD ............... 150

5.1 Introduction............................................................................. 151

5.2 Topological and structural predictions of KpsD ....................... 151

5.3 Substituted-cysteine accessibility method................................ 158
  5.3.1 Expression trial of DH5α (pBADED)............................... 159
  5.3.2 Generation of construct carrying cysteine-less kpsD........... 160
  5.3.3 Development of a protocol for substituted-cysteine accessibility method for KpsD topology analysis .................................. 163

5.4 Myc-fusion approach: C-terminal myc-epitope insertion into KpsD ...... 164
  5.4.1 Generation of construct carrying C-terminally myc-tagged kpsD .... 168
  5.4.2 Expression and location analysis of C-terminally myc-tagged KpsD .... 168

5.5 Discussion............................................................................... 172
  5.5.1 Cysteine residue 488 of KpsD is not essential for native KpsD function and K5 export.......................................................... 172
  5.5.2 Suitability of substituted-cysteine accessibility method for topology analysis of KpsD.......................................................... 172
  5.5.3 C-terminal myc-tag affects KpsD expression levels and abolishes KpsD function and K5 export...................................................... 173
  5.5.4 Future directions towards establishing the KpsD topology ............ 174
List of Figures

Figure 1.1 Genetic organisation of the *Escherichia coli* group 1-4 capsular polysaccharide gene clusters. ................................................................. 29
Figure 1.2 Current model for *Escherichia coli* group 1 capsular polysaccharide biosynthesis and export. ................................................................. 32
Figure 1.3 Phylogenetic tree of selected polysaccharide co-polymerases and outer membrane polysaccharide export proteins. ................................. 34
Figure 1.4 Crystal structure of Wza monomer and octamer of *Escherichia coli* K30....... 37
Figure 1.5 Three-dimensional reconstruction of the Wzc-Wza complex of *Escherichia coli* K30................................................................. 38
Figure 1.6 Current model for *Escherichia coli* K5 (group 2) capsular polysaccharide biosynthesis and export. ................................................................. 45

Figure 3.1 Model of the secondary structure and topology of KpsE ......................... 92
Figure 3.2 Clustal alignment of KpsE residues 279 to 297 with CtrB and BexC, and AcrA and MexA......................................................................................... 94
Figure 3.3 Colony PCR of site-directed mutagenesis product from pBADE Q289A and pBADE control ........................................................................ 96
Figure 3.4 Illustration of the plasmid pBADE Q289A. .................................................. 96
Figure 3.5 Clustal alignment of KpsE residues 352 to 382 with CtrB and BexC ........... 99
Figure 3.6 Generation of the pBADEΔCterm and pBADEΔCterm hisMT constructs....... 101
Figure 3.7 Illustrations of the plasmids pBADEΔCterm and pBADEΔCterm hisMT. ...... 102
Figure 3.8 Location of KpsD in outer membrane and soluble fraction of MSΔE (pBADEΔCterm) and controls ................................................................. 103
Figure 3.9 Dimerisation of KpsD in outer membrane from MSΔE (pBADEΔCterm) ... 103
Figure 3.10 Expression and membrane location of KpsE and KpsM in DH5α (pBADEhisMT) and DH5α (pBADEΔCterm hisMT). ........................................ 107
Figure 3.11 Co-purification study of hisKpsM with KpsE and KpsEΔCterm .................. 107

Figure 4.1 Concept of on-column proteolytic cleavage following nickel-affinity chromatography. ..................................................................................... 119
Figure 4.2 Generation of pBADEhisTEVMT construct. ............................................ 120
Figure 4.3 Illustration of the plasmid pBADEhisTEVMT ........................................... 121
Figure 4.4 Expression analysis of KpsE and hisKpsM from DH5α (pBADEhisTEVMT). ............................................................................................... 124
Figure 4.5 Co-purification of KpsE and KpsM using TEV protease cleavage from total membranes of DH5α (pBADEhisTEVMT)................................. 124
Figure 4.6 Chromatogram of size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). ........................................ 127

Figure 4.7 Analysis of KpsE and KpsM in elution fractions from size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). ................................................................. 127

Figure 4.8 Analysis of total protein from elution fractions from size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). ................................................................. 130

Figure 4.9 Analysis of KpsT in whole cells and elution fractions from size exclusion chromatography of total membranes of DH5α (pBADEhisMT). .......... 130

Figure 4.10 Expression trial of DH5α (pBADEhisMT) using various induction conditions. .................................................................................. 132

Figure 4.11 Generation of plasmid pETEhisMT. .......................................................... 135

Figure 4.12 Illustration of the plasmid pETEhisMT. .................................................. 135

Figure 4.13 Expression and membrane localisation of KpsE and hisKpsM from BL21-AI™ (pETEhisMT)................................................................. 136

Figure 4.14 Nickel-affinity chromatography of total membranes from BL21-AI™ (pETEhisMT)................................................................. 136

Figure 4.15 Comparison of expression profiles of KpsE, hisKpsM and KpsT from different vectors and hosts using various induction conditions. .............. 138

Figure 5.1 Clustal alignments of homologous regions between KpsD and Wza........ 153

Figure 5.2 Model of the secondary structure and topology of KpsD. ....................... 154

Figure 5.3 Three-dimensional model of KpsD86-555 based on KpsD-Wza alignment. 156

Figure 5.4 Expression trial of KpsD from DH5α (pBADED). ........................................ 161

Figure 5.5 Generation of plasmid pBADEDΔcys.......................................................... 161

Figure 5.6 Illustration of the plasmid pBADEDΔcys.................................................. 162

Figure 5.7 Localisation of KpsD in total membranes and biotin-labelling trial using DH5α (pBADEHisMT). ................................................................. 165

Figure 5.8 Biotin-labelling of total membranes from DH5α (pBADED) and DH5α (pBADEDΔcys) ................................................................................. 166

Figure 5.9 Biotin-labelling of outer membranes from DH5α (pBADED) and DH5α (pBADEDΔcys) ................................................................................. 167

Figure 5.10 Generation of plasmid pBADEDCTmyc. Plasmid pBADEDCTmyc .......... 169

Figure 5.11 Illustration of the plasmid pBADEDCTmyc. ............................................. 169

Figure 5.12 Expression of KpsDCTmyc in whole cells. .............................................. 171

Figure 5.13 Location of KpsDCTmyc in total membranes, outer membranes and periplasm. ................................................................. 171
## List of Tables

Table 1.1 Homologues of capsular polysaccharide biosynthesis enzymes of *Escherichia coli* K1 and *Neisseria meningitidis* group B. ............................................. 42

Table 1.2 *Escherichia coli* group 2 capsular polysaccharide transport proteins and homologues. ........................................................................................................ 49

Table 2.1 List of strains used during the study. .............................................................. 71

Table 2.2 List of plasmids used during the study. ....................................................... 71

Table 2.3 List of primers used during the study. ......................................................... 72

Table 2.4 Composition of SDS-PAGE gels used during the study. ......................... 89

Table 2.5 Antibodies and other conjugated proteins used during the study. .............. 89

Table 3.1 K5 bacteriophage assays of KpsE mutations and controls. ......................... 97

Table 4.1 K5 bacteriophage assays of KpsEhisTEVKpsMT construct and controls. 121

Table 4.2 Predicted homo- and hetero-oligomeric complexes resulting from size exclusion chromatography of DH5α (pBADE) and DH5α (pBADEhisMT). .................................................................................. 128

Table 5.1 K5 bacteriophage assays of KpsD mutants and controls. .............................. 162
Abstract


The expression of polysaccharide capsules is common in bacteria and associated with virulence in some pathogenic strains. Strains of the Gram-negative bacterium Escherichia coli express a structurally diverse range of capsular polysaccharides. E. coli strains expressing group 2 capsules are associated with a number of extra-intestinal infections, including sepsis, urinary tract infections, and neonatal meningitis.

Group 2 capsular polysaccharides are synthesised on the cytoplasmic face of the inner membrane. Evidence from previous work suggests that export of polysaccharides across the Gram-negative membranes involves four transport proteins which interact to form a continuous membrane-spanning translocation complex (the KpsMTED translocon). Polysaccharide translocation across the inner membrane requires the ABC transporter KpsMT, in which KpsM is the integral inner membrane component and KpsT is the ATPase. Transport across the periplasmic space and outer membrane involves the integral inner membrane protein KpsE and the outer membrane protein KpsD, respectively. This thesis addressed some of the key areas in the study of group 2 polysaccharide transport by employing the K5 capsule as a model system. Using biochemical and molecular genetics approaches, the study focused on establishing functional and structural characteristics of the translocon members and analysing protein-protein interactions within the complex.

This study demonstrated that KpsE can self-associate as dimers, tetramers and possibly higher order oligomers in the absence of other capsule gene products and the K5 substrate. A mutagenesis study of KpsE revealed that the periplasmic, membrane-associated C-terminus is essential for correct protein function. Work presented here confirmed previous data, which suggested a direct interaction between KpsE and KpsM, by alternative methods, and demonstrated that the C-terminal domain of KpsE is required for this interaction. Further experiments suggested that KpsE and KpsM can both form higher order oligomers when interacting as a complex. The C-terminus of KpsE is not required for an interaction between KpsE and KpsD, and the two proteins are thus more likely to interact via their respective periplasmic domains. Generation of a theoretical model of the secondary structure and topology of KpsD predicted that KpsD is made primarily of β-sheets with some interspersed α-helices, including a larger coiled coil region. The theoretical topology model proposed an N-terminal transmembrane domain made of eight membrane-spanning regions, and a large periplasmic domain. Substituted-cysteine accessibility method and myc-epitope insertion analysis were both assessed for their suitability for topology analysis of KpsD. Myc-epitope insertion was identified as the recommended approach for future topology study. Myc-epitope tagging of the periplasmic C-terminus of KpsD revealed that a native C-terminus is essential for correct KpsD function.

In conclusion, this thesis contributes to the model of group 2 polysaccharide export in E. coli, and, more generally, provides clues about the transport of high-molecular weight molecules across Gram-negative membranes. It is hoped that a thorough understanding of polysaccharide transport might reveal therapeutic targets to block capsule export in pathogenic E. coli in the future.
Lay Summary

This work looks at the molecular mechanisms behind bacterial infections of the blood and urinary tract as well as meningitis in the newborn.

Bacteria are microscopic living organisms made of single cells and either live independently or as parasites, that is, they are dependent on other organisms such as humans. Like human cells, bacterial cells have a selective barrier around them, called the membrane, which maintains the shape of the cell and protects the environment inside the cell from undesired effects such as the entering of foreign particles. It is important for the survival of the bacterium that the membrane is permeable to some particles. This, for example, allows the bacterium to take up nutrients for growth or release toxins into the environment for protection.

Researchers have discovered that a high percentage of *Escherichia coli* (*E. coli*) bacteria which cause meningitis in the newborn and blood and urinary tract infections decorate their bacterial cell surface with a capsule made of sugar molecules. This capsule forms an additional layer around the membrane and provides the bacterium with extra protection from the outside environment. For example, the capsule helps to keep *E. coli* hydrated and allows the bacterium to hide from the immune system when inside the human body. This means that bacteria with a capsule are less likely to be killed by the human immune system and can cause an infection.

The sugar molecules which make the capsule are produced by *E. coli* inside the cell. In order to assemble the capsule on the cell surface, *E. coli* needs to transport the sugar molecules across a layer of two membranes. Researchers have established that the bacterium uses a group of four proteins embedded into the membranes to transport the sugar molecules to the surface. My research tries to understand how these proteins work together to form a transport system. Conducting a series of experiments, I studied these transport proteins separately as well as together and was able to establish interesting molecular details about the structure and function of these proteins. These data contribute to the current understanding of how *E. coli* transports sugar molecules to the cell surface. In the future, researchers hope to use this information to find drug targets in the sugar transport system. This would be the first step towards developing therapies against infectious diseases caused by capsule-producing *E. coli*.
Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other 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, trade marks 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 (see http://documents.manchester.ac.uk/DoculInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus-regulations) and in The University’s policy on Presentation of Theses.
Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Ian Roberts, for having provided me with the opportunity to conduct this PhD research project in his laboratory, and for all the invaluable guidance, support and encouragement he has given me during my PhD. I am also very grateful to Dr. Dennis Linton for his constructive comments and advice throughout my doctoral training.

I would like to thank Dr. James Thompson for sharing his wealth of expertise in membrane proteins and providing invaluable technical advice, and for his help with constructing a theoretical model of KpsD. I also gratefully acknowledge Dr. Thomas Hudson for providing several pBAD28 plasmids, which served as key starting points for my research, as well as for his technical guidance during the first months of my PhD.

A big thank you goes to all members, past and present, of the Manchester Microbiology group for their unlimited support during the past four years. In particular, I would like to thank Alison, Julie, Tom, Warren, Ashley, Sean, Marie, Jane, Steffi, Helen, and Sharaz for their continuous encouragement as well as banter in and outside the laboratory, and for their friendship. I could not have done it without you.

Finally, I would like to express my deepest gratitude to my family for their unwavering support and continued belief in me.

This research project was funded by the Biotechnology and Biological Sciences Research Council, UK.
This work is dedicated to my grandmother Rosa
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ω</td>
<td>Ohm, unit of electrical resistance</td>
</tr>
<tr>
<td>A5P</td>
<td>Arabinose 5’-phosphate</td>
</tr>
<tr>
<td>ABC</td>
<td>Adenosine 5’-triphosphate-binding cassette</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine 5’-diphosphate</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5’-triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>Complement protein</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>Calcium chloride</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>CMP</td>
<td>Cytidine 5’-monophosphate</td>
</tr>
<tr>
<td>C-terminus</td>
<td>Carboxyl or COOH-terminus</td>
</tr>
<tr>
<td>DC-SIGN</td>
<td>Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Distilled water</td>
</tr>
<tr>
<td>DM</td>
<td>n-Decyl-β-D-maltopyranoside</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide triphosphates</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>F</td>
<td>Farad, unit of capacitance</td>
</tr>
<tr>
<td>g</td>
<td>G-force</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcA</td>
<td>Glucuronic acid</td>
</tr>
<tr>
<td>GlcN</td>
<td>Glucosamine</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetylglucosamine</td>
</tr>
<tr>
<td>GTPase</td>
<td>Guanosine 5’-triphosphatase</td>
</tr>
<tr>
<td>HA</td>
<td>Hyaluronic acid</td>
</tr>
<tr>
<td>Hib</td>
<td><em>H. influenzae</em> serotype b</td>
</tr>
<tr>
<td>his</td>
<td>Histidine</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>Ig</td>
<td>Immunoglobulin</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thio-galactoside</td>
</tr>
<tr>
<td>JUMPstart</td>
<td>Just upstream of many polysaccharide starts motif</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilodalton</td>
</tr>
<tr>
<td>Kdo</td>
<td>3-deoxy-D-manno-octulosonic acid</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>Monopotassium phosphate</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LOS</td>
<td>Lipooligosaccharide</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>Man</td>
<td>Mannose</td>
</tr>
<tr>
<td>ManNAc</td>
<td>N-acetylmannosamine</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>Magnesium sulphate</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino) propanesulfonic acid</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium chloride</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>Disodium hydrogen phosphate</td>
</tr>
<tr>
<td>Na₂SO₄</td>
<td>Sodium sulphate</td>
</tr>
<tr>
<td>NDP</td>
<td>Nucleotide diphosphate</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>Ammonium chloride</td>
</tr>
<tr>
<td>NBD</td>
<td>Nucleotide-binding domain</td>
</tr>
<tr>
<td>NeuNAc</td>
<td>N-acetylneuraminic acid</td>
</tr>
<tr>
<td>NMP</td>
<td>Nucleotide monophosphate</td>
</tr>
<tr>
<td>N-terminus</td>
<td>Amino or NH₂-terminus</td>
</tr>
<tr>
<td>OD₆₀₀</td>
<td>Optical density measured at wavelength of 600 nm</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCP</td>
<td>Polysaccharide co-polymerase</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PES</td>
<td>Polysaccharide export sequence</td>
</tr>
<tr>
<td>PFO</td>
<td>Perfluoro-octanoic acid</td>
</tr>
<tr>
<td>psi</td>
<td>Pounds per square inch</td>
</tr>
<tr>
<td>PVDF</td>
<td>Polyvinylidene difluoride</td>
</tr>
<tr>
<td>qBBBr</td>
<td>Monobromo(trimethylammonio)bimane bromide</td>
</tr>
<tr>
<td>RbCl₂</td>
<td>Rubidium chloride</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>Ru5P</td>
<td>Ribulose 5’-phosphate</td>
</tr>
<tr>
<td>SDM</td>
<td>Site-directed mutagenesis</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulphate-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>T1SS</td>
<td>Type 1 secretion system</td>
</tr>
<tr>
<td>TAE</td>
<td>Buffer made of Tris, acetic acid and EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>Buffer made of Tris and EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N,N’-tetramethylene-diamine</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco etch virus</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane</td>
</tr>
<tr>
<td>TNF</td>
<td>Tumour necrosis factor</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>U</td>
<td>Enzyme activity unit</td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine 5’-diphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume of solute per volume of solvent</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight of solute per volume of solvent</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1 Bacterial polysaccharides

The expression of polysaccharides is important for bacterial physiology. Polysaccharides are high molecular weight carbohydrates made of sugar units linked by glycosidic bonds. Bacterial polysaccharides are either incorporated into the cell membrane, attached to the bacterial surface or excreted into the environment. Peptidoglycan, teichoic acids, lipopolysaccharides (LPS) and lipooligosaccharides (LOS) are all crucial structural components of the bacterial cell. Some bacteria also excrete extracellular polysaccharides or express them on their cell surface, and thus interact with the outside environment via these macromolecules. In addition to providing structural support for the bacterial cell, polysaccharides can prevent bacterial desiccation and loss of nutrients. The two types of surface-associated polysaccharides can be classified into exopolysaccharide slime - which is loosely linked to the cell – and polysaccharide capsules – which have tight association with the bacterial cell surface.

This introduction provides an overview of the role of capsular polysaccharides in pathogenic bacteria and then discusses the genetic organisation, synthesis and export of capsular polysaccharides of Escherichia coli in detail. The remainder of the review describes capsules expressed by other bacteria (with focus on those exported via an adenosine 5’-triphosphate-binding cassette (ABC) transporter-dependent pathway) and discusses parallels between the ABC transporter-dependent capsule export in E. coli and the bacterial type 1 secretion system (T1SS) which excretes proteins.

1.2 Capsular polysaccharides in pathogenic bacteria

The expression of polysaccharide capsules is common in Gram-negative and Gram-positive bacteria. The capsule forms a fibrous matrix around the bacterial cell surface; the average thickness of Gram-negative capsules varies from 300 nm (E. coli K29) to 10 nm (E. coli K1) (Amako et al., 1988). Capsular polysaccharides (K antigens) display a vast structural diversity and vary by the nature of their monosaccharide units, the number of saccharide repeats and by their diverse glycosidic linkage (Roberts, 1996). The capsule has diverse functions, including the prevention of desiccation, as well as advantages in host colonisation and evasion of the host immune system in pathogenic bacteria.
There is strong evidence that the capsule is a virulence determinant in many bacterial pathogens and plays a role in the pathogenesis of disease. In extra-intestinal pathogenic *E. coli* expressing the K1 capsule, the wild-type can outcompete the acapsular mutant in colonising the kidney and liver of mice in an experimental model of urinary tract infection (Buckles *et al.*, 2009). In comparative infection studies in mice *Haemophilus influenzae* expressing capsule serotypes a, b, or c showed much higher potential to invade the blood stream and meninges than their acapsular variants (Moxon and Vaughn, 1981). In a mouse model of pneumonia, only encapsulated *Klebsiella pneumoniae* were able to cause infection, while their acapsular mutants were avirulent (de Astorza *et al.*, 2004). Similarly, capsule expression is crucial for the development of meningitis caused by *Streptococcus pneumoniae* in mice (Peppoloni *et al.*, 2010). And in *S. pyogenes* the capsule is an important determinant for causing necrotic lesions in mice (Ashbaugh *et al.*, 1998b). Thus, expression of a capsule appears to provide some pathogenic bacteria with a distinct advantage above their acapsular variants.

Regulation of capsule expression is an important factor in the ability to cause virulence in some pathogens, and phase variation is a common occurrence in encapsulated pathogens such as *H. influenzae* and *Neisseria meningitidis*. In the discussion below it will become apparent that up- and down-regulation of capsule expression at the right times during infection is crucial for successful invasion and persistence.

### 1.2.1 Pathogen-pathogen interactions: the role of capsules in biofilm formation

Biofilm formation can play an important role in bacterial infection (see Peters *et al.* (2012) for a review). Evidence from several encapsulated bacterial species suggests that the expression of a capsule inhibits biofilm formation. Encapsulated *S. pneumoniae* and *E. coli* K1 were less successful biofilm formers than their acapsulated variants (Schembri *et al.*, 2004; Munoz-Elias *et al.*, 2008), and strains deficient in *kpsM*, a gene essential for capsule expression, showed an increased ability to form biofilms compared to the encapsulated wild-type *E. coli* (Ong *et al.*, 2008). This suggests that the expression of a capsule either inhibits expression of surface molecules required for biofilm formation or that the capsule somehow interferes with the function of these surface molecules. In *K. pneumoniae* the adhesive properties of type 1 fimbriae, which are important for the initial stage of biofilm formation by adhering to surfaces, were lost in
the presence of a capsule (Schembri et al., 2005). However, the amount of cell-surface type 1 fimbriae was not reduced in encapsulated strains (Schembri et al., 2005), suggesting the capsule might physically interfere with type 1 fimbriae during the initial stage of biofilm formation. Another study demonstrated that the ability of the capsule to inhibit cell-surface interactions during the initial stages of biofilm formation is induced by the additional release of soluble polysaccharide into the environment, resulting in the modification of the physiochemical properties of the surface (Valle et al., 2006).

The role of the capsule in the subsequent steps of biofilm formation, the biofilm maturation, is less clear. It was reported that mutants lacking basA, a gene crucial for capsule expression in S. pyogenes, were able to make initial cell-surface contacts but failed to form cell-cell aggregates required for biofilm maturation (Cho and Caparon, 2005). This suggested that the capsule is crucial for biofilm maturation. However, a study on the self-recognising adhesin antigen 43, which promotes cell-cell agglutination during biofilm maturation, suggests an inhibitory role of capsules during the maturation stage. It was demonstrated that the presence of the K1 capsule of E. coli inhibits the cell-cell interactions mediated by antigen 43 molecules (Schembri et al., 2004). A further study confirmed that aggregation through antigen 43, as well as conjugative pili, the adhesive fimbriae curli and other adhesins, was inhibited by capsule expression in E. coli (Valle et al., 2006). It was proposed that the inhibitory effect of capsules might be due to shielding of the other surface molecules. These contradictory results might either reflect the different techniques employed to analyse biofilm maturation, or account for the varied role of capsular polysaccharides in different species. However, it also suggests that expression of capsule must be tightly regulated to ensure advantageous coordination of capsule expression with the expression of other cell-surface antigens. Indeed, analysis of capsule gene expression in S. pneumoniae revealed that the cpsA gene, which is important for capsule expression, is down-regulated in biofilm forming cells compared to the planktonic variants (Hall-Stoodley et al., 2008).

1.2.2 Pathogen-host interactions: the role of capsules in host cell adhesion

Adherence to host cells is a crucial step in the infection by many encapsulated bacterial pathogens, and there has been extensive research about the interaction between host
cells and polysaccharide capsules. Research on encapsulated *K. pneumoniae* has shown that capsule-expressing variants adhere better to mucus-producing cells but have a reduced ability to attach to epithelial cells (Favre-Bonte *et al.*, 1999). In concordance with the studies on biofilms discussed in section 1.2.1, the majority of literature reports that expression of a capsule reduces the ability of pathogens to adhere to epithelial cells. Studies on *K. pneumoniae*, *N. meningitidis*, *H. influenzae* and *S. pyogenes* showed that the encapsulated wild-types adhere less efficiently to epithelial cells than their capsule-deficient variants (St Geme and Falkow, 1991; Hammerschmidt *et al.*, 1996; Cywes *et al.*, 2000; Sahly *et al.*, 2000; Cortes *et al.*, 2002; Struve and Krogfelt, 2003; Hammerschmidt *et al.*, 2005; Spinosa *et al.*, 2007). Similar to the evidence from biofilm studies, it was proposed that adherence of cell-surface adhesins to epithelial cells is masked by the presence of capsular polysaccharides. Capsule expression by *N. meningitidis* significantly reduced the ability of pili to adhere to epithelial cells and abolished adhesion by adhesin Opc completely (Virji *et al.*, 1995).

These results raise questions about the role of capsules as virulence factors and their ability to adhere to host cells. However, there is evidence that exactly this inhibition of adhesion is used by encapsulated pathogens to evade the host immune system. As well as adhering more efficiently to epithelial cells, acapsulated variants have also been shown to be more easily internalised by epithelial cells (Hammerschmidt *et al.*, 1996; Sahly *et al.*, 2000; Cortes *et al.*, 2002; Struve and Krogfelt, 2003; Spinosa *et al.*, 2007). It was shown that capsule-deficient *K. pneumoniae* mutants exhibited a 10-fold increase in adhesion to and internalisation into epithelial cells than encapsulated wild-type and were avirulent in a mouse model of pneumoniae, compared to the highly virulent wild-type (Cortes *et al.*, 2002). This suggests that internalisation of bacteria is a host response to clear infection.

Studies on *N. meningitidis* and *S. pyogenes* provide more evidence. Scanning electron microscopy demonstrated that all *S. pyogenes* cells internalised by epithelial cells were acapsulated, while the extracellular environment contained mainly encapsulated variants (Hammerschmidt *et al.*, 2005). The expression profile of *S. pyogenes* capsule during infection appears to depend on the location of the capsule and its association of the capsule with epithelial cells. While bacteria in immediate contact with host cells lost their capsule, bacteria not directly interacting with epithelial cells expressed a capsule
(even if they were associated to capsule-deficient cells in a chain) (Hammerschmidt et al., 2005). This suggests that mechanisms which allow a reversible variation of capsule expression during infection are crucial for the virulence of encapsulated pathogens. A study on *N. meningitidis* found that the capsule was primarily expressed by bacteria located extracellularly from epithelial cells, rather than by intracellular bacteria (Hammerschmidt et al., 1996). In contrast, analyses of *N. meningitidis* encapsulation within human brain microvascular endothelial cells (Nikulin et al., 2006) and HeLa (Spinosa et al., 2007) cells suggest that the meningococcal capsule can be expressed inside host cells to aid immune evasion (cf. section 1.2.3.1). Taken together, these findings illustrate the role for a reversible on/off switching of capsule expression and might, in addition, imply that the type of host cell can affect the strategy for capsule phase variation.

### 1.2.3.1 Intracellular survival

There is evidence that the capsule enhances the ability of some bacterial pathogens to survive inside host cells. This strategy can be important for pathogens in order to translocate to areas of the host where natural immune responses are limited, such as across the blood-brain barrier. Encapsulated *N. meningitidis* serogroup B can reside and...
replicate within a protective vacuole in human brain microvascular endothelial cells, while acapsular mutants are killed (Nikulin et al., 2006). Similarly, experiments on E. coli K1 demonstrated that encapsulated bacteria residing inside endosomes of human brain microvascular endothelial cells modulate the endosome maturation process, preventing endosome-lysosome fusion and subsequent bacterial killing by endothelial cells (Kim et al., 2003). The same mechanism has been proposed for encapsulated S. pneumoniae in brain macrophages. The expression of a capsule by S. pneumoniae confers resistance to intracellular killing within microglial cells (Peppoloni et al., 2010). Capsule-specific gene transcription levels were elevated during replication of S. pneumoniae inside microglial cells, compared to extracellular replication (Peppoloni et al., 2010). Enhanced intracellular survival and replication within macrophages was also reported for encapsulated H. influenzae serotype b (Williams et al., 1991) and Salmonella enterica serovar Typhi (Hirose et al., 1997). In S. Typhi the ability for survival inside macrophages has been attributed to the suppression of the release of pro-inflammatory tumour necrosis factor (TNF)-α by capsule expression (Hirose et al., 1997).

A different strategy to aid intracellular survival has been described in N. meningitidis serogroup B infecting HeLa cells, a human epithelial carcinoma cell line. Acapsulated meningococci rapidly decreased in numbers inside HeLa cells, while encapsulated variants could recover and multiply (Spinosa et al., 2007). In contrast to acapsulated meningococci, bacteria expressing a capsule showed a higher level of resistance to cationic antimicrobial peptides such as protegrins, which promote bacterial cell lysis by pore-formation. As in S. pneumoniae, capsule-specific genes were up-regulated during intracellular growth of N. meningitidis (Spinosa et al., 2007). Yet another mechanism for intracellular survival has been reported, which is associated with similar mechanisms as biofilm formation described in section 1.2.1. A study on bladder epithelial cells infected by E. coli K1 and its acapsular mutant demonstrated that the wild-type formed tightly-aggregated intracellular bacterial communities, while the mutant was detected in dispersed formations (Anderson et al., 2010). These intracellular bacterial communities showed biofilm-like properties, and as aggregation was capsule-dependent it was suggested that the capsule acts as matrix component to form the multicellular structure. Neutrophil tracking experiments provided evidence that formation of these bacterial communities increases inhibition of neutrophil killing (Anderson et al., 2010).
1.2.3.2 Modulation of immune mediators and immune cell recruitment

In addition to *E. coli* K1, other encapsulated pathogens such as *N. meningitidis* serogroup B, *K. pneumoniae* and *E. coli* K29 (group 1) have also been reported to display enhanced resistance to phagocytic killing by neutrophils or dendritic cells (Horwitz and Silverstein, 1980; Unkmeir *et al.*, 2002; Regueiro *et al.*, 2006; Evrard *et al.*, 2010). There is evidence that in some encapsulated strains this observed resistance is associated with the ability of capsules to modulate the recruitment and activation of immune cells by modulating the mediators which are responsible for immune cell recruitment and activation. Compared to its acapsular mutant, the uptake of encapsulated *K. pneumoniae* by dendritic cells is associated with a decrease in the down-regulation of the expression of dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DCSIGN) on dendritic cells and a decrease in pro-inflammatory cytokines interleukin (IL)-12 and TNF-α and anti-inflammatory cytokine IL-10 (Evrard *et al.*, 2010). This suggests that encapsulated *K. pneumoniae* can interfere with maturation of dendritic cells and suppress the secretion of cytokines from dendritic cells.

Expression of the Vi capsule by *S. Typhi* reduces the recruitment of neutrophils and natural killer cells (Jansen *et al.*, 2011). It was demonstrated that encapsulation decreases the release of pro-inflammatory mediators TNF-α and interferon (IFN)-γ, while leading to an increase in the anti-inflammatory cytokine IL-10 (Jansen *et al.*, 2011). The direct interaction of encapsulated *S. Typhi* with epithelial cells has been identified as a further strategy of the pathogen to suppress the immune system, as this interaction inhibits release of pro-inflammatory cytokine IL-8 (Sharma and Qadri, 2004). Indeed, another study confirmed that encapsulated *S. Typhi* impaired Toll-like receptor-mediated production of IL-8 from epithelial cells compared to acapsular mutants (Raffatellu *et al.*, 2005). In addition it was shown that the *S. Typhi* capsule interferes with macrophage-like cells and down-regulates the secretion of pro-inflammatory IL-17(Raffatellu *et al.*, 2007). In *K. pneumoniae*, capsule expression can reduce the surface expression of cluster of differentiation (CD) 54 by epithelial cells and the levels of IL-8 released by these cells, and thus inhibits the adhesion of neutrophils to infected epithelial cells as well as the recruitment of new neutrophils (Regueiro *et al.*, 2006).
1.2.3.3 **Resistance to the complement pathway**

Evidence from a number of bacterial species suggests that capsules can confer resistance to complement. The complement system is a complex system made of serum, enzymes and cell-surface proteins which produce effectors of both innate and adaptive immune response, and, in turn, results in a cascade of opsonisation, phagocytosis, complement-mediated lysis, and stimulation of inflammatory responses. Activation of the complement system occurs via either of three routes: the classical pathway (which is activated by antibodies), the alternative pathway (which is activated by microbial surface molecules in the absence of antibodies), and the lectin pathway (which is activated by the binding of plasma lectin to microbial mannose (Man); this pathway will not be of further relevance here). All three routes result in the generation of variants of complement protein 3 (C3)-convertase, which cleave C3 into C3a and C3b complement fragments. C3a has chemotactic and histamine-releasing activity, recruiting phagocytes and histamine-releasing leukocytes to the site of infection. C3b is a principal complement opsonin. It induces bacterial uptake by phagocytes by binding to the bacterial cell surface (complement deposition) or by binding to antigen-bound antibody. In subsequent steps complement component C5 initiates the formation of pore-forming membrane-attack complexes (made of C5b, C6, C7, C8 and C9) integrated into the membrane of host cells or bacterial target cells, resulting in bacterial killing.

The ability to resist complement-mediated immune response, including the ability to resist a specific complement pathway, is capsule-specific. For instance, in *S. pneumoniae* some capsule types confer complement-resistance, while others do not, and it has been demonstrated that capsule type is a more crucial determinant for resistance to complement than the bacterial genetic background (Melin *et al.*, 2010). The majority of literature has identified the alternative pathway as target for resistance by encapsulated bacteria. It was demonstrated that encapsulated variants resisted the bacteriocidal actions activated by the alternative pathway, while acapsular *N. meningitidis* serogroup B were killed (Jarvis and Vedros, 1987). Equally, the K54 capsule expressed by *E. coli* confers resistance to the alternative pathway (Russo *et al.*, 1993). Similarly, *E. coli* either expressing the K1 or K5 capsule display resistant to the alternative but not to the classical pathway (Pluschke *et al.*, 1983; Burns and Hull, 1998).
Inhibition of the actions of complement components C3 and C3b has been identified as one mechanism by which encapsulated bacteria resist the complement pathway. This mechanism has been mainly associated with inhibiting the alternative route, but has also been described as classical pathway-related in some cases. In a study on meningococcal interaction with C3, C3 was found to bind in higher efficiency to acapsular meningococci than the encapsulated variant, and an increase in binding efficiency directly correlated to a down-regulation in capsule expression (Jarvis and Vedros, 1987). The same relationship between C3 binding and capsule expression was found in *Staphylococcus aureus* (Cunnion et al., 2001). Encapsulated *K. pneumoniae* and *S. Typhi* also bind less efficiently to C3 than their acapsular variants (Cortes et al., 2002; Wilson et al., 2011). Similarly, quantifying C3b deposition on bacterial surfaces identified a higher amount of bound C3b on acapsular than on encapsulated *S. pneumoniae* serotypes 2 and 4 (Hyams et al., 2010). Acapsular variants were increasingly phagocytosed by neutrophils as a result of enhanced complement activation (Hyams et al., 2010). Further evidence from *S. pneumoniae* serotypes 7 and 12 implies that the capsule prevents the recognition of pneumococcal-bound C3b by C3b-receptors on phagocytic cells (Brown et al., 1983).

A study on *E. coli* K1 identified a different mechanism which modulates the complement cascade. Encapsulated but not acapsular *E. coli* bound to the C4b-binding protein (Wooster et al., 2006). This protein acts a co-factor for factor I-mediated cleavage of C3b and C4b. Binding of the C4b-binding protein by K1 resulted in the degradation of C3b and C4b, consequently inhibiting opsonisation and killing via membrane-attack complexes (Wooster et al., 2006).

*N. meningitidis* and *S. pneumoniae* are both able to inhibit the classical complement-pathway. Acapsular *N. meningitidis* showed an increased ability to bind immunoglobulin (Ig) A than encapsulated variants and the presence of a capsule was required for IgA-dependent activation of the classical pathway (Vogel et al., 1997). In *S. pneumoniae* serotypes 2 and 4 the capsule conferred resistance to both IgG-dependent and IgG-independent activation of the classical pathway (Hyams et al., 2010). This mechanism of complement evasion might be associated with masking of bacterial antibody-receptors by capsule expression.
Although some species can evade the classical pathway, the majority of encapsulated bacteria are unable to modulate the complement pathway in the presence of antibodies. As discussed above, encapsulated *S. pneumoniae* serotypes 2 and 4 are resistant to both classical and alternative pathway. However, a study on serotypes 7 and 12 revealed that strains expressing these capsules are sensitive to actions of the classical pathway (Brown *et al.*, 1983). Equally, encapsulated *E. coli* K29 and *N. meningitidis* serogroup B were cleared by phagocytosis after opsonisation with IgM (Horwitz and Silverstein, 1980; Jarvis and Vedros, 1987), while encapsulated *S. Typhi* are opsonised by IgG (Wilson *et al.*, 2011). This suggests that activation of the classical pathway by antibody-complement complexes bound to the capsule surface is an important mechanism in clearing infection by encapsulated pathogens.

In summary, the evidence above indicates that encapsulated bacteria employ a variety of mechanisms to modulate and evade the host immune responses, depending on route of infection, stage of infection, type of capsule and bacterial surface-antigens co-expressed with the capsule.

### 1.3 Polysaccharide capsules of *Escherichia coli*

*E. coli* is the model organism for the study of capsular polysaccharide genetic organisation, regulation, biosynthesis and export in Gram-negative bacteria. Over 80 different capsules types have been identified in *E. coli* and the current classification system distinguishes between four groups (group 1-4), according to genetic and biosynthetic criteria (Whitfield and Roberts, 1999). The four capsule groups differ in characteristics such as gene cluster organisation, thermo-regulation, and polysaccharide synthesis and translocation system. Notably, group 1 and 4 capsular polysaccharides are exported via a Wzy-dependent route, while group 2 and 3 are transported by an ABC transporter-dependent pathway.

### 1.4 Group 1 and 4 capsules of *Escherichia coli*

Both group 1 and 4 capsules are expressed in two different forms: a low-molecular oligosaccharide termed K\textsubscript{LPS} which is attached to the lipid A core of rough LPS, and the high-molecular LPS-independent K antigen which forms a proper capsule (Maclachlan
et al., 1993). K_{LPS} are structurally identical to O polysaccharides (O-antigens) of LPS. Capsular K antigens and K_{LPS} of the same strain carry identical saccharide repeat units and share the same enzymes for synthesis (Drummelsmith and Whitfield, 1999). However, their final step during substrate translocation differs (Dodgson et al., 1996).

Group 4 capsules differ from group 1 capsules by the presence of amino acids or amino sugars as part of their polysaccharide repeat structure. Research on group 4 capsules has been limited but as they share common genes with group 1, the mechanisms for synthesis and export are predicted to be parallel. K40 is the best studied group 4 polysaccharide and has the structure of 1,4-α-D-N-acetylglucosamine (GlcNAc)-1,4-β-D-glucuronic acid (GlcA)-1,6-α-D-GlcNAe, with a serine residue attached to position 6 of GlcA (Dengler et al., 1986). In group 4, two different gene clusters are essential for capsule expression (Figure 1.1 D). One is identical to region 2 of the group 1 gene cluster and encodes genes wzy, wzx and genes for several glycosyltransferases (Amor and Whitfield, 1997; Amor et al., 1999). In contrast to group 1, where the remaining genes are encoded by region 1, genes wza, wzb and wzc are carried by the so-called “22-minute locus” (or group 4 capsule cluster, G4C cluster), located at minute 22 on the E. coli chromosome (Peleg et al., 2005).

Group 1 capsules have been extensively studied and capsule K30 is the archetype of group 1 polysaccharides. The K30 polymer consists of 1,2-α-D-Man-1,3-β-D-galactose (Gal) chains which carry a 1,3-β-D-GlcA-1,3-α-D-glucose (Glc) branch at position 3 of Man (Chakraborty et al., 1980). K30 of E. coli is identical to the K20 capsule of K. pneumoniae. Expression of group 1 capsules requires a single gene cluster, which is located near the histidine-biosynthesis operon his and O-antigen gene cluster rfb on the bacterial chromosome (Figure 1.1 C). In contrast to group 2 and 3, the cps gene cluster of group 1 only comprises two regions. Region 1 carries genes wzi, wza, wzb, and wzc, encoding components required for polymerisation and is conserved among all group 1 strains. Region 2 is serotype-specific and encodes glycosyltransferase genes (wbaP, wcaN, wcaO, and wbaZ in K30) as well as wzy and wzx. The serotype-specific region is identical to the gene cluster for Wzy-dependent LPS O-antigen synthesis and export. The primary promoter of the capsular gene cluster is located upstream of a regulatory region 5′ of cps known as the JUMPstart (just upstream of many polysaccharide gene starts) motif (i.e., the promoter lies upstream of wzy) (Rahn and Whitfield, 2003). It has been
Figure 1.1 Genetic organisation of the *Escherichia coli* group 1-4 capsular polysaccharide gene clusters. The 1 kb marker gives an indication of the gene sizes within each gene cluster. (A) Genetic and transcriptional organisation of the K5 (group 2) capsular polysaccharide gene cluster. Regions 1 and 3 highlighted in blue-green are conserved among group 2 capsular gene clusters. Genes highlighted in blue are responsible for polysaccharide translocation, and genes in green are involved in polysaccharide modifications. Region 2 highlighted in orange is serotype-specific and encodes genes for polysaccharide synthesis. Broken arrows denote the start sites of transcription on the chromosome. V1 denotes a promoter upstream fragment located between -316 and +9 base pairs relative to the transcriptional start site of region 1. The V1 region acts as binding site of transcriptional regulators SlyA and H-NS. (B) Genetic organisation of the K10 (group 3) capsular polysaccharide gene cluster. Region 1 (blue) and 3 (green) are conserved among group 3 capsular gene clusters. Genes highlighted in blue are involved in polysaccharide translocation, and genes in green are involved in polysaccharide modifications. Region 2 highlighted in orange is serotype-specific and encodes genes for polysaccharide synthesis. Region 2 genes of K10 are unknown. (C) Genetic organisation of the K30 (group 1) capsular polysaccharide gene cluster. Region 1 highlighted in blue is conserved among group 1 capsular gene clusters. Region 2 highlighted in orange is serotype-specific. Region 2 of the K30 gene cluster is identical to the Wzy-dependent O-antigen gene cluster. orf denotes open reading frame with unknown function. (D) Genetic organisation of the K40 (group 4) capsular polysaccharide gene cluster. The “22-minute locus” (or group 4 capsule cluster, G4C cluster) highlighted in light green is conserved among group 4-producing strains. Genes highlighted in blue are involved in polysaccharide translocation, and genes in yellow are uncharacterised genes involved in group 4 polysaccharide synthesis and/or export. The biosynthesis locus highlighted in orange is serotype-specific and carries genes involved in polysaccharide synthesis. orf denotes open reading frame with unknown function. Note that the figure does not represent the true distance between the two loci (indicated by the black bar between the loci).

Please see next page for figure.
Figure 1.1 Genetic organisation of the *Escherichia coli* group 1-4 capsular polysaccharide gene clusters.
established that RfaH-dependent antitermination is required for capsule gene cluster transcription. The sequence of the operon polarity suppressor \textit{ops} (upstream of \textit{cps}) mediates the interaction between RfaH and an RNA polymerase, resulting in read-through transcription from the start to the end of the cluster via a stem-loop structure just upstream of \textit{wbaP} (which decreases transcription in the absence of RfaH) (Rahn and Whitfield, 2003). The \textit{cps} gene cluster for polysaccharide biosynthesis and export in group 1 is homologous to that of \textit{K. pneumoniae}. Additionally, the JUMPstart motif is conserved in the two species (Rahn \textit{et al.}, 1999). The highly conserved capsular cluster organization supports the theory of horizontal gene transfer between encapsulated bacterial species.

\subsection{1.4.1 Group 1 synthesis and export}

The biosynthesis and export of group 1 capsular polysaccharides is an interlinked process (Figure 1.2). The glycosyltransferase WbaP initiates the assembly of polymer repeat units on an undecaprenyl diphosphate lipid acceptor in the cytoplasm close to the cytoplasmic membrane (Drummelsmith and Whitfield, 1999). The assembly of polymer repeats is also likely to involve other glycosyltransferases which utilise nucleotide mono- or di-phosphate-linked sugar donors for transfer. In \textit{E. coli} K30, WbaZ, WcaN, and WcaO are predicted to catalyse the \(\alpha-1,3\) Man-Gal-, \(\beta-D-1,3\) Glc-Gal-, and \(\alpha-D-1,3\) Gal-Man linkages of the polysaccharide repeat units, respectively (Drummelsmith and Whitfield, 1999). Transfer of the repeat units across the cytoplasmic membrane occurs via the integral cytoplasmic membrane protein Wzx. The underlying mechanism is poorly understood but evidence suggests that Wzx ‘flips’ the substrate across the membrane in an adenosine 5’-triphosphate (ATP)-independent manner (Liu \textit{et al.}, 1996). Once on the periplasmic face of the cytoplasmic membrane, the polymer is further synthesised by glycosyltransferase Wzy, which is an integral cytoplasmic membrane protein. The common pathway of K\textsubscript{LPS} and LPS-independent K antigens ends here. While K\textsubscript{LPS} are attached to lipid A core and exported via an LPS-typical route, the high-molecular form is further processed by the Wzy-dependent pathway (Dodgson \textit{et al.}, 1996). The determinants for chain length or export pathway are unknown. However, Wzy might play a role, as Wzy mutants express a shorter K\textsubscript{LPS} version and completely lack LPS-independent capsular K antigen (Drummelsmith and Whitfield, 1999).
Figure 1.2 Current model for *Escherichia coli* group 1 capsular polysaccharide biosynthesis and export. Assembly of polymer repeat units on the undecaprenyl diphosphate lipid acceptor (labelled as lipid) is initiated by glycosyltransferase WbaP. Three additional glycosyltransferases (labelled as GTs; WbaZ, WcaN, and WcaO in *E. coli* K30) are required for further synthesis of the polymer units by transferring nucleotide mono- or di-phosphate (NMP or NDP)-linked sugar donors to the reducing end of the growing chain. Putative flippase Wzx is predicted to translocate the repeat units across the cytoplasmic membrane. Further polymerisation of the periplasmic repeat units requires glycosyltransferase Wzy, tyrosine autokinase Wzc (((P) indicates phosphorylation at the C-terminus), and Wzb, which dephosphorylates Wzc. Outer membrane protein Wza translocates the polysaccharide across the periplasm and outer membrane in interaction with Wzc. Wzi is predicted to be involved in polymer attachment to the cell surface.
In addition to WbaP, Wzx and Wzy, synthesis of the high-molecular form of group 1 polysaccharides also involves the integral cytoplasmic membrane protein Wzc and the cytoplasmic protein Wzb. Wzc and the outer membrane lipoprotein Wza are required for export of the completely synthesised polymer across the periplasm and outer membrane. The role of a second outer membrane protein, Wzi, is also important for capsule expression. Wzi is not required for polysaccharide export (Rahn et al., 1999), however, Wzi-mutants display half the amount of surface-bound polysaccharide, while expressing three times higher levels of excreted polysaccharide than the wild-type (Rahn et al., 2003). This suggests that Wzi is involved in attachment of exported polysaccharide to the cell surface. Wzi is unique to E. coli group 1 and Klebsiella pneumoniae and is likely to act as a monomeric β-barrel (Rahn et al., 2003; Bushell et al., 2010).

1.4.2 Inner membrane protein Wzc and outer membrane protein Wza

Wzc is a polysaccharide co-polymerase protein (PCP) type 2a. Please refer to Figure 1.3 for a phylogenetic tree of Wzc and related PCP proteins. Unique to PCP-2a proteins, Wzc contains ATP-binding Walker A and B motifs as well as a tyrosine-rich cluster at its carboxyl (C)-terminal domain (Wugeditsch et al., 2001). Indeed, it was experimentally confirmed that Wzc has tyrosine autokinase activity and auto-phosphorylates at the tyrosine-rich cluster and the tyrosine reside 569 at the C-terminus (Wugeditsch et al., 2001). The tyrosine phosphatase Wzb dephosphorylates Wzc (Hagelueken et al., 2009), and thus regulates the phosphorylation activity of Wzc. Mutagenesis experiments demonstrated that the Walker A motif as well as the phosphorylation of the tyrosine-rich cluster are essential for Wzc function and correct polysaccharide expression (Wugeditsch et al., 2001). The dephosphorylation of Wzc by Wzb is also required for polysaccharide expression (Wugeditsch et al., 2001). This suggests that Wzc is required in both phosphorlylated and non-phosphorylated forms during capsule assembly and indicates that Wzc might cycle between these two states with the aid of Wzb-regulation. Site-directed mutagenesis experiments of tyrosine 569 and the tyrosine residues of the tyrosine-rich cluster highlight that phosphorylation and functional capsule assembly does not require the presence of specific tyrosine residues but rather relies on a crucial level of phosphorylation (Paiment et al., 2002). Even though the autokinase-specific regions are located in the C-terminal domain of the protein, the complete protein is
Figure 1.3 Phylogenetic tree of selected polysaccharide co-polymerases and outer membrane polysaccharide export proteins. Multiple sequence alignments of complete protein sequences were generated using ClustalW (Thompson et al., 1994). The trees were constructed using ProDist (neighbour-joining method) of PHYLIP, version 3.5c (Felsenstein, 1993). The bar indicates distance of branch lengths, expressed in arbitrary units. (A) Phylogenetic tree of polysaccharide co-polymerases (PCP). All proteins belong to the PCP-3 family, except Wzc which belongs to the PCP-2 family. PCP-3 KpsE of *E. coli* K5 (gene bank accession number P62586), PCP-2a Wzc of *E. coli* K30 (Q9X4B9), PCP-3 BexC of *H. influenzae* serotype b (P22930), PCP-3 CtrB of *N. meningitidis* serogroup B (P32014), and PCP-3 VexD of *S. Typhi* (P43111). (B) Phylogenetic tree of outer membrane polysaccharide export proteins. KpsD of *E. coli* K5 (P42213), Wza of *E. coli* K30 (Q9X4B7), BexD of *H. influenzae* serotype b (P22236), CtrA of *N. meningitidis* serogroup B (P0A0V8), and VexA of *S. Typhi* (Q04976).
required for phosphorylation activity of Wzc (Wugeditsch et al., 2001). A further study identified Wzc as bifunctional enzyme, as it was discovered that Wzc also has ATPase activity, converting ATP to adenosine 5'-diphosphate (ADP) and free phosphate (Soulat et al., 2007). The role of ATP hydrolysis in Wzc function and capsule assembly was not studied, however, the dual function of Wzc is likely to play a role.

Wzc membrane topology was elucidated using lacZ and phoA fusions. Typical for PCP proteins, the N- and C-terminal of Wzc are located in the cytoplasm, with two transmembrane helices on either side of a periplasmic domain (Doublet et al., 2002). The amino (N)-terminal region is short, while the C-terminal carrying the tyrosine autokinase regions is considerably longer. In vivo cross-linking experiments indicated that Wzc can act as trimer, and possibly a hexamer (Doublet et al., 2002). Neither the presence of the tyrosine-rich cluster nor autophosphorylation is required for Wzc oligomerisation. However, the C-terminal domain is essential for the ability of Wzc to form oligomers (Doublet et al., 2002), suggesting an important structural role for the C-terminal region unrelated to enzyme function. Later studies utilising cryo-electron microscopy to study the three-dimensional structure of Wzc strongly suggests that Wzc acts as a tetramer rather than trimer (Collins et al., 2006; Collins et al., 2007). However, the oligomerisation state of Wzc is far from clear, as the crystal structure of the cytoplasmic C-terminal domain in non-phosphorlyated form suggests that this domain forms a ring-shaped octamer (Bechet et al., 2010). Site-directed mutagenesis of several amino acids identified a conserved EX2RX2R motif between residues 508-514 which is important for wild-type folding and thus stability of the oligomeric state of the C-terminal domain. There is evidence that this region is required for wild-type phosphorylation levels (Bechet et al., 2010). The study further identified a conserved RK-cluster between residues 473 and 496 to be important for capsule assembly, suggesting it might be involved in intra- or inter-protein interactions (Bechet et al., 2010). Further three-dimensional analysis of Wzc in high-resolution are required to provide a clearer picture of the oligomeric state of Wzc.

Wza is the outer membrane component of the Wzy-dependent translocation system. Please refer to Figure 1.3 for a phylogenetic tree of Wza and related polysaccharide outer membrane exporters. Wza shares the conserved polysaccharide export sequence (PES) motif (cf. Figure 1.4) with other outer membrane polysaccharide export proteins.
(Dong et al., 2006). An electron microscopy study on crystallised lipid-reconstituted protein provided the first three-dimensional structural analysis of Wza (Nesper et al., 2003). Wza was identified as octamer forming a ring-like structure with a central cavity. Negative-staining cryo-electron microscopy confirmed the octameric state of Wza, but visualised a 4-fold rotational symmetry in contrast to an 8-fold symmetry identified by electron crystallography (Nesper et al., 2003). After further crystallography studies of Wza confirmed the 8-fold rotational symmetry (Figure 1.4 (Dong et al., 2006)). The discrepancies concerning the oligomeric state were attributed to a possibly effect of negative staining on the native conformation of the protein.

The crystal structure of Wza constitutes four rings (octameric rings 1-4, monomeric domains 1-4) forming a central cavity which is closed at the periplasmic end and opens into a narrow neck in the outer membrane (Dong et al., 2006). As previously shown by electron crystallography, the ring structure consists of eight monomers. The overall structure of Wza was later confirmed by further cryo-electron microscopy analyses (Ford et al., 2009). Domain 1 (residues 89-169) contains the conserved PES motif and forms an anti-parallel β-sandwich with an α-helix on one side. Domain 2 (residues 68-84, 175-252) forms a mixed β-sheet made of five strands and three α-helices on one side. Domain 3 (residues 46-64, 255-344) is a structural duplication of domain 2. Domain 4 (residues 345-376) is the C-terminal region and forms an amphipathic α-helix. In octameric Wza the eight domains of domain 4 form an α-helical barrel in the outer membrane, sitting on top of the three periplasmic rings (Dong et al., 2006). The formation of a helical barrel in the transmembrane region makes Wza an unusual outer membrane protein, as most outer membrane proteins, including TolC, form a β-barrel. Domain 4 is indeed important for Wza oligomerisation and function. A truncation mutant lacking the last 35 residues at the C-terminus was unable to form octamers and was impaired in polysaccharide export (Ford et al., 2009).

Cross-linking experiments suggested that Wzc and Wza interact to form a protein complex (Nesper et al., 2003). Transmission electron microscopy visualised the Wzc-Wza complex three-dimensionally (Figure 1.5 (Collins et al., 2007)). As predicted, the complex spans both membranes, reaching from the cytoplasm to the cell-surface. The structure was described as resembling an elongated molar tooth with a central cavity in the crown at the top (Wza) and four roots at the bottom (Wzc). The two proteins are
Figure 1.4 Crystal structure of Wza monomer and octamer of *Escherichia coli* K30. The crystal structure of the Wza monomer (left) shows the protein’s four domains (D1-4). Domain 1 (residues 89-169) forms an anti-parallel β-sandwich with an α-helix on one side and contains the PES domain. Domain 2 (residues 68-84, 175-252) and domain 3 (residues 46-64, 255-344) are structural duplications and both form a mixed β-sheet made of five strands and three α-helices on one side. Domain 4 (residues 345-376) forms an amphipathic α-helix. The octamer of Wza (right) is composed of rings 1-4 (R-R4). Rings 1-3 are located in the periplasm and form a central cavity, while ring 4 is inserted into the outer membrane and forms an α-helical membrane barrel. Reprinted with permission from Macmillan Publishers Ltd. on behalf of Cancer Research UK (Nature, Dong *et al.* (2006), p.227, copyright 2012).
Figure 1.5 Three-dimensional reconstruction of the Wzc-Wza complex of *Escherichia coli* K30. This figure visualises the three-dimensional structure of the Wzc-Wza complex based on electron microscopy analysis. The complex has the proposed structure of an elongated molar tooth, with tetrameric Wzc forming the roots and octameric Wza forming the crown of the structure. A vertical cut through the complex reveals the central cavity and horizontal cuts illustrate the proposed 4-fold and 8-fold rotational symmetry of Wzc and Wza, respectively. Adapted with permission from National Academy of Sciences (Proceedings of the National Academy of Sciences of the United States of America, Collins *et al.* (2007), p. 2392, copyright 2007, National Academy of Sciences, U.S.A.).
thought to interact via their periplasmic domains but details remain unclear. This structural analysis provides the basis of a model for polysaccharide translocation. It is anticipated that the synthesised polymer somehow interacts with Wzc, which then binds to Wza. Upon interaction with Wzc, Wza then changes its conformation of the central cavity from a closed to an open state, translocating the polymer through the periplasm and outer membrane. However, new data on the crystal structure of cytoplasmic C-terminus of Wzc (Bechet et al., 2010), discussed above might propose a different export process. Here, the polymer is not flipped across the membrane by Wzx but enters a cavity made by the ring-like structure of Wzc, which enables translocation across the cytoplasmic membrane. The continuous translocon complex formed by Wzc-Wza would then allow transport of the polymer across the periplasm and outer membrane. Further studies on the three dimensional structure of Wzc, the mechanisms of Wzx, as well as on the interaction of the capsule proteins with the polymer are required to elucidate the transport process.

1.5 Group 2 capsules of *Escherichia coli*

K1 and K5 capsules are the best studied capsules in group 2 and are associated with extra-intestinal infections. Both K1 and K5 are frequently found in patients with sepsis and urinary tract infections (Cross et al., 1984; Kaijser and Jodal, 1984; Devine et al., 1989). K1 strains are also a common cause of neonatal meningitis, with 80% of cases being associated with K1-expressing strains (Robbins et al., 1974; Cross et al., 1984).

Unlike group 1, 3 and 4 capsules, the expression of group 2 polysaccharides is thermoregulated. Group 2 capsules are not expressed below 20°C (Simpson et al., 1996; Rowe et al., 2000; Navasa et al., 2011). As in group 3, group 2 polysaccharides are transported via an ABC transporter-dependent system. Figure 1.1 A illustrates the organisation of the K5 capsular gene cluster as an example for a group 2 cluster. Like the group 3 cluster, the *kps* cluster of group 2 is located near *serA* on the bacterial chromosome (Vimr, 1991) and is structured into three regions. Region 1 and 3 are conserved among group 2-expressing strains and encode genes responsible for capsule assembly and export. Region 2 is serotype-specific central region, which is flanked by region 1 and 3 on either side and carries genes for polysaccharide biosynthesis.
Transcription of \( kps \) genes is driven by two temperature-regulated promoters located upstream of regions 1 and 3 (cf. Figure 1.1 A (Cieslewicz and Vimr, 1996; Simpson et al., 1996; Stevens et al., 1997; Rowe et al., 2000)). The polycistronic message resulting from transcription from region 1 promoter covers all region 1 genes. In addition, a \( kpsS \)-specific transcript is produced by processing of the region 1 transcript (Simpson et al., 1996). Activation of transcription from region 1 promoter is dependent upon binding of the transcriptional regulator Integration Host Factor to the promoter (Rowe et al., 2000). Transcriptional termination of region 1 and 2/3 is likely to occur via a Rho-dependent mechanism, as one Rho-dependent transcriptional terminator was identified within the \( kpsF \) gene and another between \( kpsT \) and \( kfiA \) (Simpson et al., 1996).

The global regulator H-NS and the tyrosine-phosphorylated guanosine 5'-triphosphatase (GTPase) BipA mediate temperature-regulated transcription of the K5 capsular gene cluster from region 1 as well as region 3 promoter (Rowe et al., 2000). While H-NS and BipA enhance \( kps \) transcription at 37°C, they repress transcription at 20°C. In addition, the transcriptional regulator SlyA promotes transcription from the region 1 promoter by binding to two sites of the promoter's V1 region, a promoter upstream fragment (cf. Figure 1.1 A (Corbett et al., 2007)). It was established that H-NS also binds to the V1 region using partially overlapping binding sites, and that H-NS and SlyA can bind to the V1 region at the same time. Furthermore, H-NS is essential for SlyA-regulated activation of transcription (Corbett et al., 2007). These observations suggest that SlyA and H-NS interact, possibly by forming a higher order complex with the DNA to promote RNA polymerase binding to the region 1 promoter, consequently resulting in transcription of region 1 genes.

Transcription from the region 3 promoter results in a large polycistronic mRNA sequence covering both region 3 and 2 genes (cf. Figure 1.1 A (Stevens et al., 1997)). The regulator RfaH, which is also responsible for the regulation of group 1 \( cps \), drives the transcription of region 2 and 3 genes but not of region 1 genes (Stevens et al., 1997). A JUMPstart motif is located upstream of \( kpsM \) (Stevens et al., 1994a; Stevens et al., 1997). This supports the hypothesis of an RfaH-dependent regulation of regions 2 and 3. Indeed, experimental data suggest that this JUMPstart element is essential for complete \( kfiA \) transcription and read-through transcription starting from region 3 promoter and proceeding into region 2 (Stevens et al., 1997). These results strongly
indicate that the JUMPstart sequence promotes anti-termination activity of RfaH, which results in a read-through of transcription into region 2. At 37°C both H-NS and SlyA are required for maximal transcription from region 3 promoter, and, as in region 1, binding of H-NS to regions upstream and downstream of the transcription start site represses transcription from the region 3 promoter (Xue et al., 2009).

### 1.5.1 Group 2 synthesis

#### 1.5.1.1 Synthesis of K1 polysaccharides

The K1 polysaccharide is a poly-α-2,8-linked N-acetylneuraminic acid (NeuNAc) and is structurally identical to the capsular polysaccharide of *N. meningitidis* serogroup B (Steenbergen and Vimr, 1990). The biosynthetic operon of K1 carries the genes *neuDBACES*. K1-degradation assays demonstrated that synthesis and export of K1 polysaccharides occurs within a sub-cellular compartment in the cytoplasm (Steenbergen and Vimr, 2008). Due to structurally identical polysaccharides, *E. coli* K1 and *N. meningitidis* serogroup B express synthesis enzymes which share homology with each other (Table 1.1). NeuC is essential for NeuNAc synthesis and has been identified as uridine 5’-diphosphate (UDP)-GlcNAc 2-epimerase, catalysing the formation of N-acetylmannosamine (ManNAc) from UDP-GlcNAc (Ringenberg et al., 2003; Vann et al., 2004). NeuC is homologous to SiaA of *N. meningitidis* serogroup B. The second step in K1 synthesis involves NeuB – a cytoplasmic sialic acid synthase and homologue of *N. meningitidis* enzyme SiaC. NeuB generates NeuNAc by condensating phosphoenolpyruvate with ManNAc (Annunziato et al., 1995; Vann et al., 1997). This is followed by the activation of NeuNAc to cytidine 5’-monophosphate (CMP)-NeuNAc by putative CMP-NeuNAc synthetase NeuA, a homologue of SiaB of *N. meningitidis* (Edwards and Frosch, 1992; Vimr et al., 2004). NeuS is the sialyltransferase, which uses CMP-NeuNAc as precursors for K1 synthesis and elongates the growing chain by sequentially adding monomer repeats (Steenbergen and Vimr, 1990). This peripheral cytoplasmic membrane protein is homologous to SiaD of *N. meningitidis* (Steenbergen et al., 1992). Mutagenesis studies identified residues in the N-terminus of NeuS for being responsible for the linkage-specificity during chain synthesis (Steenbergen et al., 1992).

Evidence suggests that NeuS interacts with KpsC$_{K1}$ (Steenbergen and Vimr, 2008). This is concordant with the observation that, in addition to the *neu* genes, expression of
<table>
<thead>
<tr>
<th>N. meningitidis group B</th>
<th>Identity (%)</th>
<th>Similarity (%)</th>
<th>E. coli K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiaA</td>
<td>28</td>
<td>43</td>
<td>NeuC</td>
</tr>
<tr>
<td>SiaB</td>
<td>18</td>
<td>25</td>
<td>NeuA</td>
</tr>
<tr>
<td>SiaC</td>
<td>33</td>
<td>44</td>
<td>NeuB</td>
</tr>
<tr>
<td>SiaD</td>
<td>25</td>
<td>37</td>
<td>NeuS</td>
</tr>
</tbody>
</table>

Table 1.1 Homologues of capsular polysaccharide biosynthesis enzymes of *Escherichia coli* K1 and *Neisseria meningitidis* group B. This table details the percentages of identical and similar residues in homologues of *E. coli* K1 and *N. meningitidis* group B. Data of sequence identities and similarities were obtained through analysis of sequence alignments produced by ClustalW (Thompson *et al.*, 1994) using the SIAS Server (Reche, 2008).
KpsS\textsubscript{K1} from region 1 is also important for K1 synthesis (Andreishcheva and Vann, 2006b). While expression of KpsS\textsubscript{K1} is not essential and results in increased K1 being expressed, KpsC\textsubscript{K1} is required for \textit{de novo} synthesis of K1 (Andreishcheva and Vann, 2006b). Interaction between NeuS and KpsC\textsubscript{K1} might be relevant for K1 synthesis but further studies are required to elucidate the specific role of KpsC\textsubscript{K1} in synthesis. NeuE is the CtrG homologue of \textit{N. meningitidis} serogroup B, but, unlike cytoplasmic CtrG, it is a putative cytoplasmic membrane protein (Steenbergen \textit{et al.}, 1992). NeuE is not necessary for synthesis, but NeuE mutants show impaired K1 export (Steenbergen and Vimr, 2008). NeuE might either be involved in linking the synthesis to the export process as proposed for CtrG (Hobb \textit{et al.}, 2010), or might be responsible for NeuS stability or facilitating NeuS function (Steenbergen and Vimr, 2008). Expression of NeuD is essential for synthesis (Daines \textit{et al.}, 2000) and NeuD mutants accumulate NeuNAc and CMP-NeuNAc intracellularly (Annunziato \textit{et al.}, 1995). NeuD belongs to the family of acetyltransferases but its role in K1 synthesis remains unclear (Annunziato \textit{et al.}, 1995; Daines \textit{et al.}, 2000).

KpsF\textsubscript{K1} is not necessary for K1 expression. However, as it is transcribed as part of region 1, it is sensible to assume a role in capsular polysaccharide expression. KpsF\textsubscript{K1} carries a Walker A box which is involved in nucleotide binding as well as a region homologous to glucosamine (GlcN)-binding domains (Cieslewicz and Vimr, 1996). It was suggested that KpsF\textsubscript{K1} might have a regulatory role by binding GlcN to modulate the synthesis of GlcN-containing cell wall components which competes with K1 synthesis for intermediates (Cieslewicz and Vimr, 1996). However, as KpsF\textsubscript{K1} shares 99% of its residues with KpsF\textsubscript{K5}, a similar role during polysaccharide synthesis is also likely. As will be discussed below (section 1.5.1.2), there is evidence that KpsF\textsubscript{K5} is involved in synthesis of 3-deoxy-D-manno-octulosonic acid (Kdo), which is attached to the reducing end of K5 polysaccharides (Finke \textit{et al.}, 1991). Even though no Kdo residue has yet been detected on K1, the close structural resemblance of Kdo and NeuNAc might have made unequivocal identification of Kdo in K1 strains difficult. The hypothesis that KpsF\textsubscript{K1} is involved in Kdo synthesis is supported by the identification of Kdo opolysaccharides produced by \textit{N. meningitidis} serogroup B, which expresses an identical polymer to K1.
1.5.1.2 Synthesis of K5 polysaccharides

A model of the current understanding of K5 synthesis and export is illustrated in Figure 1.6. The K5 polymer is made of the repeated disaccharide structure 4-β-GlcA-1,4-α-N-GlcNAc (Vann et al., 1981). As K1 polymers, K5 polysaccharide synthesis occurs in the cytoplasm close to the cytoplasmic membrane (Finke et al., 1991). However, unlike K1 polysaccharides, K5 polymers are not synthesised in a subcellular compartment (Hudson et al., 2009). The biosynthetic region of K5 encodes the genes kfiABCD. KfiD is a 44 kilodalton (kDa) cytoplasmic membrane-associated protein with UDP-Glc dehydrogenase activity (Petit et al., 1995; Sieberth et al., 1995; Rigg et al., 1998). KfiD converts UDP-Glc into UDP-GlcA. KfiA and KfiC are glycosyltransferases and elongate the growing polymer by the alternate addition of repeat units (Sugiura et al., 2010). KfiA is a 27 kDa cytoplasmic membrane-associated enzyme. It was established that KfiA acts as an α-UDP-GlcNAc glycosyltransferase by catalysing the addition of GlcNAc by means of an α-1,4-glycosidic bond (Hodson et al., 2000; Chen et al., 2006; Sugiura et al., 2010). The activity of KfiA is specific to GlcNAc as the enzyme specifically recognises functional groups of the GlcNAc residues (Chen et al., 2006). KfiC is a 60 kDa cytoplasmic membrane-associated protein (Griffiths et al., 1998; Rigg et al., 1998). The enzyme uses UDP-GlcA to catalyse the addition of GlcA by 4-β-glycosidic bonding. Protein expression studies revealed that enzyme activity of KfiC requires the co-expression of KfiA (Sugiura et al., 2010). Further analysis identified the C-terminal region of KfiA as essential domain for KfiA function and required for the activation of KfiC glycosyltransferase activity (Sugiura et al., 2010).

Further evidence for the complexity of the interactions between proteins involved in K5 synthesis is provided by studies on KfiABC and KpsEDUC. Cytoplasmic-membrane association of KfiA and KfiC requires the expression of the respective other (Hodson et al., 2000) as well as the expression of KpsC_k5 (Rigg et al., 1998). In addition, membrane association of KfiC relies on the expression of KfiB (Hodson et al., 2000). The role of KfiB remains unclear (Petit et al., 1995) and to date its function in K5 expression remains confined to a structural role. A basic local alignment search tool (BLAST) analysis (Altschul et al., 1990) of KfiB for homologues revealed some sequence similarity with putative sugar transferases of Campylobacter, but KfiB does not carry any conserved regions which would indicate enzyme activity.
Figure 1.6 Current model for *Escherichia coli* K5 (group 2) capsular polysaccharide biosynthesis and export. UDP-Glc dehydrogenase KfiD converts UDP-Glc into UDP-GlcA. This is followed by the alternate addition of GlcA and GlcNAc from UDP-GlcA and UDP-GlcNAc to the non-reducing terminus of the growing polysaccharide chain by glycosyltransferases KfiC and KfiA, respectively. The conversion of Ru5P into A5P is catalysed by A5P isomerise KpsF and provides the CMP-Kdo synthetase KpsU with an intermediate to produce the Kdo precursor CMP-Kdo. The Kdo unit is attached to the reducing terminus of the completely synthesised polysaccharide; this step might involve KpsS and KpsC. Kdo-attachment or a different, unknown process initiates the translocation of the polymer across the membranes. KpsMTED are shown as dimers. KpsM and KpsT form an ABC transporter complex, translocating the substrate across the cytoplasmic membrane using energy from ATP hydrolysis. KpsE and KpsD are involved in exporting the polysaccharide across the periplasm and outer membrane, where the polymer attaches to the bacterial cell surface.

Roles of the proteins involved in K5 biosynthesis and export

KfiD: UDP-Glc dehydrogenase
KfiC: β-UDP-GlcA glycosyltransferase
KfiA: α-UDP-GlcNAc glycosyltransferase
KfiB: function unclear, but required for membrane stability of KfiC

KpsF: putative arabinose 5-phosphate isomerase
KpsU: CMP-Kdo synthetase
KpsC: possible role in Kdo-attachment
KpsS: possible role in Kdo-attachment

KpsT: nucleotide binding domain of ABC transporter KpsMT
KpsM: membrane-spanning domain of ABC transporter KpsMT
KpsE: putative polysaccharide co-polymerase
KpsD: polysaccharide outer membrane exporter

Please see next page for figure.
Figure 1.6 Current model for *Escherichia coli* K5 (group 2) capsular polysaccharide biosynthesis and export.
As in K1 synthesis, region 2 proteins are not sufficient for K5 synthesis but require the additional expression of region 1 proteins (Bronner et al., 1993b). KpsF\textsubscript{K5} is a 35 kDa cytoplasmic protein and 99% identical to KpsF\textsubscript{K1} (Simpson et al., 1996). As in K1, KpsF\textsubscript{K5} is not essential for K5 capsule expression. As mentioned above (section 1.5.1.1), in contrast to K1 polysaccharides, a Kdo residue has been identified attached to the reducing end of K5 and NeuNAc from \textit{N. meningitidis} (Finke et al., 1991). As its \textit{N. meningitidis} homologue, KpsF\textsubscript{K5} can act as isomerise, catalysing the conversion of ribulose 5'-phosphate (Ru5P) to arabinose 5'-phosphate (A5P), a Kdo-precursor (Simpson et al., 1996; Meredith and Woodard, 2006). Phosphatidyl-Kdo might act as acceptor for the initiation of K5 synthesis. KpsU\textsubscript{K5} is also thought to be involved in Kdo synthesis. The 27 kDa cytoplasmic protein acts as synthetase by synthesising CMP-Kdo from arabinose 5-phosphate (Pazzani et al., 1993; Rosenow et al., 1995b).

The role of KpsS\textsubscript{K5} and KpsC\textsubscript{K5} might be different to their homologues in K1 and \textit{N. meningitidis}. The observation that, as in K1 strains, KpsC\textsubscript{K5} is required for the membrane stability of the glycosyltransferase(s) suggests a similar role of KpsC\textsubscript{K5}. KpsC and KpsS are homologues of LipA and LipB of \textit{N. meningitidis} serogroup B, which are thought to be involved in linking synthesis to export by transferring the polymer from the synthesis machinery to the export complex (Tzeng et al., 2005). An equivalent role for KpsS\textsubscript{K5} and KpsC\textsubscript{K5} appears plausible. The two proteins might be required to hold the synthesis complex together in the correct manner and then subsequently translocate the polymer towards KpsMT. However, in contrast to the \textit{N. meningitidis} homologues, the evidence for an involvement of KpsS\textsubscript{K5} and KpsC\textsubscript{K5} in Kdo-attachment is compelling. KpsS\textsubscript{K5} mutants accumulate the completely synthesised polymer in the cytoplasm (Bronner et al., 1993a). Furthermore, KpsS\textsubscript{K5} as well as KpsC\textsubscript{K5} mutants produced the K5 wild-type structure but lacked the attached phosphor-lipid and Kdo.

### 1.5.2 Coupling of K5 synthesis and export

As briefly indicated above, there is evidence that K5 synthesis and export are coupled. Immunofluorescence microscopy revealed that both K5 polysaccharide biosynthesis and export are located at the poles of the bacterial cell (McNulty et al., 2006; Hudson, 2009). The translational coupling of region 2 and 3 of the capsular gene cluster
supports the hypothesis that synthesis and export are directly linked. As discussed earlier, region 2 proteins KfiA and KifC lose their membrane association in the absence of region 1 proteins KpsEDUC(K5 (Rigg et al., 1998). This suggests that there is a direct interaction between the proteins encoded by the two different regions. Furthermore, KpsS(K5, which might be involved in Kdo-attachment to the polysaccharide before export, is required for both polysaccharide synthesis and export (Bronner et al., 1993a). Indeed, chemical cross-linking studies demonstrated that his-tagged KpsS(K5 can cross-link with the export proteins KpsM, KpsT, KpsE, and KpsD (McNulty et al., 2006). This further suggests that polysaccharide synthesis and export are coupled and implies a linker role for KpsS(K5. Recent evidence for the involvement of a protein not encoded by the capsular gene cluster in biosynthesis and export strengthens the hypothesis of a closely linked biosynthesis-export process. Strains carrying a mutated version of the predicted outer membrane protein RhsA exhibited a change in KpsE and KpsD location (both being more dispersed around the cell than in the wild-type) and at the same time exhibited reduced polysaccharide synthesis due to decreased UDP-GlcA glycosyltransferase activity. This suggests that RhsA might mediate the coupling of the biosynthesis-export complex (McNulty et al., 2006). In contrast to the situation in the T1SS where export is initiated by recognition of the substrate's secretion signal by the ABC transporter, no known secretion signal exists in group 2 polysaccharides. This makes the need for a tight coupling of synthesis and export plausible.

1.5.3 Group 2 export – an ABC transporter-dependent process

In contrast to the Wzy-dependent system, transport of group 2 polysaccharides across the membranes occurs after completion of substrate synthesis. Unlike in the case of the Wzy-dependent system and many other substrate exporters, no crystal structures are yet available of any components of the group 2 polysaccharide exporter system. However, a range of studies using molecular approaches have been used to develop a model for group 2 export. The KpsMTED proteins responsible for export of group 2 polysaccharides in E. coli show high sequence similarities with the export proteins of N. meningitidis group B, H. influenzae serotype b, and S. Typhi (Table 1.2; cf. sections 1.7, 1.8 and 1.9).

It is well established that group 2 capsular polysaccharides are transported via an ABC transporter-dependent mechanism. ABC membrane transporters play a significant role
### Table 1.2 *Escherichia coli* group 2 capsular polysaccharide transport proteins and homologues

This table details the percentages of identical and similar residues (left and right side of each table, respectively) of *E. coli* KS transport proteins (KpsTMED) and homologues of *N. meningitidis* group B (CtrDCBA), *H. influenzae* type b (BexABCD), and *S. Typhi* (VexCBDA). Data of sequence identities and similarities were obtained through analysis of sequence alignments produced by ClustalW (Thompson *et al.*, 1994) using the SIAS Server (Reche, 2008).

<table>
<thead>
<tr>
<th>KpsT homologues</th>
<th>KpsM homologues</th>
<th>KpsT</th>
<th>CtrD</th>
<th>BexA</th>
<th>VexC</th>
<th>KpsM</th>
<th>CtrC</th>
<th>BexB</th>
<th>VexB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>63</td>
<td>63</td>
<td>37</td>
<td></td>
<td>39</td>
<td>38</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>CtrD</td>
<td></td>
<td>48</td>
<td>86</td>
<td>41</td>
<td></td>
<td>CtrC</td>
<td>23</td>
<td>80</td>
<td>34</td>
</tr>
<tr>
<td>BexA</td>
<td></td>
<td>48</td>
<td>82</td>
<td>40</td>
<td></td>
<td>BexB</td>
<td>25</td>
<td>70</td>
<td>33</td>
</tr>
<tr>
<td>VexC</td>
<td></td>
<td>22</td>
<td>27</td>
<td>27</td>
<td></td>
<td>VexB</td>
<td>20</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>KpsT</td>
<td>CtrD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BexA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VexC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsT homologues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KpsM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CtrD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BexA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VexC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CtrB</td>
<td></td>
<td>70</td>
<td>70</td>
<td>61</td>
<td></td>
<td>CtrA</td>
<td>21</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>BexC</td>
<td></td>
<td>62</td>
<td>86</td>
<td>61</td>
<td></td>
<td>BexD</td>
<td>10</td>
<td>66</td>
<td>56</td>
</tr>
<tr>
<td>VexD</td>
<td></td>
<td>64</td>
<td>80</td>
<td>63</td>
<td></td>
<td>VexA</td>
<td>10</td>
<td>48</td>
<td>46</td>
</tr>
<tr>
<td>KpsE</td>
<td>CtrB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BexC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VexD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsE homologues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KpsD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CtrA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BexD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>VexA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>KpsE</th>
<th>CtrB</th>
<th>BexC</th>
<th>VexD</th>
<th>KpsD</th>
<th>CtrA</th>
<th>BexD</th>
<th>VexA</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>52</td>
<td>55</td>
<td></td>
<td>70</td>
<td>86</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>80</td>
<td>63</td>
<td></td>
<td>10</td>
<td>66</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>52</td>
<td>55</td>
<td></td>
<td>10</td>
<td>48</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>
in the transport of a variety of substrates, either being involved in importing (ABC importers) or exporting (ABC effluxers) substrates across the bacterial membrane.

Structurally, all ABC transporters share common characteristics (Locher et al., 2002; Schmitt et al., 2003; Kos and Ford, 2009). The four protein subunits comprise two hydrophobic membrane-spanning domains (signalling domains) which form a channel-like structure across the membrane, and two hydrophilic nucleotide-binding domains (NBD) which are located in the cytoplasm. The NBD resemble the ABC region, which carries highly conserved ATP-binding motifs forming an ATP-binding pocket. The ATP-binding pocket supplies the energy required for active transport of substrates by ATP binding and hydrolysis. The conserved motifs in the ABC comprise of the Walker A (or P loop) and Walker B motif, the LSGGQ motif (signature motif or linker peptide), the Q loop and the switch region (H motif). While Walker A and signature motif are engaged in ATP binding and hydrolysis, Walker B motif and Q loop bind the magnesium and water required for the conversion of ATP to ADP. The Q motif has been shown to interact with the membrane-spanning domains of the transporter. The histidine residue in the H motif forms a hydrogen bond to the γ-phosphate of ATP lost during ATP hydrolysis. In addition to conserved regions, the ABCs also contain a structurally diverse region which interacts with the membrane-spanning domains. As this region is specific to each ABC it is likely to be responsible for the substrate selectivity of the ABC transporter (Schmitt et al., 2003).

1.5.3.1 The ABC transporter KpsMT

In group 2 export the ABC transporter is made of the two proteins KpsT and KpsM, which form the NBD and the membrane-spanning domain, respectively. As in bacterial T1SS, the ABC transporter translocates the substrate across the cytoplasmic membrane. The two region 3 encoded genes kpsT and kpsM are organised as one transcriptional unit (cf. Figure 1.1 A), with a two-nucleotide-overlap between the kpsM stop codon and the kpsT start codon (Pavelka et al., 1991). There is evidence that the two genes are translationally coupled (Smith et al., 1990; Pavelka et al., 1991). Taken together, this suggests that the two gene products interact with each other and are likely to be expressed in equal amounts.
KpsT is a 25 kDa peripheral cytoplasmic membrane protein (Pavelka et al., 1994) and expresses the ABC (Smith et al., 1990). KpsT shares 63% similarity with BexA of *H. influenzae*, considering identical and conserved residues (Table 1.2). Evidence that KpsT is essential for polysaccharide export comes from studies on KpsT-deficient mutants which accumulated polysaccharide in the cytoplasm (Pavelka et al., 1994). Indeed, site-directed mutagenesis identified ATP-binding as crucial for KpsT function and polysaccharide export (Pavelka et al., 1991; Pavelka et al., 1994). A structural model predicts that KpsT is a typical ABC protein (Pavelka et al., 1994). The ABC region is thought to consist of a 5-stranded β-sheet with 5 intervening α-helices, carrying the Walker A and Walker B motifs. This region is interrupted by a variable helical domain after the second β-strand. The helical domain comprises three α-helices and carries the signature motif towards the end of the domain. This helical domain is thought to interact with the membrane-bound domains of the transporter, KpsM. The signature motif is predicted to ‘transmit’ the conformational change ‘signal’ upon ATP binding from the ABC to the helical domain. This then leads to a conformational change in the helical region, and, in turn, results in conformational change of KpsM. This model of action is supported by evidence from mutagenesis studies. Glutamic acid at position 150 of KpsT is located within the third intervening α-helix of the ABC and is essential for ATP hydrolysis (Bliss et al., 1996). Proteolysis experiments support the hypothesis that KpsT and KpsM undergo conformational change upon ATP binding and KpsT interaction, respectively (Nsahlai and Silver, 2003).

KpsM is a 29.5 kDa integral cytoplasmic membrane protein with cytoplasmic N- and C-termini and 6 transmembrane domains (Smith et al., 1990; Pigeon and Silver, 1994). KpsM shares 38% similarity with BexB of *H. influenzae* (Table 1.2). The observation that KpsM can act as a dimer in the cytoplasmic membrane supports the evidence that KpsM is part of the ABC transporter complex (Pigeon and Silver, 1994). Mutagenesis experiments demonstrated that functional polysaccharide export requires the expression of the first cytoplasmic loop, a linker region within the third periplasmic loop, as well as the N-terminus of KpsM (Pigeon and Silver, 1994). A mutation of native glycine 93 within the first cytoplasmic loop to glutamic acid resulted in a reduction of polysaccharide surface expression and transport. This indicates that the first cytoplasmic loop is involved in interaction with either the substrate or one of the other translocon members, possibly KpsT.
1.5.3.2 The inner membrane protein KpsE

The 43 kDa protein KpsE (Pazzani et al., 1993) belongs to the PCP-3 family (Tocilj et al., 2008) and is required for group 2 polysaccharide export (Bronner et al., 1993a). Please refer to Figure 1.3 for a phylogenetic tree of KpsE and other PCP members. KpsE shares 70% sequence similarity with both CtrB and BexC of N. meningitidis serogroup B and H. influenzae type b, respectively (Table 1.2). Protease analysis revealed that KpsE is integrated into the cytoplasmic membrane and expresses a large periplasmic domain (Rosenow et al., 1995a). Its N-terminus is located in the cytoplasm. However, in contrast to PCP-1 and PCP-2 proteins, KpsE has only a single transmembrane domain, and its topology thus closely resembles that of membrane fusion proteins such as HlyD of the T1SS. Protease digestion analysis revealed that the C-terminus is located in the periplasm and is predicted to form an α-helix which is closely associated with, but not integrated into, the cytoplasmic membrane (Arrecubieta et al., 2001; Phoenix et al., 2001). Interestingly, a recent paper studying the topology of CtrB showed that even the closely related CtrB expresses two transmembrane domains, following the classical topology of PCP proteins (Larue et al., 2011). KpsE can act as a dimer independently of the polysaccharide substrate and other Kps proteins (Arrecubieta et al., 2001). It remains to be elucidated whether this is the oligomeric state KpsE adopts during polysaccharide export.

1.5.3.3 The outer membrane protein KpsD

KpsD is a 60 kDa outer membrane protein (Pazzani et al., 1993) required for group 2 export (Bronner et al., 1993a). It carries an N-terminal signal sequence typical for proteins secreted via a Sec-dependent secretion system (Wunder et al., 1994). Please refer to Figure 1.3 for a phylogenetic tree of KpsD and other polysaccharide outer membrane exporters. KpsD does not have any significant known homologues, not even in functionally close polysaccharide export systems such as N. meningitidis or H. influenzae (cf. Table 1.2 (Pazzani et al., 1993)). However, KpsD shares limited homology with Wza (cf. Figure 5.2 (Drummelsmith and Whitfield, 1999)). KpsD-Wza homology and KpsD structure will be further discussed in Chapter 5. Analysis of the periplasm and cytoplasmic and outer membrane fractions showed that KpsD is present in the all of them but primarily in the periplasm (Arrecubieta et al., 2001). This initially implied that KpsD is responsible for the periplasmic transport of polysaccharides. However, a
further study using fluorescence activated cell sorting demonstrated that KpsD is also exposed to the cell surface, indicating that KpsD is a proper membrane protein integrated into the outer membrane (McNulty et al., 2006). This suggests that KpsD is involved in outer membrane transport. KpsD forms dimers in the outer membrane (McNulty et al., 2006), but has so far not been shown to form any higher oligomeric complexes as common in Wza or TolC.

1.5.3.4 Evidence for a group 2 translocon complex

In the past decade evidence for a group 2 translocon complex with interacting export proteins has accumulated. It has been observed that KpsM cannot be detected by SDS-PAGE when co-expressed with KpsT alone (Smith et al., 1990; Pavelka et al., 1991). Recent evidence demonstrated that stable membrane association of KpsM requires the co-expression of KpsE (Hudson, 2009). That this interaction is likely to be a direct physical one is suggested by the observation that KpsM can co-purify KpsE (Hudson, 2009). This interaction appears independent of other capsule proteins, including KpsT, and the polysaccharide substrate. This implies that KpsE and KpsM might form a pre-complex independent of ATP binding and hydrolysis. A parallel observation has been made in the T1SS, where the KpsMT-equivalent HlyB and the KpsE-equivalent HlyD can cross-link in the absence of ATP binding and hydrolysis (Thanabalu et al., 1998).

Proteinase K sensitivity assays demonstrated that KpsD and KpsE were more sensitive to enzyme degradation in the absence of the other protein, respectively (Arrecubieta et al., 2001). This suggests that the two proteins might stabilise each other. It was further shown that KpsD requires expression of KpsE for outer membrane localisation and oligomerisation (Hudson, 2009). Evidence for direct interaction comes from co-purification studies. KpsE can co-purify KpsD (Hudson, 2009). As this interaction is dependent on the co-expression of KpsM, the formation of a KpsE-KpsD pre-complex is less likely. More plausible is the interaction, or the induction of conformational change, of the two proteins upon binding of KpsD with a KpsTM-KpsE complex after substrate interaction. Recent analysis indeed demonstrated that KpsMT-KpsE-KpsD can be co-eluted together in a purification experiment (Hudson, 2009) and thus provided the first evidence of a continuous translocon complex which spans the cytoplasm membrane, periplasm and outer membrane. Interestingly,
KpsMTED interaction is independent of the substrate, which contrasts with the T1SS where the transport machinery assembles upon substrate binding.

The current model of the group 2 polysaccharide translocon complex proposes that ATP-binding by KpsT leads to conformational change in KpsT and the insertion of the protein into the cytoplasmic membrane (Pigeon and Silver, 1997; Silver et al., 2001; Nsahlai and Silver, 2003). The subsequent interaction of KpsT with KpsM might result in a conformational change in the KpsM-KpsE pre-complex. KpsD might then change its conformation upon interaction with KpsE, possibly switching from a closed to an open state, allowing the transport of the substrate across the outer membrane (similarly as suggested for the Wzc-Wza complex).

1.6 Group 3 capsules of *Escherichia coli*

Group 3 capsular polysaccharides (formerly either group I/II or III) are represented by capsules K10 and K54. The K10 is made of 1,3-α-rhamnose-1,3-β-4(2-carboxyacetamido)-4,6-dideoxyglucose, with an O-acetylated position 2 of rhamnose (Sieberth et al., 1993). The capsular gene clusters of K10 and K54 are, like group 2 gene clusters, located near *serA* on the bacterial chromosome (Pearce and Roberts, 1995). As group 2 capsular gene clusters, the gene cluster of group 3 is structured into three regions, with one central, serotype-specific region flanked by two conserved regions (Figure 1.1 B (Clarke et al., 1999)). However, the genetic organisation of the group 3 gene cluster differs to that of group 2. The genes *kpsDMTE* and *kpsCS* are encoded by regions 1 and 3, respectively (Russo et al., 1998; Clarke et al., 1999). Region 2 carries gene for capsule synthesis. This organisation is identical to the capsular gene cluster organisation in *N. meningitidis* and *H. influenzae*, where all export genes are organised as one transcriptional unit and the putative polysaccharide modification genes are grouped independently. DNA hybridisation analysis of the group 3 gene cluster was unable to identify nucleotide homologues among capsule genes from other *E. coli* groups (Russo et al., 1998). Group 3 KpsMTED and KpsCS share 46%, 40%, 30% and 34%, and 34% and 87% amino acid identity with group 2 equivalent proteins, respectively (Clarke et al., 1999). In addition, KpsMTED are significantly homologous to the export proteins of *N. meningitidis*, *H. influenzae* and *S. Typhi* (Russo et al., 1998). These observations suggest that group 3 capsular polysaccharides share a similar translocation mechanism with *E.*
coli group 2, N. meningitidis, H. influenzae and S. Typhi, and are thus most likely exported via an ABC transporter-dependent pathway.

1.7 Neisseria meningitidis serogroup B capsule

The Gram-negative bacterium N. meningitidis expresses a number of immunologically and chemically distinct polysaccharide capsules. Serogroups A, B, C, W-135 and Y are the medically most important capsule types. All serogroups share a common organisation of the capsule gene cluster and show homologies across all capsule genes except for those in the polysaccharide-specific synthesis region (Frosch et al., 1991; Swartley et al., 1997). Meningococcal capsule expression is most widely studied in group B strains. The group B meningococci capsule is an α-2,8-linked NeuNAc polymer, identical to the K1 (group 2) capsule of E. coli. The identical structure of the two capsules suggests similar synthesis and export mechanisms. As in group 2 capsular gene clusters in E. coli, three regions required for capsular synthesis and export have been identified in group B meningococci (Frosch et al., 1989; Frosch et al., 1991; Edwards et al., 1994; Ganguli et al., 1994). These regions have been termed regions A, B, and C and encode the sia, lip, and ctr operons respectively. Transcription of the sia and ctr operons occurs constitutively via promoters located in the 134-bp region between the two operons, with transcription of sia being about four times higher than transcription of ctr (Tzeng et al., 2001). Complementation experiments showed that transcriptional regulation of the sia operon is specific to N. meningitidis, as the regulatory elements involved could not replace the ones present in E. coli K1 (Tzeng et al., 2001). The regulatory protein CrgA is proposed to negatively regulate transcription of the sia genes during bacterial adhesion with host epithelial cells (Deghmane et al., 2002) in order to avoid internalisation of the encapsulated bacterium. Capsule-switching can occur through reversible insertion of the transposable element IS1301 into the siaA gene, reversibly inactivating capsule synthesis (Hammerschmidt et al., 1996).

The genes siaABCD (synABCD) encode the enzymes involved in synthesis of the polysaccharide in the cytoplasm. Please refer to Table 1.1 for details on E. coli K1 enzyme homologues. SiaA has been identified as 40 kDa protein, catalysing the production of ManNAc (Petersen et al., 2000; Murkin et al., 2004). SiaA shares homology with the E. coli K1 GlcNAc 2-epimerase NeuC, which uses UDP to form
ManNAc (Vann et al., 2004). However, the substrate identity for SiaA remains unclear as Petersen et al. (2000) and Murkin et al. (2004) report two different substrates for this conversion, UDP-GlcNAc or GlcNAc-6-phosphate. SiaB is a 25 kDa CMP-NeuNAc synthetase which activates NeuNAc to CMP-NeuNAc (Edwards and Frosch, 1992; Ganguli et al., 1994) and is located in the cytoplasm (Masson and Holbein, 1983). SiaB is a homologue of the \( E. \ coli \) K1 CMP-NeuNAc synthetase NeuA and can functionally complement NeuA (Edwards and Frosch, 1992). SiaC, which has been localised in the cytoplasm (Masson and Holbein, 1983) and has homology to NeuB of \( E. \ coli \) K1, is thought to condensate phosphoenolpyruvate with ManNAc to form NeuNAc (Edwards et al., 1994). SiaD has been identified as an \( \alpha \)-2,8 polysialyltransferase and is homologous to the \( E. \ coli \) K1 sialyltransferase NeuS (Frosch et al., 1991). SiaD is cytoplasmic membrane-associated (Masson and Holbein, 1983). Mutagenesis studies revealed amino acid motifs conserved in sialyltransferases which are essential for CMP-NeuAc binding and enzyme catalysis in SiaD (Freiberger et al., 2007). Similar to polysaccharide synthesis in \( E. \ coli \) group 2, Kdo appears to play a role in meningococcal capsule expression. Mutagenesis analysis using all five \( N. \ meningitidis \) serogroups demonstrated the requirement of the gene \( kpsF \) for capsule expression (Tzeng et al., 2002). KpsF catalyses the interconversion of Ru5P and A5P and has been demonstrated to be essential for Kdo synthesis (Tzeng et al., 2002). This suggests that Kdo is important for capsule expression in meningococci. Kdo might act as acceptor for the initiation of polysaccharide polymerisation, a role it is believed to play in \( E. \ coli \) group 2 polysaccharide synthesis.

\( N. \ meningitidis \) capsular polysaccharide transport is encoded by the \( ctr \) and, possibly, by the \( lip \) operon. Sequence homology of the \( ctr \)ABCD genes with the capsular transport genes \( bexDCA \) of \( H. \ influenzae \) and \( kpsDMET \) of group 2 \( E. \ coli \) suggest related functions of these gene products (cf. Table 1.2). To date, meningococcal capsule export has been poorly studied and the export model mainly relies on functional predictions based on studies of group 2 \( E. \ coli \). CtrD is a 25 kDa protein with an ATP-binding region (Frosch et al., 1991); it shares strong homology with BexA and significant homology with KpsT. CtrD is predicted to form an ABC transporter with CtrC. CtrC shares strong homology with BexB and significant homology with KpsM, the membrane-spanning domain of the ABC transporter KpsMT (Frosch et al., 1991). Assuming a similar mechanism of polysaccharide translocation in \( N. \ meningitidis \) and \( E. \)
coli, the CtrDC transporter is responsible for transport of the polymer across the cytoplasmic membrane. CtrB and CtrA are predicted to be tranlocating the polysaccharide chain across the periplasm and outer membrane. Please refer to Figure 1.3 for a phylogenetic tree of CtrB and CtrA and related proteins. CtrA shares significant homology with BexD, has a lipoprotein signal sequence and is located in the outer membrane (Frosch et al., 1991). CtrB is a cytoplasmic membrane protein, sharing significant homology with BexC and KpsE (Frosch et al., 1991). A recent study on CtrB was the first to analyse a ctr gene product in more detail (Larue et al., 2011). The secondary structure of CtrB and KpsE were previously predicted to be almost identical due to close sequence homology. Cysteine-scanning mutagenesis revealed that CtrB is a typical PCP protein with cytoplasmic N- and C-termini and a large central domain in the periplasm (Larue et al., 2011). As this structure differs from KpsE which has only one transmembrane region, it appears that even within the PCP-3 family membrane topologies diverge. Electron microscope analysis suggested a ring-shaped three-dimensional structure of CtrB (Larue et al., 2011).

The lip operon is the least investigated region of the N. meningitidis capsular gene cluster. All serogroups express a 1,2-diacylglycerol phospholipid anchor at the reducing end of the polysaccharide chain (Gotschlich et al., 1981), and it has been previously thought that the lipAB genes, which share significant homology with kpsC and kpsS of E. coli group 2, are involved in lipidation. However, a study revealed that lipAB and separate lipA and lipB mutants produce wild-type lipid-linked polysaccharides but accumulated the completely synthesised polymers in the cytoplasm (Tzeng et al., 2005). This suggests a role in polysaccharide transport for LipAB. It is possible that LipA and LipB play a role in linking capsule synthesis to export. A further candidate for this role has recently been identified. ORF NMB0065, designated ctrG, is located just downstream of siaD and is predicted to encode a 37 kDa cytoplasmic protein, which has homology with NeuE of E. coli K1 (Hobb et al., 2010). CtrG mutants expressed no or little capsular polysaccharide at the cell surface. (Hobb et al., 2010). However, CtrG was not involved in adding repeat units, chain length regulation or lipidation. It was thus proposed that CtrG might be involved in transferring the synthesised polymer from the site of synthesis to the export machinery.
1.8 *Haemophilus influenzae* serotype b capsule

Research on the synthesis and export of *H. influenzae* capsules is limited and efforts to elucidate the biological basis of capsulation have been focused on developing effective capsule-based vaccine strategies against *H. influenzae* infection. The Gram-negative *H. influenzae* expresses six chemically distinct capsules, termed serotypes a to f. *H. influenzae* type b (Hib) is associated with the most invasive form of *H. influenzae* infection and has been demonstrated to be an independent virulence factor by providing protection from immune response during infection (Zwahlen *et al.*, 1989). The Hib capsule is made of repeats of 5-D-ribitol-β-D-ribose-3-phosphate (Crisel *et al.*, 1975). Loss of Hib capsule expression is a frequent event and *H. influenzae* commonly switches between encapsulated and acapsulated forms (Hoiseth *et al.*, 1985).

The genetic organisation of the Hib capsular gene cluster (*cap*) is closely related to the one in *N. meningitidis* group B and *E. coli* group 2. DNA-DNA hybridisation analyses have demonstrated that *cap* is a 18 kb sequence with three distinct regions (Kroll *et al.*, 1989). As with the *E. coli* group 2 and *N. meningitidis* group B clusters, region 1 and 3 are conserved among all capsule types of *H. influenzae* and region 2 is type-specific (Kroll *et al.*, 1989). Functional capsule expression requires the partial duplication of *cap*. Encapsulated *H. influenzae* have two identical copies of region 2 as well as region 3, and additionally one complete and one truncated copy of region 3 (Kroll *et al.*, 1989; Satola *et al.*, 2003). The partially deleted region 3 contains a truncated version of *bexA* – the gene encoding the putative ATP-binding protein of the capsule ABC transporter. Early research suggested that loss in capsule expression is due to a *rec*-dependent recombination event which results in the deletion of the duplicated copies within *cap* (Hoiseth *et al.*, 1985). Indeed, more recent sequence analysis of encapsulated and acapsulated variants confirmed this (Satola *et al.*, 2003). There is some variation in *cap* expression within Hib as well as within the other types, placing the various *cap* loci into two phylogenetic categories, termed division I and II. The overall organisation of *cap* in division I and II is identical and genetic divergence is very limited (Kroll and Moxon, 1990). However, while *cap* in division I (expressed more frequently) is flanked by direct repeats of insertion element *IS1016* on either side, division II *cap* is not associated with an insertion element (Kroll *et al.*, 1991; Satola *et al.*, 2003). This suggests that division I *cap* is a compound transposon and provides evidence for common evolutionary origin.
of many capsule gene clusters of Gram-negative bacteria resulting from lateral gene transfer.

Hib capsule synthesis has not been studied very extensively and the roles of proteins involved in polysaccharide polymerisation are deduced mainly from sequence and homology analysis. The four capsular synthesis genes *bcs1234* are located on region 2 of *cap* (Vaneldere *et al.*, 1995; Satola *et al.*, 2003). Bsc1 and Bsc3 are essential for capsule expression and encode a cytoplasmic CDP-ribitol pyrophosphotylase and a protein of unknown function, respectively (Vaneldere *et al.*, 1995). Bsc2 and Bsc4 are not required for expression of Hib and encode a 42 kDa cytoplasmic protein believed to be a ribitol dehydrogenase and a further protein of unknown function, respectively (Vaneldere *et al.*, 1995).

Capsule export is encoded by genes *bexABCD* in region 1 of *cap* and export genes share homology with the export genes of *N. meningitidis* group B and group 2 *E. coli* (Table 1.2). See Figure 1.3 for the relationship of BexC and BexD to related proteins illustrated by a phylogenetic tree. Based on this homology a similar transport mechanism via an ABC transporter has been proposed, where polysaccharide export involves the putative ATP-binding protein BexA interacting with the potential inner membrane component of the ABC transporter, BexB. Translocation of the polymer across the periplasm and outer membrane is predicted to occur via the putative PCP-3 BexC and the outer membrane protein BexD. Similarly to *lipAB* in *N. meningitidis*, region 3 has been previously believed to be involved in lipid modification of the Hib capsule. However, analysis of the region 2 genes *hcsA* and *hcsB* did not find evidence for an involvement of HcsA and HcsB in lipidation (Sukupolvi-Petty *et al.*, 2006). Polysaccharides from mutants lacking *hcsA, hcsB* or both genes appeared to be still attached to a phospholipid, suggesting the absence of HcsA and/or HcsB did not affect lipidation (Sukupolvi-Petty *et al.*, 2006). Close homologies of HcsA with LipA and KpsC, and HcsB with LipB and KpsS suggest that the two proteins are responsible for linking capsule synthesis and transport. Deletion mutagenesis of *hcsAB* demonstrated that capsule cell-surface expression was reduced after *hcsA* deletion and completely abolished after deletion of either *hcaB* or *bcsAB* (Sukupolvi-Petty *et al.*, 2006). Localisation and quantification of polysaccharide in the cytoplasm, periplasm and cell surface revealed that *hcsA* and *hcsB* mutants expressed levels of the capsule comparable
to wild-type in the cytoplasm and periplasm. However, deletion of \textit{hcsA} and \textit{hcsB} resulted in a reduced amount of cell surface-associated and excreted polysaccharide (Sukupolvi-Petty \textit{et al}., 2006). This suggests that HcsA and HcsB might have a different role to LipAB and KpsCS, which might be more closely linked to capsule transport. Further studies are required to elucidate the exact role of HcsAB in Hib capsule expression.

1.9 \textit{Salmonella enterica} serovar Typhi capsule

Research on the polysaccharide capsule Vi of \textit{Salmonella enterica} serovar Typhi has primarily been focused on the role of Vi as virulence factor and as basis for a vaccine against typhoid fever. The Vi capsule is made of polymers of \textit{N}-acetyl galactosamine uronic acid (Heyns \textit{et al}., 1959). The capsule gene cluster \textit{viaB} is part on the \textit{Salmonella} pathogenicity island 7 and contains, like \textit{E. coli} group 2, \textit{N. meningitidis} group B and \textit{H. influenzae} type b, three distinct regions. Vi biosynthesis is encoded by region 2 genes \textit{tviBCDE} (Hashimoto \textit{et al}., 1991; Kolyva \textit{et al}., 1992; Hashimoto \textit{et al}., 1993). Functional analysis showed that TviB is a UDP-GlcNAc 6-dehydrogenase, TviC acts as UDP-GlcNAc 4-epimerase, and TviE is a glycosyltransferase (Zhang \textit{et al}., 2006). The role of TviD remains unclear.

Regulation of Vi capsule expression is under the control of at least two different two-component regulatory systems, \textit{rcsB-rcsC} and \textit{ompE-envZ} (Pickard \textit{et al}., 1994; Arricau \textit{et al}., 1998), the negative regulator RpoS (Santander \textit{et al}., 2007). Regulation also involves \textit{tviA}, the region 1-encoded regulatory protein (Virlogeux \textit{et al}., 1995). RcsB acts as a positive regulator and can interact with TviA at the \textit{tviA} promoter to regulate capsule expression (Virlogeux \textit{et al}., 1995). Vi regulators have been shown to respond to changes in osmolarity levels, with optimal Vi expression being induced at low and medium osmolarities (Tran \textit{et al}., 2010). Osmolarity-dependent regulation is crucial for pathogenic S. Typhi, as the pathogen has to adjust its capsule expression depending on the host environment.

Region 3 of \textit{viaB} contains the genes \textit{vexABCDE}, which are responsible for transport of the Vi capsular polymers. VexABC have been shown to be essential for Vi transport (Virlogeux \textit{et al}., 1995). Sequence homology of VexABCD to BexDBAC (Hashimoto \textit{et al}., 1995).
suggests that the export model of Vi resembles that of *H. influenzae* group B (cf. Table 1.2). VexC (putative ATP-binding region) is predicted to form an ABC transporter with VexB (inner membrane component). VexD is likely to be a PCP-3 and VexA an outer membrane exporter. The function of VexE remains unsolved. Figure 1.3 illustrates the relationship of VexD and VexA to related proteins in a phylogenetic tree.

1.10 *Streptococcus pyogenes* capsule

The Gram-positive *S. pyogenes* (group A streptococcus) is a common skin and throat coloniser and the causative agent of infections like bacterial pharyngitis, scarlet fever, impetigo and streptococcal toxic shock syndrome (Cunningham, 2000). *S. pyogenes* produces a hyaluronic acid (HA) capsule made of 1,4-β-GlcNAc-1,3-β-GlcA di- and trisaccharides. In contrast to most other capsule-producing bacteria, the gene cluster responsible for HA expression appears to consist of three genes only. The three genes *hasABC* are transcribed as a single unit (Crater and Vanderijn, 1995). Transposon mutagenesis demonstrated that *hasAB* are sufficient for HA synthesis in an acapsular host strain (DeAngelis *et al.*, 1993a). HasA is a 42 kDa HA synthase, adding UDP-GlcA and UDP-GlcNAc in alternation to the reducing end of the growing polymer (Markovitz *et al.*, 1959; DeAngelis *et al.*, 1993b; Tlapak-Simmons *et al.*, 2005). HasB has been identified as a 47 kDa UDP-Glc dehydrogenase (Dougherty and Vanderijn, 1993). Mutagenesis studies revealed that HasA activity is highly specific (DeAngelis *et al.*, 1993b). Truncation mutants of HasB were still able to produce HA in the presence of wild-type HasA when an alternative dehydrogenase was simultaneously expressed. However, the loss of HasA activity due to HasA truncation could not be complemented and mutants were unable to express HA. HasC is a 36 kDa UDP-Glc pyrophosphorylase (Crater *et al.*, 1995) and is not essential for HA production (Ashbaugh *et al.*, 1998a). Due to the size of the *has* gene cluster it cannot be excluded that additional proteins, which are shared with other metabolic pathways, might contribute to HA synthesis and export.

The *has* cluster is also present in acapsulated variants of *S. pyogenes*. Primer extension studied showed that the *has* genes are under the control of a transcriptional regulators, which represses transcription in acapsulated and stationary encapsulated variants.
To date, the two-component regulatory system $csrR-csrS$ has been identified as playing an important part in transcriptional regulation of HA (Bernish and van de Rijn, 1999; Jadoun et al., 2002).

HA synthesis and export are spatially coupled processes (Hubbard et al., 2012). To date, two contrasting models for HA transport have been proposed. Model 1 involves an ABC transporter, while model 2 suggests a HasA-dependent process. Insertional mutagenesis in the region upstream of the $has$ operon identified seven open reading frames (Ouskova et al., 2004). The two gene products encoded directly adjacent to $hasA$ both carry Walker A and B motifs and one of them shares 54% similarity with KpsT, the $E. coli$ group 2 ATP-binding component of the capsular ABC transporter. A third gene encodes a protein with four transmembrane domains and 46% similarity to KpsM, the cytoplasmic membrane component of the $E. coli$ group 2 ABC transporter. There is evidence that HA is transported via this putative ABC transporter, as deletion of these three genes (and a putative $N$-acetyl glucosylase involved in peptidoglycan synthesis) significantly reduced HA production and HA release (Ouskova et al., 2004). However, supporters of the HasA-dependent export model suggest that the identified ABC transporter complex might have alternative substrates such as nucleotides and that the observed decrease in HA expression might be a response to a modified concentration of UDP-sugars required for wild-type HA synthesis (Medina et al., 2012).

Topology analysis of HasA using alkaline phosphatase and $\beta$-galactosidase fusion proteins identified four membrane spanning domains and two membrane-associated regions (Heldermon et al., 2001). The N- and C-termini of HasA are located in the cytoplasm. HasA has three cytoplasmic loops and two smaller loops in the periplasm. Further characterisation of HasA provided evidence that the active HasA enzyme complex consists of monomeric HasA interacting with around 16 cardiolipin molecules (Tlapak-Simmons et al., 1998). Evidence for an $in vivo$ involvement of cardiolipins is supported by the observation that high enzyme activity of purified HasA required supplementary cardiolipins (Tlapak-Simmons et al., 1998). Taken together, these studies provide evidence that HasA might be a pore-forming protein. A recent study further investigated the hypothesis of HasA-dependent export (Medina et al., 2012). They tested the ability of purified HasA to efflux the fluorescent dye Cascade Blue across liposome membranes. It was demonstrated that release of Cascade Blue from
liposomes was HasA-dependent and that active, rather than denatured HasA, was required for successful substrate efflux. Indeed, another recent study showed that the lipid-reconstituted version of the HasA protein of *S. equisimilis* was able to translocate the *S. equisimilis* HA polymer across the lipid membrane (Hubbard *et al.*, 2012). Further analyses are required to elucidate the *in vivo* mechanisms of HA export.

1.11 Type 1 secretion system: parallels to *Escherichia coli* group 2 capsular polysaccharide export

The mechanism of exporting polysaccharides across the two membranes in Gram-negative bacteria shares similar obstacles to the transport of proteins and other substrates. Saier *et al.* (2008) summarise the diversity of protein insertion and secretion systems in bacteria and eukaryotic organelles in their comprehensive review. As the bacterial T1SS shares many characteristics with the ABC transporter-dependent translocation pathway of *E. coli* group 2 polysaccharides, the characteristics of its protein secretion machinery will be briefly discussed.

1.11.1 α-Haemolysin export system

The expression of T1SSs is common in Gram-negative bacteria and is involved in transporting substrates such as adhesion factors, lipases and toxins. As *E. coli* group 2 polysaccharide export systems, the T1SS consists of an ABC transporter component responsible for substrate transport across the cytoplasmic membrane as well as a cytoplasmic membrane fusion protein and an outer membrane protein. One of the most studied T1SS is the α-haemolysin (HlyA) transporter of *E. coli*. HlyA is a 110 kDa exotoxin which can lyse erythrocytes and plays a role as virulence factor in extra-intestinal infections (Welch *et al.*, 1981). The tripartite exporter of HlyA consists of the HlyB-HlyD-TolC complex.

HlyB is the ABC transporter component of the HlyA secretion system. Topology analysis of HlyB using fusion proteins identified a short N-terminal cytoplasmic domain, 8 transmembrane regions and a large C-terminal cytoplasmic domain (Wang *et al.*, 1991). The periplasmic regions between the α-helical transmembrane domains are rather short, while the cytoplasmic regions are larger (Wang *et al.*, 1991; Gentschev and
Goebel, 1992). Mutagenesis studies demonstrated that the first 25 residues at the N-terminus are not required for HlyB function but site-directed mutations in the conserved periplasmic loop 5 were toxic (Blight et al., 1994). The C-terminal domain of HlyB contains the NBD and its crystal structure has been resolved (Schmitt et al., 2003). This region is made of two domains which together form the L-shape characteristic for NBDs of ABC transporters, and are connected by the conserved Q and Pro loops. The catalytic domain 1 contains the Walker A and B motifs and other residues including the H and D loops conserved among NBDs and forms two \( \beta \)-sheets and 7 \( \alpha \)-helices. It was suggested that residues within the Walker A motif act as switch to regulate ATPase activity. The signalling domain 2 contains the signature motif and is made of 5 \( \alpha \)-helices. This domain is also thought to contain a motif responsible for the interaction between the NBD and the transmembrane region of the protein.

In contrast to group 2 exporters, information about the initial recognition stages at the beginning of the transport process exist. There is evidence that recognition of HlyA by HlyB initiates export (Gentschev and Goebel, 1992). Indeed, the final 62 amino acids at the C-terminus of HlyA were identified as secretion signal which is recognised by HlyB (Zhang et al., 1993; Jarchau et al., 1994; Sheps et al., 1995). It has been proposed that the cytoplasmic loops of HlyB are involved in this interaction (Zhang et al., 1993). In addition to the secretion signal sequence, the efficiency of substrate secretion is dependent on the folding rate of the substrate (Bakkes et al., 2010). It has been demonstrated that HlyB expression in the absence of other components of the transporter is sufficient for HlyA transport across the cytoplasmic membrane (Thomas et al., 1992). Nevertheless, expressing HlyB to detectable levels has been proven difficult in the absence of the membrane fusion component HlyD (Blight et al., 1995; Pimenta et al., 1999). This phenomenon is also seen in KpsM of \( \text{E. coli} \) group 2, which requires KpsE for expression. In the HlyA system, this might indicate that HlyB requires HlyD for its membrane stability. Blight et al. (1995) proposed an alternative explanation in which HlyD is required to post-translationally regulate HlyB via a HlyB-HlyD translational regulatory complex.

HlyD is a typical member of the family of membrane fusion proteins with a cytoplasmic N-terminus, a single cytoplasmic transmembrane domain and a large periplasmic domain (Wang et al., 1991; Schulein et al., 1992). Deletion of the N-
terminus abolished HlyA export but did not disrupt the native trimeric state of HlyD (Balakrishnan et al., 2001). The C-terminal region of HlyD is also essential for export. The final 30 residues at the C-terminus are predicted to form a helix-loop-helix motif (Schulein et al., 1994). The expression of this motif is required for functional HlyA export and is therefore likely to be involved in interaction of HlyD with other translocon members. As leucine 475, glutamic acid 477 and arginine 479 within the second helix are essential for export (Schulein et al., 1994), these residues might play a particularly significant role in protein-protein interaction. The region between leucine 127 and leucine 170 are predicted to play a role in the interaction with TolC, the outer membrane component of the transporter (Schlor et al., 1997). This region is located at the beginning of the periplasmic region and is homologous to region valine 231-leucine 274 in TolC (Schulein et al., 1994). This homologous region is essential for export (Schulein et al., 1994).

TolC is a versatile outer membrane transporter involved in export of various substrates. TolC was first identified as the outer membrane component of the HlyA transporter in 1990 (Wandersman and Delepelaire, 1990) and its crystal structure was resolved in 2000 (Koronakis et al., 2000). TolC has a three-folded symmetry and the three monomers of TolC form a β-barrel in the outer membrane and a large α-helical channel in the periplasm. The TolC monomer is made of a set of structural repeats containing α-helix (H) 1, H2, β-sheet (S) 1, S2, H3, H4 and H5, H6, S4, S5, H7, H8. A short mixed α-β structure, termed equatorial domain, is packed against the middle region of the α-helical domain (Koronakis et al., 2000). The HlyD-homologous region forms part of H6, all of S4 and part of the extracellular loop between S4 and S5, and thus is located close to the periplasmic face of the outer membrane and within the outer membrane and extracellularly. A simulation study identified two flexible domains of TolC which would potentially allow a conformational change to occur, regulating the opening and closing of the channel (Vaccaro et al., 2008). The first region is the extracellular loops connecting the β-barrel-forming sheets. This region can change between a collapsed (closed) and non-collapsed (open) state, and regulation of conformational change of this region would allow an opening and closing of the top region of the β-barrel. The second region is the β-barrel region itself. TolC can change its barrel from a cylindrical to a triangular prism conformation. This conformational change might be induced by interaction with one or both translocon members and might function in guiding the
substrate towards the channel entrance. The crystal structure of two open states of TolC demonstrated that the helical coiled coils in the periplasmic entrance of the channel change from a closed, resting state to an open conformation by expanding in an iris-like rotation (Pei et al., 2011). This expansion opens the TolC tunnel. It remains to be investigated whether conformational changes in TolC during export are limited to just this expansion, but it is likely that other changes, similar to the ones described by the simulation study, occur simultaneously.

Many parallels can be identified when comparing the protein interactions within the group 2 polysaccharide and HlyA exporters. As noted earlier, in the absence of HlyD expression levels of HlyB are greatly reduced (Pimenta et al., 1999), suggesting an interaction between the two proteins in some way. Indeed, cross-linking experiments suggest that HlyB and HlyD interact in the absence of other proteins (Thanabalu et al., 1998). This interaction is also independent of the substrate (Thanabalu et al., 1998; Balakrishnan et al., 2001), indicating that the two proteins might form a pre-complex, prior to interaction with TolC. Similar evidence is available for KpsM and KpsE in group 2 polysaccharide export. In the absence of KpsE, KpsM is only expressed at low levels and co-purification experiments suggest that a direct interaction between the two proteins independent of the K5 substrate (Hudson, 2009). The wild-type state of HlyD is altered in the absence of HlyB as well as TolC. In the absence of either of the proteins an increased amount of lower molecular HlyD breakdown products were detected (Pimenta et al., 1999). This suggests that HlyD requires both HlyB and TolC for stability. In group 2 polysaccharide export, the absence of KpsMT leads to the increased formation of breakdown products of KpsE (Hudson, 2009). It has not yet been investigated whether lack of KpsD also leads to KpsE breakdown products. Co-purification studies provided further evidence of protein-protein interactions in the HlyA export system. HlyB as well as HlyD can both co-purify TolC, and co-purification of HlyD-TolC is dependent on the presence of HlyB (Thanabalu et al., 1998). Similar observations were made in the group 2 export system. KpsM was shown to co-purify KpsE and KpsD (Hudson, 2009), while co-purification of KpsD by KpsE was dependent on the presence of KpsM (Hudson, 2009). The first 59 residues at the N-terminal cytoplasmic region of HlyD are not crucial for HlyD-HlyB interaction as mutants lacking this region were still able to interact with HlyB (Balakrishnan et al., 2001). A similar situation is likely to be found in group 2 exporters. Deletion of the first
N-terminal amino acids of KpsE did not abolish polysaccharide export (Hudson, 2009), which suggests that this region is not crucial for KpsE-KpsM interaction. Future work on both T1SS and group 2 polysaccharide exporters will help to identify further parallels within, and differences between the two systems.
1.12 Aims and objectives of this study

Despite extensive research into the export process of *E. coli* group 2 capsular polysaccharides, many questions about the protein components of the translocon complex remain unsolved. The overall aim of this project is to investigate the functional and structural characteristics of KpsM, KpsE, and KpsD, as well as their protein-protein interactions. To address this overall aim, the following objectives were set:

1. Determine residues and domains in KpsE which are crucial for protein function and interaction with KpsM and KpsD

2. Investigate strategies to purify KpsE-KpsM as higher order complex

3. Investigate strategies to determine the membrane topology of KpsD

It is anticipated that an in-depth understanding of the group 2 transport machinery will aid the development of therapeutic tools against infections caused by encapsulated *E. coli* strains in the future.
Chapter 2

Materials and Methods
2.1 Chemicals and reagents

Unless otherwise stated, all chemicals and reagents were manufactured by Sigma-Aldrich and all media were obtained from Oxoid. Where necessary, reagents were autoclaved (121°C, 15 psi, for 20 minutes) or filtered through a 0.22 µm PES filter disc (Milipore). Oligonucleotide primers were custom-made by Sigma-Aldrich. A list of primers used during this study is given in Table 2.3.

2.2 Bacterial strains, growth conditions and plasmids

The bacterial strains used during this study are listed in Table 2.1. Bacterial starter cultures were grown overnight by shaking at 200 revolutions per minute (rpm) in Luria-Bertani (LB) broth (1% w/v tryptone, 0.5% w/v yeast extract, 1% w/v sodium chloride (NaCl)), supplemented with the appropriate antibiotic as required (100 µg ml⁻¹ ampicillin; 50 µg ml⁻¹ streptomycin, 30 µg ml⁻¹ kanamycin). Except for encapsulated strains, which were grown at 25°C, all strains were routinely grown at 37°C. When strains were grown for experimental assays, a starter culture was grown overnight in LB at 200 rpm shaking (at 25°C for encapsulated strains, and at 37°C for acapsular strains). The overnight culture was then diluted 1:100 into fresh LB and grown at 37°C, 200 rpm shaking, until the desired growth phase was reached (determined by optical density measurements at a wavelength of 600 nm (OD₆₀₀ nm)). Induction of protein expression in bacterial cultures is described in sections 2.7.1-2.7.3. For growth on solid medium, strains were grown on LB agar (LB broth supplemented with 1.5% w/v agar), supplemented with the appropriate antibiotic, and incubated at the appropriate temperature. For short term storage, strains grown on LB agar were stored at 4°C. For long term storage, strains were stored as glycerol stocks at -80°C (900 µl of overnight culture of the required strain was mixed with 600 µl of sterile 50% v/v glycerol). A list of plasmids used during this study is given in Table 2.2.

2.3 Preparation of competent cells

2.3.1 Chemically competent cells

An overnight culture of the desired strain was diluted 1:100 into 100 ml fresh LB and grown at 37°C, 200 rpm shaking, until a cell density of OD₆₀₀: 0.4 was reached. Cells
### Table 2.1 List of strains used during the study.

<table>
<thead>
<tr>
<th><strong>Escherichia coli</strong> strain</th>
<th>Relevant genotype</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td><em>deoR endA1 gyrA96 hsdR17 (rK mK</em>) recA1 relA1 supE44 thi-1 Δ(lacZYA-argFV169) Φ80lacZΔM15, F*</td>
<td>(Hanahan, 1983)</td>
</tr>
<tr>
<td>BL21-Δ1&lt;sup&gt;TM&lt;/sup&gt;</td>
<td>F-<em>ompT hsdSB (rB-mB-) gal dcm araB::T7RNAPtetA</em></td>
<td>Invitrogen; derived from BL21 (DE3) (Studier and Moffatt, 1986; Grodberg and Dunn, 1988)</td>
</tr>
<tr>
<td>MS101</td>
<td>PA360 (from <em>E. coli</em> K-12) serA&lt;sup&gt;+&lt;/sup&gt; K5&lt;sup&gt;+&lt;/sup&gt;</td>
<td>(Stevens <em>et al.</em>, 1994b)</td>
</tr>
<tr>
<td>MSΔE</td>
<td>MS101 ΔkpsE</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>MSΔD</td>
<td>MS101 ΔkpsD</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>MSΔEΔMT</td>
<td>MS101 ΔkpsE ΔkpsM ΔkpsT</td>
<td>(Hudson, 2009)</td>
</tr>
</tbody>
</table>

### Table 2.2 List of plasmids used during the study.

<table>
<thead>
<tr>
<th><strong>Plasmid</strong></th>
<th>Relevant properties</th>
<th>Reference/Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBAD28</td>
<td>Arabinose-inducible cloning vector, 5.8 kb, Cm&lt;sup&gt;r&lt;/sup&gt;, Amp&lt;sup&gt;r&lt;/sup&gt;, p15a replication origin.</td>
<td>(Guzman <em>et al.</em>, 1995)</td>
</tr>
<tr>
<td>pBADE</td>
<td>pBAD28 containing the full length kpsE (encoding amino acids 1-382) cloned into SacI and XmaI restriction sites. Includes engineered ribosome binding site (RBS) prior to kpsE start codon.</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>pBADE Q289A</td>
<td>pBADE with mutation in KpsE amino acid 289 from glutamine to alanine.</td>
<td>This study</td>
</tr>
<tr>
<td>pBADE A287V</td>
<td>pBADE with mutation in KpsE amino acid 287 from alanine to valine.</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>pBADEΔCterm (pBADEΔTM2)</td>
<td>pBADE lacking C-terminal membrane anchor of kpsE (encoding amino acids 357-377) replaced with KpnI restriction site (encoding glycine and threonine).</td>
<td>This study</td>
</tr>
<tr>
<td>pBADhisMT</td>
<td>pBAD28 containing full length kpsM and kpsT and 589 base pairs (bp) downstream of kpsT cloned into SacI and XmaI sites. Includes engineered RBS prior to kpsM start codon plus N-terminal 6x histidine (his)-tag after kpsM start codon.</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>pBADEhisMT</td>
<td>pBADE containing kpsM and kpsT downstream of kpsE cloned into XbaI and HindIII restriction sites. Includes engineered RBS prior to kpsM start codon plus N-terminal 6x his-tag after kpsM start codon.</td>
<td>(Hudson, 2009)</td>
</tr>
<tr>
<td>pBADEΔCterm,hisMT (pBADEΔTM2hisMT)</td>
<td>pBADE&lt;sub&gt;ΔCterm&lt;/sub&gt; containing full length kpsM and kpsT downstream of kpsE&lt;sub&gt;ΔCterm&lt;/sub&gt; cloned into XbaI and HindIII restriction sites. Includes engineered RBS prior to kpsM start codon plus N-terminal 6x his-tag after kpsM start codon.</td>
<td>This study</td>
</tr>
<tr>
<td>pBADEhisTEVMT</td>
<td>pBADE containing full length kpsM and kpsT downstream of kpsE, cloned into XbaI and HindIII restriction sites. Includes engineered RBS prior to kpsM start codon, N-terminal 6x his-tag.</td>
<td>This study</td>
</tr>
</tbody>
</table>
his-tag after \textit{kpsM} start codon and N-terminal Tobacco etch virus (TEV) protease cleavage site downstream of his-tag.

\begin{tabular}{|l|l|l|}
\hline
\textbf{pBADED} & pBAD28 containing full length \textit{kpsE} and \textit{kpsD} cloned into SacI and Xmal restriction sites. Includes engineered RBS prior to \textit{kpsE} start codon. (Hudson, 2009) & \\
\hline
\textbf{pBADED}_{\text{Δcys}} (pBADED C488S) & pBADED containing full length \textit{kpsE}, and full length \textit{kpsD} lacking native cysteine (cysteine to serine mutation at amino acid 488). This study & \\
\hline
\textbf{pBADE}_{\text{CTmyc}} & pBADE containing full length \textit{kpsD} plus c-myc-tag before \textit{kpsD} stop codon, cloned into XbaI and HindIII sites. This study & \\
\hline
\textbf{pBADE}_{\text{hisMT}} & pBADE containing full length \textit{kpsM} and \textit{kpsT} downstream of \textit{kpsED} cloned into XbaI and HindIII restriction sites. Includes engineered RBS prior to \textit{kpsM} start codon plus N-terminal 6x his-tag after \textit{kpsM} start codon. (Hudson, 2009) & \\
\hline
\textbf{pET-24(+) T7 promoter primer} & IPTG-inducible expression vector, 5.2 kb, Kan', f1 replication origin, T7 RNA polymerase gene, C-terminal 6x his-tag. Novagen & \\
\hline
\textbf{pET-24(+) T7 terminator primer} & pET-24(+) containing full length \textit{kpsE} cloned into SacI and Xmal restriction sites, and full length \textit{kpsM} and \textit{kpsT} cloned into XbaI and HindIII sites. Includes engineered RBS prior to \textit{kpsE} and \textit{kpsM} start codons plus N-terminal 6x his-tag after \textit{kpsM} start codon. Contains stop codon downstream of \textit{kpsMT}, so construct does not express C-terminal his-tag. This study & \\
\hline
\end{tabular}

\textbf{Table 2.3 List of primers used during the study.} Restriction enzyme cleavage sites are underlined.

\begin{tabular}{|l|l|l|}
\hline
\textbf{Primer name} & \textbf{Sequence (5'-3')} & \textbf{Explanation} \\
\hline
\textbf{Screening primers} & & \\
\textbf{pBAD28 F} & cttctccatatccgggttc & Forward multiple cloning site (MCS) primer (binds downstream of MCS at bp 1274) of pBAD28. \\
\hline
\textbf{pBAD28 R} & ggctgaaaatcttctct & Reverse MCS primer (binds upstream of MCS at bp 1396) of pBAD28. \\
\hline
\textbf{pET-24(+)} T7 promoter primer & taatagctcactagag & Forward MCS primer (binds downstream of MCS at bp 327) of pET-24(+). \\
\hline
\textbf{pET-24(+)} T7 terminator primer & gctagttatggcaggg & Reverse MCS primer (binds upstream of MCS at bp 69) of pET-24(+). \\
\hline
\textbf{Cloning & mutagenesis primers} & & \\
\textbf{KpsM and KpsT} & & \\
\textbf{hisKpsMT XbaI F} & ggtcctagggagtgaatatcatgccagcactaccatc & Amplification of \textit{hiskpsMT} for generation of pBADE_{ΔCterm}hisMT and pETEhisMT. \\
\hline
\end{tabular}
<table>
<thead>
<tr>
<th><strong>hisTEVKpsMT XbaI F</strong></th>
<th>ggttcatagaggagtatacatgtgcatcaccaatcac</th>
<th>Amplification of <em>hiskpsMT</em> for generation of pBADEhisTEVMT.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KpsT HindIII R</strong></td>
<td>cttgaagtctttctaatataatctgtagcattattataactatctatcac</td>
<td>Amplification of <em>hiskpsMT</em> for generation of pBADEΔCtermhisMT, pBADEhisTEVMT and pETEhisMT.</td>
</tr>
<tr>
<td><strong>KpsE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KpsE Q289A F</strong></td>
<td>gcacaatatcactgcccagcgggtgacaagcttaacc</td>
<td>Site-directed mutagenesis (SDM) of glutamine to alanine at position 289 in KpsE for generation of pBADE Q289A.</td>
</tr>
<tr>
<td><strong>KpsE Q289A R</strong></td>
<td>gtttagctgtcaccggctggcgcagtgattttgc</td>
<td>SDM of glutamine to alanine at position 289 in KpsE for generation of pBADE Q289A.</td>
</tr>
<tr>
<td><strong>KpsE P288A F</strong></td>
<td>gcacaatatcactgcccagcgggtgacaagcttaacc</td>
<td>SDM of proline to alanine at position 288 in KpsE for generation of pBADE P288A.</td>
</tr>
<tr>
<td><strong>KpsE P288A R</strong></td>
<td>gtttagctgtcaccggctggcgcagtgattttgc</td>
<td>SDM of proline to alanine at position 288 in KpsE for generation of pBADE P288A.</td>
</tr>
<tr>
<td><strong>KpsE SacI F</strong></td>
<td>ggggagctcggagtatacatgtgataaagtgagaagtc</td>
<td>Amplification of <em>kpsE</em> and C-terminal deletion in <em>kpsE</em> for generation pETEhisMT and pBADEΔCtermV.2, respectively. Also used by T. Hudson to generate pBADEΔTM2v.1 (creation of N-terminal side) (Hudson, 2009).</td>
</tr>
<tr>
<td><strong>KpsE R</strong></td>
<td>gggcccggtgtagctctggtgatcttaataacagcgc</td>
<td>Amplification of <em>kpsE</em> fragment for generation of pETEhisMT. Contains XmaI site.</td>
</tr>
<tr>
<td><strong>KpsE TM2 XmaI R</strong></td>
<td>gcgcgcttcagtaaccgagctgcgactaaccgcgc</td>
<td>Amplification of C-terminal deletion in KpsE for generation of pBADEΔCtermV.2.</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Primer Sequence</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>KpsD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsD C488S F</td>
<td>gaggattacatcagaaaaagcgggtgcctgacg</td>
<td>SDM of cysteine to serine at position 488 in KpsD for generation of pBADEΔcys.</td>
</tr>
<tr>
<td>KpsD C488S R</td>
<td>cgtcaggccaccgctttctcgtatgtacctc</td>
<td>SDM of cysteine to serine at position 488 in KpsD for generation of pBADEΔcys.</td>
</tr>
<tr>
<td>KpsD Xbal v2 F</td>
<td>gggccccttagaatgaaatattatatcaattttactgattgcgccctgtacgc</td>
<td>Amplification of C-terminally myc-tagged KpsD for generation of pABDEΔCTmyc.</td>
</tr>
<tr>
<td>KpsD-CT myc HindIII R</td>
<td>gcggcaagctttacagatcttctcagaaatctttgtctttctcagaaatctttgtcttcagaaatcactttgcac</td>
<td>Amplification of C-terminally myc-tagged KpsD for generation of pABDEΔCTmyc.</td>
</tr>
<tr>
<td><strong>Sequencing primers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsM and KpsT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsM seq 180bp R</td>
<td>gtctggcatcgtgcggtgcataatg</td>
<td></td>
</tr>
<tr>
<td>KpsM seq 189bp F</td>
<td>cccggtgttttactaatgtgctctg</td>
<td></td>
</tr>
<tr>
<td>KpsM seq 513bp R</td>
<td>catttcaggggaaggttacctacc</td>
<td></td>
</tr>
<tr>
<td>KpsT seq 228bp F</td>
<td>gccctgccgggttgattcc</td>
<td></td>
</tr>
<tr>
<td>KpsT seq 300bp R</td>
<td>ctctttgcttgcgtataacgc</td>
<td></td>
</tr>
<tr>
<td><strong>KpsE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsE seq 223bp R</td>
<td>gacccaaatttccaggctgcgcctg</td>
<td></td>
</tr>
<tr>
<td>KpsE seq 235bp F</td>
<td>gcctttaaccccaagtccggcacag</td>
<td></td>
</tr>
<tr>
<td>KpsE seq 435bp R</td>
<td>ggaccgtatcaaacgctctg</td>
<td></td>
</tr>
<tr>
<td>KpsE seq 501bp F</td>
<td>gtttaacccaaacgcgtgctgaaag</td>
<td></td>
</tr>
<tr>
<td>KpsE seq 793bp F</td>
<td>gtgaatgcgcgtatgcgtatccag</td>
<td></td>
</tr>
<tr>
<td>KpsE seq 1140bp R</td>
<td>gtgatcttcaataacgcgcagcaac</td>
<td></td>
</tr>
<tr>
<td><strong>KpsD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpsD Start F</td>
<td>caacctttacctgatgccgcctg</td>
<td></td>
</tr>
<tr>
<td>KpsD Start R</td>
<td>tcaacccgttaacgcacacatc</td>
<td></td>
</tr>
<tr>
<td>KpsD Mid F</td>
<td>gcgcgactacattacag</td>
<td></td>
</tr>
<tr>
<td>KpsD Mid R</td>
<td>tttgcaacataacgcctg</td>
<td></td>
</tr>
<tr>
<td>KpsD End F</td>
<td>atcgagaaatgcggtgcgc</td>
<td></td>
</tr>
<tr>
<td>KpsD End R</td>
<td>aagacagaaatcacttttgcaacc</td>
<td></td>
</tr>
</tbody>
</table>
were kept on ice for 10 minutes and then pelleted by centrifugation at 3000 g for 15 minutes at 4°C. The cell pellet was resuspended in 25 ml ice-cold transformation buffer 1 (30 mM potassium acetate, 10 mM rubidium chloride (RbCl), 20 mM calcium chloride (CaCl$_2$), 50 mM manganese chloride, 15% v/v glycerol, pH 5.8) and left on ice for 10 minutes. The cells were centrifuged again at 3000 g for 10 minutes at 4°C and the pellet resuspended in 1 ml ice-cold transformation buffer 2 (10 mM 3-(N-morpholino)propane sulfonic acid (MOPS), 150 mM CaCl$_2$, 10 mM RbCl$_2$, 15% v/v glycerol pH 6.5). Cells were stored on ice for 2 to 3 hours, and then separated into 100 µl aliquots, which were then either used immediately or were snap-frozen using liquid nitrogen and stored at -80°C.

2.3.2 Electrocompetent cells
An overnight culture of the desired strain was diluted 1:100 into 5 ml fresh LB and grown at 37°C, 200 rpm shaking, until a cell density of OD$_{600}$: 0.2-0.5 was reached. Cells were kept on ice for 5 minutes. A cell pellet was obtained by centrifugation at 3000 g for 10 minutes at 4°C. The pellet was resuspended in 1 ml of ice-cold sterile distilled water and centrifuged at 6000 g for 3 minutes. This wash step was repeated a further four times. The cell pellet was resuspended in a final volume of 500 µl ice-cold sterile dH$_2$O, or resuspended in sterile 20% v/v glycerol for long-term storage. Cells were separated into 50 µl aliquots, which were then either used immediately or were snap-frozen using liquid nitrogen and stored at -80°C.

2.4 Transformation of competent cells
2.4.1 Transformation of chemically competent cells
Frozen aliquots of competent cells were thawed on ice. 100 µl of ice-cold competent cells were mixed with either 5 µl of ligation mixture, 15 µl of QuikChange reaction, or 100 ng of plasmid DNA. The mixture was left on ice for 5 minutes; it was then heat-shocked at 42°C for 60 seconds and returned onto ice for 5 minutes. Cells were allowed to recover by adding 900 µl of pre-warmed LB broth and incubating them at the appropriate temperature and for the the appropriate time (either at 37°C for 45 minutes to 1 hour, or at 25°C for 1 to 1.5 hours). The recovered cells were then centrifuged at 6000 g for 3 minutes and the pellet was resuspended in 200 µl of its supernatant. The
cells were plated onto the appropriate selective solid medium, left to dry, and were incubated overnight for cells with a growth optimum of 37°C or for 36 to 48 hours for cells with a growth optimum of 25°C at the appropriate temperature.

2.4.2 Transformation of electrocompetent cells
Frozen aliquots of competent cells were thawed on ice. 50 µl of ice-cold competent cells were mixed with 5 µl of plasmid DNA and left on ice for 5 minutes. The mixture was then transferred into the bottom of a 0.2 cm electroporation cuvette (Bio-Rad) and the cells were electroporated at 2500 V, 200 ohm (Ω), and 25 µF. Cells were immediately returned onto ice and mixed with 1 ml of fresh LB broth. Cells were allowed to recover by incubating them at the appropriate temperature and for the appropriate time (either at 37°C for 45 minutes to 1 hour, or at 25°C for 1 to 1.5 hours). The cells were plated onto the appropriate selective solid medium, left to dry, and were incubated overnight for cells with a growth optimum of 37°C or for 36 to 48 hours for cells with a growth optimum of 25°C at the appropriate temperature.

2.5 DNA analysis and manipulation

2.5.1 Polymerase chain reaction
DNA amplifications by polymerase chain reaction (PCR) were prepared in a total volume of 50 µl per reaction and performed in a thermal cycler (PxE 0.5 [Thermo Electron Corporation] or Tc-412 [Techne]). The annealing temperatures were adjusted depending on the primers used, setting the annealing temperature 5°C below the melting temperature of the primer.

2.5.2 PCR-based manipulation of DNA fragments
Amplifications of DNA fragments were carried out by PCR. The 50 µl reaction volume included 1 µM of both forward and reverse primer, at least 100 ng of plasmid DNA template, 1x PCR buffer (Roche Diagnostics), and 1 mM deoxynucleotide triphosphates (dNTP) mix (Roche Diagnostics). In addition, 5 units of Taq polymerase (Roche Diagnostics) and 0.25 units of PfuUltra polymerase (Stratagene) were added to the reaction mixture to facilitate cloning and reduce polymerase error rate, respectively.
The reaction mixtures were denatured for 5 minutes at 95°C, followed by 35 amplification cycles (1 minute at 95°C, 1 minute at annealing temperature, 2 minutes at 72°C) and a final extension period of 10 minutes at 72°C. The PCR reactions were purified using a MinElute PCR Purification kit (Qiagen), according to the manufacturer’s instructions.

2.5.3 Site-directed mutagenesis

Site-directed mutagenesis (SDM) was performed according to the QuikChange® Site-Directed Mutagenesis method (Stratagene). Forward and reverse primers coding for the point mutations were designed and used for PCR-based DNA manipulation. A 50 µl of reaction volume included 0.2 µM of each primer (forward and reverse), 20-50 ng of plasmid DNA template, 0.2 mM dNTP mix (Roche Diagnostics), 1x PfuUltra buffer (Stratagene) and 2.5 units of PfuUltra polymerase (Stratagene). In addition, a reaction without PfuUltra polymerase was set up to serve as positive control. The following thermal cycles were used: an initial denaturation step for 30 seconds at 95°C, followed by 18 amplification cycles (1 minute at 95°C, 1 minute at 55°C, 16 minute at 68°C) and a final extension period of 5 minutes at 68°C. Following PCR, 1.5 µl of DpnI restriction enzyme (10 U/µl) per reaction was added. In addition, a reaction identical to the above but without DpnI was set up to serve as negative control. The reactions were incubated for at least 1 hour at 37°C to digest non-mutated supercoiled double-stranded DNA. The reaction mixture was then ligated into a suitable vector (section 2.5.7), before being transformed into chemically competent cells (section 2.4.1).

2.5.4 Colony PCR

Colony PCR was carried out to screen for mutations and to verify successful cloning. The 50 µl of reaction volume included 1 µM of each primer (forward and reverse), 1x PCR buffer (Roche Diagnostics), 1 mM dNTP mixture (Roche Diagnostics) and 1 unit of Taq polymerase (Roche Diagnostics). A single colony was added as DNA template by transferring a single colony into the PCR mix using the end of a sterile pipette tip. The reaction mixtures were denatured for 5 minutes at 95°C, followed by 35 amplification cycles (1 minute at 95°C, 1 minute at annealing temperature, 2 minutes at 72°C) and a final extension period of 10 minutes at 72°C.
2.5.5 Plasmid DNA extraction and purification

Plasmids used during the course of this study are listed in Table 2.2. 10 ml of overnight culture was pelleted at 3000 g for 10 minutes. Plasmid DNA was extracted using a QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer’s instructions. Plasmid DNA was further purified by ethanol precipitation. Briefly, 10% v/v 3 M sodium acetate pH 4.5 and 10x volume of 100% v/v ethanol were added to the plasmid in 10 mM Tris pH 8.5, and the mixture incubated at -20°C for 20 minutes. Recovery of the precipitated DNA was achieved by centrifugation at 16,000 g for 20 minutes. The supernatant was decanted and the pellet washed twice in an equal volume of 70% v/v ethanol. The plasmid pellet was then air-dried at 42°C. The pellet was either stored at -20°C or resuspended in 10 mM Tris, pH 8.5. Plasmid DNA concentration and purity were determined using a Nanodrop ND-1000 ultraviolet (UV) spectrophotometer (Labtech), reading at 260 nm.

2.5.6 Restriction endonuclease cleavage

Restriction endonuclease cleavage of DNA was performed in a total volume of 20 µl per reaction, with incubation of 1 hour or more at the optimum enzyme temperature. Reactions were performed according to the manufacturer’s recommendations (Roche Diagnostics). If specified, 100 µg ml⁻¹ of bovine serum albumin was added to the reaction.

2.5.7 Ligation

Ligation reactions were performed in a total volume of 20 µl per reaction. 1 unit of T4 DNA ligase (Roche Diagnostics), 2 µl of ligase buffer (Roche Diagnostics), and a 1:3 ratio of cleaved plasmid:fragment were used. Ligation reactions were incubated overnight at 4°C. Ligation reactions were transformed into chemically competent cells (section 2.4.1).

2.5.8 DNA sequencing

DNA sequencing reactions were carried out in a total volume of 20 µl using the BigDye V1.1 Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Corporation). Reactions included 3.2 pM of sequencing primer, 2 µl BigDye v.1.1 terminator, 3 µl
BigDye Sequencing Buffer, and at least 300 ng of template plasmid DNA. The following thermal cycles were used: an initial denaturation step for 2 minutes at 96°C, followed by 30 amplification cycles (45 seconds at 96°C, 45 seconds at 50°C, 4 minutes at 60°C). The sequencing reactions were then purified by mixing them with 2.5 µl of 125 mM ethylenediaminetetraacetic acid (EDTA), 2.5 µl of 3 M sodium acetate, and 75 µl of 100% v/v ethanol, followed by brief vortexing and incubation at room temperature for 20 minutes. Sequencing products were collected by centrifugation at 16,000 g for 30 minutes. The supernatant was removed and the DNA pellet was washed with 150 µl of 70% v/v ethanol, followed by another centrifugation step for 15 minutes. The ethanol was removed and the pellet air dried at 42°C. DNA samples were analysed in an ABI PRISM 3100 Genetic Analyzer DNA sequencer (Applied Biosystems) by the University of Manchester’s Faculty of Life Sciences DNA Sequencing Core Facility.

2.5.9 Agarose gel electrophoresis

DNA samples were separated by agarose gel electrophoresis according to size. 1% w/v agarose gels were made up in TAE buffer (40 mM Tris-acetate pH 7.7, 1 mM EDTA). DNA samples were mixed with 6x DNA loading buffer (0.25% w/v xylene cyanol FFC, 0.25% w/v bromophenol blue, 15% v/v ficoll type 400) prior to loading onto the agarose gel. Hyperladder markers (Bioline) were run alongside the samples to indicate molecular weights. Samples were electrophoresed at 90 V in TAE buffer containing 0.5 µg ml⁻¹ ethidium bromide (Sigma). Gels were visualised using an UV transluminator (UVIpro Silver, UVItec).

2.5.10 Agarose gel extraction of DNA fragments

DNA fragments were separated by agarose gel electrophoresis (section 2.5.9) and excised from the gel under UV light with a sterile scalpel. DNA extraction was carried out using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer’s instructions.
2.6 K5-specific bacteriophage sensitivity assay

Bacterial strains were screened for the presence of K5 polysaccharide capsule using the K5-specific bacteriophage (Roberts lab). This bacteriophage is a lytic coliphage with a double-stranded DNA genome, an icosahedral head, a base plate and tail spikes (Gupta et al., 1982). The K5 phage degrades the K5 capsule using its tail spike protein, the lyase KflA, in order to expose phage receptors on the outer membrane of *E. coli* (Gupta et al., 1982; Thompson et al., 2010). For the bacteriophage sensitivity assay, bacterial cultures were grown overnight in their appropriate media supplemented with the appropriate antibiotics. Overnight cultures were diluted 1:100 into fresh sterile LB broth and grown at 37°C (200 rpm shaking) until a cell density of OD₆₀₀: 0.5 was reached. Cells were collected by centrifugation at 6000 g for 3 minutes and the pellet was resuspended in an equal volume of ice-cold sterile 10 mM magnesium sulphate (MgSO₄). K5 bacteriophage dilution series were made up in sterile phage suspension buffer (100 mM NaCl, 8 mM MgSO₄, 50 mM Tris, 0.01 % v/v gelatine pH 7.5). For each dilution, 100 µl of phage dilution was mixed with 100 µl of cell suspension and incubated at room temperature for 30 minutes. 3 ml of molten soft-top LB agar (LB broth supplemented with 0.75% w/v agar) was added and the mixture was poured onto pre-warmed LB agar plates supplemented with the appropriate antibiotics and incubated overnight at 37°C.

2.7 Protein expression, membrane extraction, protein purification and analysis

2.7.1 Standard induction of protein expression

For expression of proteins from pBAD28 vectors, host strains DH5α or MS101 and MS101 derivatives were used. Protein expression from pET-24(+) vectors was carried out in host strain BL21-AI. Unless cells were grown to high cell density (section 2.7.2) or according to the autoinduction method (section 2.7.3), the following protocol was used to induce protein expression. Bacterial starter cultures of the required strains were grown, unless otherwise stated, overnight in LB supplemented with the appropriate antibiotics at the appropriate temperature. Overnight cultures were diluted 1:100 into fresh sterile LB broth and grown at 37°C, 200 rpm shaking, until a cell density of OD₆₀₀: 0.4-0.5, or until a different density as specified in alternative protocols, was
reached. Unless otherwise noted, protein expression was then induced by adding either 0.01% w/v arabinose (pBAD28 vectors in MS101 and MS101 derivatives), 0.1% w/v arabinose (pBAD28 vectors in DH5α), or 1 mM w/v isopropyl-β-D-thio-galactoside (IPTG) plus 0.2% w/v arabinose (pET-24(+) vectors in BL21-AI) to the medium and incubating the cultures for 3 hours or other specified duration at 37°C, 200 rpm shaking. Cells were collected by centrifugation at 6000 g for 20 minutes. The cell pellet was stored at -20°C.

2.7.2 High cell density growth and induction of bacterial cells

This protocol was modified from Sivashanmugam et al. (2009). Overnight cultures (section 2.7.1) were diluted 1:100 into fresh sterile LB broth and grown at 37°C, 200 rpm shaking, until a cell density of OD₆₀₀: 0.4-0.5 was reached. The cells were then gently spun down and resuspended in an equal volume of fresh LB and grown again for the duration which the cells previously required to reach a cell density of OD₆₀₀: 0.4-0.5. The culture was then induced with 0.1% w/v arabinose by shaking at 200 rpm at 37 °C for 3 hours. The cells were pelleted and stored at -20°C.

2.7.3 Autoinduction of protein expression

This protocol was kindly provided by Dr. Edward McKenzie, Senior Experimental Officer in the Protein Expression Core Facility of the Faculty of Life Sciences, University of Manchester. Bacterial starter cultures were grown for approximately 8 hours at 37°C by shaking at 200 rpm in LB broth, supplemented with the appropriate antibiotic as required. The starter culture was diluted 1:100 in fresh 500 ml autoinduction base medium (1% w/v tryptone, 0.5% w/v yeast extract) plus supplements (final concentrations of 25 mM monopotassium phosphate (KH₂PO₄), 25 mM disodium hydrogen phosphate (Na₂HPO₄), 50 mM ammonium chloride (NH₄Cl), 5 mM sodium sulphate (Na₂SO₄), 2 mM magnesium sulphate (MgSO₄), 2 mM calcium chloride (CaCl₂), 0.5% v/w glycerol, 2.5 mM glucose, 0.5% v/w lactose) and 0.05% w/v final concentration of arabinose (BL21-AI strain only) The culture was grown overnight at 37°C by shaking at 200 rpm. The cells were pelleted and stored at -20°C.
2.7.4 Preparation of periplasmic and cytoplasmic fractions

Periplasmic and cytoplasmic fractions were obtained by converting cells into spheroplasts. The method used was adapted from Koshland and Botstein (1980) which had originally been taken from Withold et al. (1976). 1 ml of cells (section 2.7.1) was centrifuged for 15 minutes at 18,000 g. All subsequent steps were carried out at 4°C. Cell pellets were resuspended in 75 µl of spheroplasting buffer (100 mM Tris, 500 mM EDTA, 0.5 M sucrose pH 8.0) and 7.5 µl of 2 mg/ml lysozyme (Sigma). This was followed by the addition of 75 µl dH₂O. The sample was then incubated on ice for 25 minutes. To stabilise the spheroplasts, 3 µl of 1M MgCl₂ were added and the sample was then centrifuged for 15 minutes at 18,000 g. The supernatant was saved as the periplasmic fraction. Any residual fluid was drained off and the pellet was resuspended in 100 µl dH₂O and saved as the cytoplasmic fraction. Periplasmic and cytoplasmic fractions were stored at 4°C for short term storage or at -20°C for long term storage.

2.7.5 Cell lysis by French Pressing and preparation of total membrane fractions

Thawed whole cell pellets (sections 2.7.1-2.7.3) were resuspended in an appropriate volume of ice-cold French press buffer (20 mM Tris, 150 mM NaCl, 20% v/v glycerol pH 8.0) supplemented with protease inhibitor cocktail (Sigma). Cell suspensions were passed twice through a pre-cooled French pressure cell press (Thermo Electron Corporation) at 9000 psi. The cell debris was removed by centrifugation at 10,000 g for 10 minutes. The supernatant was transferred to 3.2 ml polycarbonate ultracentrifuge tubes (Beckman) and centrifuged in an Optima Ultracentrifuge (Beckman Coulter) at 260,000 g for 1 hour at 4°C. The supernatant was saved as soluble fraction and stored on ice. Total membrane pellets were stored on ice or at -20°C.

2.7.6 Preparation of inner and outer membrane fractions

Total membrane pellets (section 2.7.5) were resuspended in 200 µl sterile TE buffer (10 mM Tris, 1 mM EDTA pH 8.0). A sucrose gradient made up in TE buffer was composed of the following layers (from bottom to top): 0.7 ml of 72% w/v sucrose, 1.2 ml of 49% w/v sucrose, and 0.9 ml of 26% w/v sucrose. 200 µl of total membrane pellets were layered on top of the sucrose gradient. Sucrose gradient preparations were centrifuged at 260,000 g for 22 hours at 4°C. Separation of inner and outer membrane
was visually defined by a whitish band (outer membrane) at the interface between the 72% w/v and 49% w/v sucrose layer and a diffused yellow band (inner membrane) above the whitish band. Inner and outer membranes were extracted using a needle and syringe and diluted up to 3 ml with 10 mM Tris pH 8.0 in fresh ultracentrifuge tubes. Membranes were pelleted by centrifugation at 260,000 g for 1 hour at 20°C. Inner and outer membrane pellets were stored on ice or at -20°C.

2.7.7 Cysteine-scanning labelling of proteins

Bacterial strains were grown and induced with 0.1% w/v arabinose for 22 hours as described in section 2.7.1. Cells were pelleted by centrifugation at 6000 g for 20 minutes and washed once with PBS. 250 ml cells were resuspended in 25 ml PBS pH 7.5. For each test strain two independent samples were prepared: biotin-labelled cells and unlabelled control cells. Future experiments would also prepare a third sample simultaneously: biotin-labelled cells pre-blocked with monobromo(trimethyl-ammonio)bimane bromide (qBBr). Biotin-labelled cells were incubated with biotin maleimide (Sigma) at a final concentration of 0.2 mM for 10 minutes at room temperature. The reaction was stopped by incubating the sample with 2% v/v 2-mercaptoethanol for 10 minutes at room temperature. The qBBr-pre-blocked cells would be incubated with a final concentration of 1 mM qBBr (Sigma) for 10 minutes at room temperature. qBBr would be removed by washing the cells three times with PBS. Pre-blocked samples would then be incubated with 0.2 mM biotin maleimide for 15 minutes at room temperature and the reaction would be stopped as described for biotin-labelled cells. All samples were then washed three times in an equal volume of PBS and subsequently used to either prepare total or outer membranes (section 2.7.5 or 2.7.6). Biotinylation was detected by Western blotting using near-infrared dye-labelled streptavidin (section 2.7.13).

2.7.8 Nickel-affinity chromatography of histidine-tagged proteins

Histidine (his)-tagged proteins were extracted from preparations of inner membranes as follows. Thawed total or inner membranes (section 2.7.5 or 2.7.6) pellets were resuspended in 200 µl sterile 50 mM Tris pH 8.0 and mixed with n-Decyl-β-D-maltopyranoside (DM) (Merck Chemicals) to a final concentration of 2% w/v. The
suspension was incubated on ice for 40 minutes. Suspensions were mixed with loading buffer (57.5 mM Tris, 58.7 mM NaCl, 12.5 mM imidazole pH 8.0) to a final concentration of 0.4% v/w DM (designated load). 1 ml HisTrap FF nickel columns (GE Healthcare) were washed with 5 column volumes of equilibration buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, 0.2% w/v DM pH 8.0). The load was then gradually passed through the column and the effluent was collected (designated flow through). The columns were washed with 5 column volumes of wash A (50 mM Tris, 500 mM NaCl, 10 mM imidazole pH 8.0) and then washed with 5 column volumes of wash B (50 mM Tris, 500 mM NaCl, 50 mM imidazole pH 8.0). The remaining proteins were eluted with elution buffer (50 mM Tris, 500 mM NaCl, 250 mM imidazole pH 8.0) and collected in 1 ml fractions (designated E1 – E5). Purification samples were stored on ice for further use. Where appropriate, samples were desalted using PD-10 desalting columns (GE Healthcare) according to the manufacturer's instructions.

2.7.9 Nickel-affinity chromatography combined with TEV protease cleavage

Proteins were extracted from preparations of total membranes (section 2.7.5) as follows. Thawed total membrane pellets were resuspended in 200 µl filter-sterilised 50 mM Tris pH 8.0 and mixed with DM to a final concentration of 2% w/v. The suspension was incubated on ice for 40 minutes and then mixed with loading buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole) to a final concentration of 0.4% v/w DM (designated load). 1 ml HisTrap FF nickel columns (GE Healthcare) were washed with 5 column volumes of equilibration buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.2% w/v DM pH 8.0). The load was then gradually passed through the column and the effluent was collected (designated flow through). The columns were washed with 5 column volumes of washing buffer (50 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole) and the effluent was collected (designated 20 mM imidazole wash). The columns were equilibrated with 5-10 column volumes of TEV protease cleavage buffer (50 mM Tris pH 8.0, 300 mM NaCl, 1mM dithiothreitol (DTT), 15% v/v glycerol). For on-column cleavage of his-tagged proteins with Tobacco etch virus (TEV) protease, 1 ml of TEV protease (Roberts lab) diluted in cleavage buffer (80µg protease:8mg substrate) was loaded onto the columns. The reaction was incubated for 16 hours at 4°C. The column was then washed with 5 column volumes of protease cleavage buffer and the released proteins were collected.
(designated protease cleavage fraction). The remaining proteins were eluted off with 5 column volumes of elution buffer (50 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole) and collected in 1 ml fractions (designated E1 – E5). Purification samples were stored on ice for further use.

2.7.10 Size exclusion chromatography

All buffers used subsequently were filtered through a 0.2 µm membrane filter disc (Whatman) and degassed prior to use. Total membranes (section 2.7.5) were resuspended in 50 mM Tris, 100 mM NaCl pH 8.0 containing 2% w/v DM and were incubated for 40 minutes on ice. The solubilised membranes were centrifuged at 260,000 g in an ultracentrifuge for 20 minutes to remove debris. A Superdex 200 gel filtration column (GE Healthcare) with a column volume of 25 ml, connected to a fast protein liquid chromatography ÄKTApurifier UPC 10, was equilibrated with equilibration buffer (50 mM Tris, 100 mM NaCl, pH 8.0). 500 µl of the solubilised membranes were loaded onto the column. The column was washed with size exclusion buffer (50 mM Tris, 100 mM NaCl pH 8.0, 0.2% w/v DM). The potential protein complex was eluted at a flow rate of 0.5 ml/minute. Fractions were collected in 0.5 ml volumes and a chromatogram was recorded measuring the UV absorbance at 280 nm. Samples were stored on ice for further use. The molecular weights of the observed peaks were determined by first calculating the distribution coefficient \( K_{av} \) using the formula \( K_{av} = (V_e - V_o)/(V_t/V_o) \), where \( V_e \) is the elution volume, \( V_o \) is the void volume of the column and \( V_t \) is the total column volume. \( K_{av} \) was then used to derive the molecular weight of the peak from the standard calibration curve of the column.

2.7.11 Determination of protein concentration

Protein concentration was determined using the Bradford protein assay. Briefly, four dilutions (between 2.5 and 20 µl/ml) of the protein standard bovine serum albumin (BSA) were prepared in dH₂O, and duplicates of the test samples were diluted 1:100 in dH₂O. 800 µl of standard dilution or sample were mixed with 200 µl of Bradford reagent (Sigma), respectively. Mixtures were incubated for 5 minutes at room temperature and absorbance was measured at 595 nm. The protein concentration of the test samples was deduced from a standard curve of protein concentration (mg/ml)
against absorbance from the protein standard; values were multiplied by the dilution factor.

### 2.7.12 SDS-polyacrylamide gel electrophoresis

Proteins were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using the Mini Protean 3 Electrophoresis System (Bio-Rad). Table 2.4 shows the various compositions resolving gels and the composition of the stacking gel used during the study. A resolving gel with a constant polyacrylamide concentration according to the size of the proteins to be examined was used. The resolving gel was covered with a layer of isopropanol. The resolving gel was allowed to polymerise and the isopropanol was removed after complete polymerisation. The stacking gel was then poured on top of the resolving gel and allowed to polymerize. Protein samples were resuspended in 2x or 10x SDS-loading buffer (100 mM Tris pH 6.8, 2% v/v SDS, 20% v/v glycerol, 0.02% v/v bromophenol blue, 0.7% v/v β-mercaptoethanol, or 2 M Tris pH 6.8, 4% v/v SDS, 40% v/v glycerol, 0.2% v/v bromophenol blue, 0.5% v/v β-mercaptoethanol, respectively) and were boiled for 5 minutes unless otherwise stated. Pre-stained molecular weight markers (Bio-Rad) were run alongside the samples to indicate molecular weights. Samples were electrophoresed in SDS-PAGE running buffer (25 mM Tris, 192 mM glycine, 0.1% w/v SDS) at 80 V for 10 minutes and at 180 V for the remaining time. Gels were visualized either by staining with Coomassie Brilliant Blue stain solution (0.8 g L⁻¹ Coomassie Brilliant Blue R-250, 40% v/v methanol, 10% v/v acetic acid, 50 v/v dH₂O) for at least 2 hours, rocking and were subsequently destained with destain solution (40% v/v methanol, 10% v/v acetic acid, 50% v/v dH₂O), or gels were stained with Coomassie-based Instant Blue stain (Invitrogen) for 2 hours, rocking and destained using dH₂O.

### 2.7.13 Western Blotting

For the purpose of Western blotting, proteins separated by SDS-PAGE (section 2.7.12) were transferred from the gel to a polyvinylidene difluoride (PVDF) membrane (Westram S, Whatman). Prior to transfer, the PVDF membrane was pre-soaked in methanol for at least 1 minute and then soaked in Western transfer buffer (48 mM Tris, 40 mM glycine, 20% v/v methanol) for at least 10 minutes. The SDS-PAGE gel and 4 pieces of Whatman 3MM paper were soaked in Western transfer buffer for at least 20
minutes. The following layers were then placed onto a Semi-Dry Transfer Cell (Bio-Rad) (from bottom to top): two pieces of Whatman 3 MM paper, PVDF membrane, SDS-PAGE gel, and two pieces of Whatman 3 MM paper. Proteins were transferred at 15 V for 20 minutes. The membrane was then processed for either chemiluminescent or near-infrared fluorescence detection (sections 2.7.13.1 and 2.7.13.2, respectively). Table 2.5 provides details about the antibodies and conjugated proteins used in this study.

### 2.7.13.1 Chemiluminescent detection

Following protein transfer onto the PVDF membrane (section 2.7.13), the membrane was soaked in PBS-Tween-4% BSA (PBS, 0.1% v/v Tween-20 supplemented with 4% w/v BSA) overnight with rocking at 4°C. The membrane was washed three times for 10 minutes in PBS-Tween (PBS, 0.1% v/v Tween-20) before being sealed in a bag containing 6 ml of PBS-Tween-4% BSA supplemented with the appropriate primary antibody (Table 2.5). The membrane was incubated for 1 hour with rocking at room temperature or overnight with rocking at 4°C. After incubation, the membrane was washed three times as before, and was then sealed in a bag containing 6 ml of PBS-Tween-4% BSA with the appropriate secondary antibody (Table 2.5). The membrane was incubated for 1 hour with rocking at room temperature. After three washing steps, the membrane was incubated with 1-2 ml of horseradish peroxidase (HRP) chemiluminescent substrate (Western Lightning Kit, Perkin Elmer) for 1-5 minutes. The membrane was then exposed to light-sensitive film (Biomax XAR, Kodak) and developed using an X-ray film processor (Xonograph Imaging Systems Compact XA).

### 2.7.13.2 Near-infrared fluorescence detection

Following protein transfer onto the PVDF membrane (section 2.7.13), the membrane was soaked in PBS-5% milk (PBS supplemented with 5% w/v skimmed milk powder) overnight with rocking at 4°C (blocking step). The membrane was washed three times for 10 minutes in PBS-Tween (PBS, 0.1% v/v Tween-20) before being sealed in a bag containing 6 ml of PBS-Tween-5% milk (PBS, 0.1% v/v Tween-20 supplemented with 5% w/v skimmed milk powder) supplemented with the appropriate primary antibody (Table 2.5). The membrane was incubated for 1 hour with rocking at room temperature or overnight with rocking at 4°C. After incubation, the membrane was washed three
times as before, and was then sealed in a bag containing 6 ml of PBS-Tween and LI-COR Blocking Buffer (1:1) supplemented with the appropriate secondary antibody (Table 2.5). The membrane was incubated for 1 hour with rocking at room temperature in darkness. After incubation, the membrane was washed three times as before, in darkness. To remove residual Tween-20, the membrane was rinsed in PBS, in darkness. The membrane was scanned using the Odyssey® Infrared Imaging System (LI-COR), according to the manufacturer's instructions. When near-infrared dye-labelled streptavidin was used, the membrane was washed three times after the blocking step and incubated in a bag containing 6 ml of PBS and LI-COR Blocking Buffer (1:1) supplemented with streptavidin. The membrane was incubated for 1 hour with rocking at room temperature and then washed three times and visualised as described for antibody detection above.
Table 2.4 Composition of SDS-PAGE gels used during the study.

<table>
<thead>
<tr>
<th>Component</th>
<th>Resolving gels (2 gels, 5 ml/gel)</th>
<th>Stacking gel (2 gels, 2.5 ml/gel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide percentage</td>
<td>8%</td>
<td>10%</td>
</tr>
<tr>
<td>Approximate protein size range</td>
<td>40-250 kDa</td>
<td>30-200 kDa</td>
</tr>
<tr>
<td>Component</td>
<td>ml</td>
<td>ml</td>
</tr>
<tr>
<td>1.0 M Tris pH 6.8</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>1.5 M Tris pH 8.8</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>30% v/v polyacrylamide (National Diagnostics)</td>
<td>2.7</td>
<td>3.3</td>
</tr>
<tr>
<td>Sterile dH₂O</td>
<td>4.6</td>
<td>4.0</td>
</tr>
<tr>
<td>10% w/v sodium dodecyl sulphate (SDS)</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>10% w/v ammonium persulphate (APS)</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>N,N,N,N'-tetramethylene-diamine (TEMED) (Biorad)</td>
<td>0.004</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 2.5 Antibodies and other conjugated proteins used during the study.

<table>
<thead>
<tr>
<th>Primary antibody</th>
<th>Secondary antibody (HRP or IRDye®)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibody</td>
<td>Working dilution</td>
</tr>
<tr>
<td>Anti-KpsE (Rabbit polyclonal, Roberts lab)</td>
<td>1:3000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-KpsD (Rabbit polyclonal, Roberts lab)</td>
<td>1:3000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-KpsT (Rabbit polyclonal, Roberts lab)</td>
<td>1:3000</td>
</tr>
<tr>
<td>Anti-his (Penta-his mouse monoclonal IgG1, Qiagen)</td>
<td>1:1000</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-myc (c-Myc-tag mouse monoclonal, New England Biolabs)</td>
<td>1:1000</td>
</tr>
<tr>
<td>Anti-β-lactamase (β-lactamase mouse monoclonal IgG1, Fisher Scientific)</td>
<td>1:100</td>
</tr>
</tbody>
</table>

Other conjugated proteins

<table>
<thead>
<tr>
<th>Target</th>
<th>IRDye®-conjugated protein</th>
<th>Working dilution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotin maleimide (Sigma) [biotinylated proteins]</td>
<td>Streptavidin IRDye® (Licor)</td>
<td>1:10,000</td>
</tr>
</tbody>
</table>
Chapter 3

Analysis of the interaction of KpsE with KpsM and KpsD
3.1 Introduction

Previous work on the topology of cytoplasmic membrane protein KpsE has identified a cytoplasmic N-terminus, a short N-terminal transmembrane region, a large periplasmic domain, and a periplasmic C-terminus which is closely associated with, but not integrated into the cytoplasmic membrane (Rosenow et al., 1995a; Phoenix et al., 2001). Secondary structure predictions using the prediction software PROFsec (Rost et al., 2004) suggest that KpsE is primarily made of α-helical domains and a few short β-sheets (Figure 3.1). Co-purification studies imply that KpsE interacts directly with the inner membrane component of the ABC transporter KpsM, and with the outer membrane protein KpsD (Hudson, 2009). Previous mutagenesis analysis of KpsE revealed that the N-terminus (residues 1-26) is not essential for KpsE function and K5 export (Hudson, 2009). This is interesting, since the N-terminal as the only cytoplasmic part of KpsE had been a plausible candidate region for interaction with KpsM. Point mutations of conserved residues (proline 36, methionine 37, leucine 41, tyrosine 43, tyrosine 52, serine 54, glutamine 55, and serine 56) to alanine located in or immediately after the transmembrane region (residues 27-48) did not affect KpsE function or K5 export either (Hudson, 2009). However, complete deletion of the transmembrane domain abolished KpsE function in the transport of K5 polysaccharide (Hudson, 2009), suggesting an important role of this region in protein function and possibly protein-protein interaction. Here, mutagenesis studies were employed to further identify regions essential for KpsE function and important in the interaction with KpsM and KpsD.

3.1.1 Use of the pBAD28 vector system

The majority of work in Chapters 3, 4 and 5 uses the pBAD28 vector system. The choice of vector was based on two factors. Firstly, pBAD vectors provide a high level of control over expression levels by offering the ability to custom-select concentrations of arabinose for induction. This is important in the study of capsule transport proteins as the use of titratable concentrations of arabinose allows appropriate balancing between too low expression, which hinders analysis, and too high expression, which might affect native protein conformation. Secondly, previous generation and use of a pBAD system for KpsMTED (Hudson, 2009) illustrated the convenience of employing a system which can be customised by adding and removing Kps genes using easy
Figure 3.1 Model of the secondary structure and topology of KpsE. The secondary structure and topology were constructed on the basis of previous experimental data (Arrecubieta et al., 2001; Phoenix et al., 2001) and predictions by PROFsec from the PredictProtein Server (Rost et al., 2004). α-Helical regions are depicted in blue and β-sheets are depicted in orange. Topology information is given in green. The locations of the APQ motif (which has a homologous region in AcrA and MexA, cf. section 3.2.1) and the C-terminal α-helix (cf. section 3.3.1) are indicated in black.
cloning steps. Previous detergent screening found DM to be the best choice for solubilising membrane-bound K5 transport proteins (Hudson, 2009).

3.2 Mutagenesis in the APQ motif of KpsE

Previous work on KpsE identified alanine 287 as essential residue for KpsE function and a point of interaction with KpsM and KpsD (Hudson, 2009). Site-directed mutagenesis of alanine 287 to valine or glycine abolished the ability of KpsE to function as part of the K5 export complex. KpsE A287G did not affect co-purification of the protein by hisKpsM but rather increased the expression of mutant KpsE. However, co-purification analysis of hisKpsM and KpsE A287V resulted in reduced KpsE purification. This indicated that, while KpsM remains stable in the presence of A287V, the change from alanine to valine leads to a disruption of the interaction between KpsE and KpsM. Both mutants were also unable to co-purify KpsD, indicating that alanine 287 is involved in KpsE-KpsD interaction (Hudson, 2009).

3.2.1 Clustal alignment of APQ motif in KpsE with related proteins

Alanine 287 belongs to an alanine-proline-glutamine motif (residues 287-289) which is conserved in KpsE and the membrane fusion protein MexA of the multi-drug efflux exporter MexAB-OprM of Pseudomonas aeruginosa (Figure 3.2). The membrane fusion protein AcrA of the E. coli AcrAB-TolC multi-drug efflux pump shares homology with MexA at this region but contains a valine instead of an alanine within the motif (Figure 3.2). Like the T1SS, multi-drug effluxers are tripartite systems with an inner membrane transporter, a membrane fusion protein, and an outer membrane channel exporter (Kumar and Varela, 2012). Alignment of the region containing the APQ motif of KpsE with CtrB and BexC reveals the absence of any homology in this region (Figure 3.2). This suggests that KpsE might not share structural similarities with CtrB and BexC at residues 287-289, but might rather be similar to MexA and AcrA in this region. In order to test whether the two remaining residues of the APQ motif, proline 288 and glutamine 289, play a similarly significant role in KpsE function and protein-protein interaction, site-directed mutagenesis of the two residues was attempted.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>BexC</td>
<td>275</td>
<td>DEQTRQLSGNGNSAATQTA</td>
<td>293</td>
</tr>
<tr>
<td>CtrB</td>
<td>284</td>
<td>DQQLRAISGGGHSLSLNNQA</td>
<td>302</td>
</tr>
<tr>
<td>KpsE</td>
<td>279</td>
<td>DEEKSKitAQPQSDLKNRMA</td>
<td>297</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcrA</td>
<td>300</td>
<td>GLNPNAILVPOQGVRTPR</td>
<td>318</td>
</tr>
<tr>
<td>MexA</td>
<td>285</td>
<td>GVKQKAILAPQGVTRDLK</td>
<td>303</td>
</tr>
<tr>
<td>KpsE</td>
<td>279</td>
<td>DEEKSKitAQPQSDLKNRMA</td>
<td>297</td>
</tr>
</tbody>
</table>

Figure 3.2 Clustal alignment of KpsE residues 279 to 297 with CtrB and BexC, and AcrA and MexA. Multiple sequence alignments of selected residues were generated using ClustalW (Thompson et al., 1994). Identical residues are highlighted in red. KpsE of *E. coli* (gene bank accession number P62586), CtrB of *N. meningitidis* group B (P32014), BexC of *H. Influenzae* type b (P22930), AcrA of *E. coli* (P0AE06), and MexA of *Pseudomonas aeruginosa* (P52477).
3.2.2 Site-directed mutagenesis of glutamine 289 in KpsE

Site-directed mutagenesis of KpsE was carried out using the QuikChange® PCR-based method as described in section 2.5.3. Primers KpsE P288A F/R and KpsE Q289AF/R (Table 2.3) were used to introduce a single substitution from the native proline and glutamine to alanine in the plasmid pBADE, respectively. The QuikChange® reactions were transformed into the *E. coli* cloning strain DH5α, and the mutant plasmids were isolated (section 2.5.5). Colony PCR (Figure 3.3; section 2.5.4) and DNA sequencing (section 2.5.8) were used to identify and verify the mutations. Successful mutagenesis of glutamine to alanine resulted in the plasmid pBADE Q289A (Figure 3.4). Despite repeated attempts to mutate proline 288 to alanine, a KpsE P288A mutant could not be obtained.

A K5-specific bacteriophage sensitivity assay can be used to test for wild-type K5 polysaccharide expression and to indirectly determine whether a component of the K5 transport complex sufficiently functions for successful K5 export. The K5 bacteriophage assay was used to determine the function of KpsE Q289A in the *kpsE*-deficient encapsulated strain MSΔE strain. Assays were carried out in triplicates as described in section 2.6. MS101 and MSΔE (pBADE) were used as positive controls in the assay; MSΔE and MSΔE (pBADE A287V) served as negative controls. The plasmid pBADE Q289A was able to complement MSΔE, producing a plaque count similar to the wild-type strain (Table 3.1). This demonstrates that KpsE Q289A is able to function as part of the K5 export complex, suggesting that glutamine 289 is not essential for KpsE function. This result indicates that glutamine 289, in contrast to alanine 287, might not be required for an interaction with KpsM or KpsD, nor for interaction with other capsule proteins. As this residue appears not to play a crucial role in K5 export, no further analyses on glutamine 289 were carried out.
Figure 3.3 Colony PCR of site-directed mutagenesis product from pBADE Q289A and pBADE control. The plasmid pBADE Q289A was generated as described in sections 2.5.3-2.5.10. Samples were separated on a 1% agarose gel. Lane 1: colony PCR of pBADE using primers pBAD28 F and R, generating a 1.15 kb product; lane 2: colony PCR of pBADE Q289A using primers pBAD28 F and R, generating a 1.15 kb product.

Figure 3.4 Illustration of the plasmid pBADE Q289A. MCS: multiple cloning site; p1: arabinose-inducible pBAD promoter; p2: AmpR promoter; term: transcriptional terminator rnbT1; araC: arabinose transcriptional regulator; f1 ori and p15A ori: f1 and p15A origin of replication, respectively; AmpR and CmR: ampicillin and chloramphenicol resistant marker, respectively; NsiI: NsiI restriction site.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Plaque count at different K5 phage dilutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10^-6</td>
</tr>
<tr>
<td>MS101</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔE</td>
<td>0</td>
</tr>
<tr>
<td>MSΔE (pBADE)</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔE (pBADE A287V)</td>
<td>0</td>
</tr>
<tr>
<td>MSΔE (pBADE Q289A)</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔE (pBADEΔCterm)</td>
<td>0</td>
</tr>
<tr>
<td>MSΔEΔMT</td>
<td>0</td>
</tr>
<tr>
<td>MSΔEΔMT (pBADEhisMT)</td>
<td>53</td>
</tr>
<tr>
<td>MSΔEΔMT (pBADEΔCtermhisMT)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3.1 K5 bacteriophage assays of KpsE mutations and controls. TMTC: too many plaques to count.
3.3 Deletion of the C-terminus of KpsE

Previous theoretical analyses demonstrated that the C-terminus of KpsE (residues 357-377; cf. Figure 3.1) has membrane-interactive potential and the ability to form an amphipathic α-helix, with a wide hydrophobic face buried into the membrane and a narrow hydrophilic face on the outside interacting with the aqueous environment (Phoenix et al., 2001). Such amphipathic C-terminal membrane-anchors are also found in other inner membrane proteins, such as penicillin-binding proteins 5 and 6 (Gittins et al., 1994).

3.3.1 Clustal alignment of the C-terminus of KpsE with homologues

To date, KpsE is the only known PCP member with a predicted C-terminal amphipathic membrane anchor. Alignment of the C-terminal residues of KpsE with those of BexC and CtrB revealed that the residues forming the predicted α-helix in KpsE share 40% identity with the corresponding residues in BexC and CtrB, while BexC and CtrB share 70% identical residues in this region (Figure 3.5). Topology analysis of CtrB identified the C-terminus to be located in the cytoplasm and not in the periplasm as in the case of KpsE (Larue et al., 2011). The difference in composition of C-terminal amino acids in KpsE explains the structural divergence of the C-terminus in the two proteins and highlights the potential regional uniqueness of KpsE within members of the PCP family. It was thus hypothesised that the C-terminal α-helix of KpsE is crucial for KpsE function and might be involved in interaction with other members of the translocon complex.

3.3.2 Deletion mutagenesis of the KpsE C-terminus

To test the hypothesis of the importance of the C-terminus, a KpsE truncation mutant lacking residues 357-377 was generated. For this purpose, a truncated kpsE gene variant was amplified by PCR as described in section 2.5.2 by replacing residues 357-377 with a KpnI restriction site (encoding glycine and threonine). The first version of this fragment was kindly provided by Thomas Hudson and had been amplified using primers KpsE F and KpsE N-term del TM2 R (for N-terminal side), and KpsE C-term del TM2 F and KpsE C-term del TM2 R (for C-terminal side; (Table 2.3)). Using primers KpsE SacI F and KpsE TM2 XmaI R (Table 2.3), the resulting fragment was
Figure 3.5 Clustal alignment of KpsE residues 352 to 382 with CtrB and BexC. Multiple sequence alignments of the final 30 residues of KpsE, CtrB and BexC were generated using ClustalW (Thompson et al., 1994). Identical residues are highlighted in red. The black box highlights the predicted α-helical region in KpsE and the corresponding residues in BexC and CtrB. KpsE of *E. coli* (gene bank accession number P62586), CtrB of *N. meningitidis* group B (P32014) and BexC of *H. influenzae* type b (P22930).
used as template to amplify the second version of a C-terminal KpsE deletion fragment, kpsEΔC term (section 2.5.2), which contained the restriction sites XmaI and SacI to facilitate further cloning (Figure 3.6, lane 1). The fragment kpsEΔC term was cloned into the expression vector pBAD28, creating the construct pBADEΔCterm (Figure 3.7). The construct was transformed into DH5α and positive clones were identified by colony PCR and restriction digest (Figure 3.6, lanes 2-5) and verified by DNA sequencing. To analyse the function of truncated KpsE, pBADEΔCterm was transformed into MSΔE. The K5 bacteriophage assay was used to assess the ability of truncated KpsE to function within the K5 export complex. MS101 and MSΔE (pBADE) were used as positive controls; MSΔE served as negative control. Bacteriophage analysis showed that KpsEΔCterm was unable to complement MSΔE (Table 3.1), suggesting that the strain has a disrupted K5 export system. This indicates that the C-terminus of KpsE is crucial for KpsE function. Please note that analyses in section 3.5.2 suggest that the absence of the C-terminus of KpsE does not affect protein expression levels and membrane location.

3.4 Interaction of KpsD with KpsE lacking the C-terminus

It was hypothesised that the observed loss of function of KpsE after C-terminal deletion is caused by disrupted protein-protein interactions between KpsE and other translocon members. KpsD forms SDS-resistant, heat-labile dimers in the outer membrane (McNulty et al., 2006). The ability of KpsD to locate to the outer membrane and to dimerise is dependent on co-expression of KpsE (Arrecubieta et al., 2001; Hudson, 2009). This interaction is independent of KpsM expression (Hudson, 2009). KpsE and KpsD are likely to interact via their periplasmic domains, and the C-terminus of KpsE is a potential candidate for this interaction.

3.4.1 Localisation and dimerisation study of KpsD in the presence of KpsE lacking the C-terminus

To investigate whether the KpsE C-terminus is important for KpsE-KpsD interaction, the role of this region in KpsD outer membrane localisation and oligomerisation was analysed. KpsEΔCterm and KpsD were expressed from MSΔE (pBADEΔCterm), and soluble fractions and outer membrane preparations were analysed for the presence of KpsD. In the absence of KpsE, KpsD was absent from the outer membrane but could be detected in the soluble fraction (Figure 3.8).
Figure 3.6 Generation of the pBADE$_{ΔCterm}$ and pBADE$_{ΔCterm}hisMT$ constructs. The plasmid pBADE$_{ΔCterm}$ was generated as described in sections 2.5.2 and 2.5.4-2.5.10 (lanes 1-5). 1% agarose gel. Lane 1: PCR amplification of kpsE$_{ΔCterm}$, generating a 1.1 kb product; lane 2: colony PCR of pBADE using pBAD28 F and R primer, generating a 1.15 kb product; lane 3: colony PCR of pBADE$_{ΔCterm}$ using pBAD28 F and R primers, generating a 1.1 kb product; lane 4: pBADE$_{ΔCterm}$ digested with SacI and KpnI, generating a 1.1 kb product containing kpsE$_{ΔCterm}$ and a 4.7 kb product containing the pBAD28 vector backbone; lane 5: pBADE$_{ΔCterm}$ digested with Sacl and Xmal, generating a 1.1 kb product containing kpsE$_{ΔCterm}$ and a 5.8 kb product containing the pBAD28 vector backbone. The plasmid pBADE$_{ΔCterm}hisMT$ was generated as described in sections 2.5.2 and 2.5.4-2.5.10 (lanes 6-11). Lane 6: PCR amplification of hiskpsMT, generating a 1.5 kb product; lane 7: colony PCR of pBADE$_{ΔCterm}hisMT$ using pBAD28 F and R primers, generating a 2.6 kb product; lane 8: pBADE$_{ΔCterm}hisMT$ digested with Sacl and HindIII, generating a 2.6 kb product containing kpsE$_{ΔCterm}$ and hiskpsMT as well as a 5.8 kb product containing the pBAD28 vector backbone; lane 9: pBADE$_{ΔCterm}hisMT$ digested with KpnI and HindIII, generating a 1.5 kb product containing hiskpsMT and a 6.9 kb vector product containing the pBAD28 backbone and partial pBADE$_{ΔCterm}$; lane 10: pBADE$_{ΔCterm}hisMT$ digested with NsiI, generating a 2.5 kb product containing the region between the pBAD28 NsiI site and the NsiI site on his-tag as well a 5.9 kb product containing the region between the NsiI site on the his-tag and the pBAD28 NsiI site; lane 11: pBADE$_{ΔCterm}hisMT$ digested with Sacl and KpnI, generating a 1.1 kb product containing kpsE$_{ΔCterm}$ and a 7.3 kb vector product containing the pBAD28 vector backbone and hiskpsMT.
Figure 3.7 Illustrations of the plasmids pBADE_{ΔCterm} and pBADE_{ΔCtermhisMT}. his: 6X histidine-tag sequence; MCS: multiple cloning site; p1: arabinose-inducible pBAD promoter; p2: AmpR promoter; term: transcriptional terminator rrnB T1; araC: arabinose transcriptional regulator; f1 ori and p15A ori: f1 and p15A origin of replication, respectively; AmpR and CmR: ampicillin and chloramphenicol resistant marker, respectively; NsiI: NsiI restriction site.
**Figure 3.8** Location of KpsD in outer membrane and soluble fraction of MSΔE (pBADE\(_{ΔCterm}\)) and controls. Western blot analyses of outer membranes (top row) and soluble fractions (bottom row) using anti-KpsD antibody. Proteins were separated on an 8% SDS-PAGE gel prior to electrophoretic transfer. Lane 1: MS101; lane 2: MSΔE; lane 3: MSΔE (pBADE); lane 4: MSΔE (pBADE\(_{ΔCterm}\)). Loadings were standardised according to protein quantity (outer membranes) or optical densities of whole cells used to prepare soluble fractions.

<table>
<thead>
<tr>
<th>Lane</th>
<th>Sample Description</th>
<th>Western Blot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MS101</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MSΔE</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MSΔE (pBADE)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MSΔE (pBADE(_{ΔCterm}))</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.9** Dimerisation of KpsD in outer membrane from MSΔE (pBADE\(_{ΔCterm}\)). Western blot analysis of outer membranes from MSΔE (pBADE\(_{ΔCterm}\)) using anti-KpsD antibody. Samples were incubated at either 25°C or 100°C prior to loading onto an 8% SDS-PAGE gel. Lane 1: outer membranes incubated at 25°C; lane 2: outer membranes incubated at 100°C. Loadings were standardised according to protein quantity.
In wild-type MS101 and MSΔE (pBADE), KpsD was located in the outer membrane but was also found in the soluble fraction (Figure 3.8). Likewise, in MSΔE (pBADE_{ΔCterm}), KpsD was detected in the outer membrane and was also found in the soluble fraction (Figure 3.8). This demonstrates that KpsD locates to the outer membrane in a wild-type manner when co-expressed with KpsE_{ΔCterm}. This result implies that outer membrane localisation of KpsD is not affected by the absence of the C-terminal domain of KpsE. Previous work established that in capsule-producing wild-type MS101, KpsD forms dimers at 25°C (McNulty et al., 2006). Dimers can be broken down into monomers by denaturing KpsD at 100°C (McNulty et al., 2006). To investigate whether truncated KpsE affects KpsD dimerisation, outer membranes of MSΔE (pBADE_{ΔCterm}) were incubated at conditions supporting KpsD monomer (incubation at 100°C) and dimer (incubation at 25°C) formation before Western blot analysis. As in the wild-type, KpsD expressed in the presence of KpsE_{ΔCterm} assembled as dimers in the outer membrane when incubated at 25°C and was observed as monomers after incubating the sample at 100°C prior to SDS-PAGE. (Figure 3.9). Taken together, these results show that co-expression with KpsE_{ΔCterm} does not have an adverse affect on KpsD membrane localisation and oligomerisation. This implies that the C-terminal domain of KpsE is not crucial for KpsE-KpsD interaction.

### 3.5 Interaction of KpsM with KpsE lacking the C-terminus

Deletion of the C-terminus of KpsE abolished KpsE function as part of the K5 translocon (section 3.3). Since results from section 3.4 suggest that the KpsE C-terminus does not to play an essential role in KpsE-KpsD interaction, a role of the C-terminus in KpsE-KpsM interaction appears plausible. As mentioned earlier (section 1.5.3.4), his-tagged KpsM has been shown to co-purify KpsE (Hudson, 2009). This interaction is independent of KpsT and the K5 polysaccharide substrate (Hudson, 2009). To test the hypothesis that the observed loss of function in KpsE_{ΔCterm} is due to a disrupted protein-protein interaction with KpsM, a co-purification approach identical to the one previously used by Hudson (2009) to identify the KpsE-KpsM interaction was applied.
3.5.1 Generation of construct carrying *hisKpsM* and *kpsE* lacking the C-terminus

A previous study on KpsEDMT successfully used N-terminally his-tagged KpsM and could not detect any apparent adverse effects caused by the tag on KpsM and subsequent analyses (Hudson, 2009). For the purpose of co-purification studies, a fragment containing his-tagged *kpsM* plus *kpsT* carrying XbaI and HindIII restriction sites was amplified using primers hisKpsMT XbaI F and KpsT HindIII R (Figure 3.6, lane 6; section 2.5.2; Table 2.3). The resulting *hisKpsMT* fragment was cloned downstream of *kpsEΔCterm* in the vector pBADEΔCterm, creating plasmid pBADEΔCtermhisMT (Figure 3.7). The construct was transformed into DH5α and positive clones were identified by colony PCR and restriction digest (Figure 3.6, lanes 7-11) and verified by DNA sequencing. To confirm the functional results obtained from the pBADEΔCterm plasmid earlier, pBADEΔCtermhisMT was transformed into the capsule-producing strain MSΔEΔMT which lacks chromosomal *kpsEMT*. As expected, pBADEΔCtermhisMT was unable to complement MSΔEΔMT in a K5 bacteriophage assay (Table 3.1).

3.5.2 Expression and location study of hisKpsM and KpsE lacking the C-terminus

To test protein expression from pBADEΔCtermhisMT and compare it to the expression of native proteins from pBADEhisMT in a capsule-deficient background, whole cells of DH5α (pBADEΔCtermhisMT) and DH5α (pBADEhisMT) were induced with 0.1% arabinose for 3 hours and analysed by Western blot using anti-KpsE and anti-his antibodies (Figure 3.10, lanes 1, 2, 5 and 6). Both KpsEΔCterm and hisKpsM from pBADEΔCtermhisMT appeared to be expressed at similar levels to KpsE and hisKpsM from pBADEhisMT. This suggests that the absence of the C-terminus of KpsE did not affect expression levels of KpsEΔCterm and hisKpsM in the absence of the K5 polysaccharide substrate. To investigate whether truncated KpsE affects membrane location of KpsE and hisKpsM, induced whole cells of DH5α (pBADEΔCtermhisMT) and DH5α (pBADEhisMT) were used to prepare total membranes and inner membranes. Preparations were analysed by Western blot using anti-KpsE and anti-his antibodies (Figure 3.10, lanes 3, 4, 7 and 8). KpsE, KpsEΔCterm and hisKpsM were detected in both total and inner membrane preparations from DH5α (pBADEΔCtermhisMT) and DH5α (pBADEhisMT). The amount of KpsE, KpsEΔCterm and hisKpsM protein present appeared similar in the preparations from the two strains.
This implies that the C-terminal truncation did not abolish KpsEΔCterm native inner membrane location or expression levels. Detection of hisKpsM in the inner membrane suggests that deletion of the C-terminus might not affect the inner membrane stability of hisKpsM.

### 3.5.3 Co-purification study of hisKpsM and KpsE lacking the C-terminus

To analyse whether deletion of the KpsE C-terminus abolishes the ability of hisKpsM to co-purify KpsE, inner membrane preparations from DH5α (pBADEhisMT) and DH5α (pBADEΔCtermhisMT) were solubilised with 0.2% DM and subjected to nickel-affinity chromatography (section 2.7.8). The load (inner membranes solubilised with DM), flowthrough, washes and elution fractions were analysed by SDS-PAGE (data not shown). As the bands predicted to be the purified proteins were not easily visible, all purification steps were analysed by Western blot using anti-KpsE and anti-his antibodies. While KpsE was detected in all steps of the purification procedure, KpsEΔCterm was detected in the load, flowthrough and wash A, and faintly in wash B, but not in any elution fractions (Figure 3.11). hisKpsM from DH5α (pBADEhisMT) was only detected in load and elution fraction 1 and 2 (Figure 3.11). The failure to detect hisKpsM in flowthrough and washes was likely to be due to low levels of protein released from the nickel column at these stages. hisKpsM from DH5α (pBADEΔCtermhisMT) was detected in the load but not in any elution fractions (Figure 3.11). The presence of KpsE and hisKpsM in the same elution fraction confirmed that KpsE can be co-purified by hisKpsM. However, in the absence of the KpsE C-terminus hisKpsM appears to lose this ability, as both KpsEΔCterm and hisKpsM were absent from the elution fractions.
Figure 3.10 Expression and membrane location of KpsE and KpsM in DH5α (pBADEhisMT) and DH5α (pBADE_{ΔCterm}hisMT). Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row) of whole cells of DH5α (pBADEhisMT) (lanes 1-4) and DH5α (pBADE_{ΔCterm}hisMT) (lanes 5-8) induced for 3 hours with 0.1% arabinose. Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Lane 1: uninduced whole cells of DH5α (pBADEhisMT); lane 2: induced whole cells of DH5α (pBADEhisMT); lane 3: total membranes of DH5α (pBADEhisMT); lane 4: inner membranes of DH5α (pBADEhisMT); lane 5: uninduced whole cells of DH5α (pBADE_{ΔCterm}hisMT); lane 6: induced whole cells DH5α (pBADE_{ΔCterm}hisMT); lane 7: total membranes of DH5α (pBADE_{ΔCterm}hisMT); lane 8: inner membranes of DH5α (pBADE_{ΔCterm}hisMT). Loadings were standardised according to optical densities.

Figure 3.11 Co-purification study of hisKpsM with KpsE and KpsE_{ΔCterm}. Western blot analyses using anti-KpsE (top row) and anti-his (bottom row) antibodies of inner membrane purifications from DH5α (pBADEhisMT) (top panel) and DH5α (pBADE_{ΔCterm}hisMT) (bottom panel). Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised between KpsEhisMT and KpsE_{ΔCterm}hisMT samples. Lane 1: sample load; lane 2: flowthrough; lane 3: 10 mM imidazole wash; lane 4: 50 mM imidazole wash; lane 5: elution fraction 1.
3.6 Discussion

3.6.1 Glutamine residue 289 of KpsE is not essential for native KpsE function and K5 export

There are no mutational data available in the literature about the APQ motif or neighbouring residues of either MexA or AcrA. However, it has been established that in MexA the APQ motif, located at positions 293-295, is part of the C-terminal region which forms the membrane proximal domain together with the membrane-anchored N-terminus (Symmons et al., 2009). The membrane proximal domain is conserved among bacterial efflux pumps; this region is closely homologous in MexA of *P. aeruginosa* and AcrA of *E. coli*. The membrane proximal domain forms a β-roll which directly interacts with the inner membrane transporter component (MexB in *P. aeruginosa* and AcrB in *E. coli*) (Symmons et al., 2009).

In contrast to MexA and AcrA, the APQ motif of KpsE is located in the periplasmic loop. Alignment of the region located around the APQ motif in KpsE showed no homology with other PCP proteins. This suggests that the structural organisation of KpsE in this region differs from other PCPs. The region around the APQ motif is located in one of up to three predicted coiled-coil regions within KpsE, which are formed by residues 264 to 292 (predicted using PairCoil2 (McDonnell et al., 2006), cf. Figure 3.1). Coiled-coils are folding motifs (Burkhard et al., 2001), and in KpsE this region might be involved in folding the protein back to the inner membrane, providing structural stability. This fold might be essential for correct interaction with KpsM and, particularly with KpsD. The data presented here showed that KpsE Q298A did not impair K5 polysaccharide export and is thus not essential for KpsE function. This implies that this residue might not be involved in crucial interactions with neighbouring proteins. If this hypothesis is correct, co-purification analysis with hisKpsE Q289A and KpsD as well as KpsE Q289A and hisKpsM would be expected to result in wild-type purification profiles.

It remains unclear why generation of a KpsE P288A variant was unsuccessful. It seems most likely that the issue was due to the use of non-optimal mutagenesis primers. It might be worthwhile to investigate whether changing primer length and mutating proline to a different residue than alanine solves the problem.
The data on KpsE Q289A presented here do not necessarily exclude that glutamine 289 is involved in KpsE-KpsM or KpsE-KpsD interaction by acting together with neighbouring residues. Mutagenesis of several residues in the neighbouring region of the APQ motif might be worth considering in order to further elucidate the role of these residues in protein-protein interaction. Interestingly, a BLAST alignment (Altschul et al., 1990) of KpsE and KpsD revealed glutamine 289 is part of a QGDXXNXXXXX_{acidic}Fxe motif which the two proteins share (residues 289-301 of KpsE and 255-257 of KpsD). This motif is located in regions predicted to be periplasmic in both proteins (cf. Figures 3.1 and 5.2). In the haemolysin A exporter system, HlyD and TolC are thought to interact via a homologous region in the periplasm (cf. section 3.6.3 (Schulein et al., 1994)). In group 1 polysaccharide export, interaction between Wzc and Wza occurs via their respective periplasmic domains (Collins et al., 2007). It is predicted that interaction of KpsE and KpsD will also occur via their periplasmic residues. The results from the C-terminal deletion study suggest that the C-terminus of KpsE is not a crucial region for interaction with KpsD, indicating that other regions play a more important role. Mutagenetic removal or substitution of the complete QGD motif in KpsED might determine the significance of this region for KpsE-KpsD interaction.

3.6.2 C-terminal domain of KpsE is crucial for native KpsE function and K5 export

The results presented here revealed that the C-terminal domain of KpsE is essential for correct KpsE function. This reflects the situation in HlyD. As in KpsE, the C-terminal region of HlyD is essential for HlyD function and HlyA export and is thus predicted to be an important interaction point with other translocon proteins (Schulein et al., 1994). Recent work on KpsE generated point mutations in prolines 341, 344 and 350, which are located in a proline-rich motif (residues 341-369) conserved among PCP proteins (Larue et al., 2011). The proline-rich motif is located just before and at the beginning of the C-terminal transmembrane $\alpha$-helix of PCP proteins. The complete motif, PX$_2$PX$_4$SPKX$_{11}$GGMXGAG, is only present in some PCPs, and KpsE shares the least number of conserved amino acids, possessing only the three prolines from the motif (Larue et al., 2011). These three prolines are located immediately preceding the C-terminal domain of KpsE. The final one or two prolines in the proline-rich motif are
important to wild-type function of PCP-1 proteins ExoP of *Rhizobium meliloti* (Becker and Pühler, 1998) and Wzz of *Shigella flexneri* (Daniels and Morona, 1999), respectively. Interestingly, in KpsE neither separate nor combined proline to alanine substitutions of prolines 341, 344 and 350 impaired KpsE function (Larue et al., 2011). This result emphasizes the uniqueness of KpsE within the final 30 residues. Combined with identifying the importance of the C-terminal domain of KpsE, it suggests that the final 20 amino acids are indeed the residues which play this essential role for KpsE function.

In order to identify potential local homologues which could be used for structure-function predictions of the KpsE C-terminus, a BLAST search (Altschul et al., 1990) using the final 30 amino acids of KpsE was carried out. The search identified four proteins, a putative acyltransferase of *Pseudoalteromonas piscicida*, a drug efflux transporter of *Burkholderia cenocepacia*, a major facilitator transporter of *S. epidermidis*, and a peroxisomal fatty acyl CoA ABC transporter ATP-binding protein of *Synechococcus*. These proteins shared 57%, 58%, 58% and 63% identical residues with the C-terminal KpsE sequence, respectively. However, none of the identified regions were located at the C-termini of the respective proteins. All four proteins had been identified during whole-genome sequencing of the respective bacterium and had not been further characterised, making a structure-function deduction impossible. Interestingly though, the four identified proteins are all membrane transporters.

The C-terminal domain of KpsE is predicted to form an amphipathic α-helix. The inner membrane penicillin-binding protein 5 of *E. coli*, which forms an amphipathic α-helix with its final 20 C-terminal residues (Jackson and Pratt, 1987), can be used as model for the potential role of the C-terminus in KpsE. The C-terminal domain of the penicillin-binding protein 5 has an important functional role. The C-terminus is essential for the stability of the cell shape of *E. coli* and the lack of this domain is lethal to the bacterium (Nelson and Young, 2001). However, the C-terminus is also essential for the membrane-association of the protein, as it functions as inner membrane anchor (Jackson and Pratt, 1987). Crystal studies of the final 20 residues revealed a structure of a helix-bend-helix-turn-helix motif, and not of a single contiguous helix as previously predicted (O’Daniel et al., 2010). The amphipathic nature of the C-terminus provides a very strong association with the periplasmic face of the inner membrane, which is not completely disrupted when the C-terminal residues are removed (O’Daniel et al., 2010).
In the native protein, the hydrophobic core of the C-terminus is positioned into the membrane in a parallel (horizontal) manner which allows most hydrophobic residues of the domain to interact with the interior of the membrane lipids. Due to the bend in domain structure, the C-terminal and N-terminal ends of the domain are bent upwards (vertically), interacting with the aqueous periplasmic environment (O'Daniel et al., 2010). Even though the primary sequences of KpsE and penicillin-binding protein 5 differ, both proteins might share a similar C-terminus structure, and the structural information of final 20 residues of the penicillin-binding protein 5 might be useful in predicting the structure of the KpsE C-terminus.

### 3.6.3 C-terminal domain of KpsE is not required for interaction of KpsE with KpsD

Previous work has shown that the ability of KpsD to locate to the outer membrane and to dimerise is dependent on KpsE but not on KpsM (Arrecubieta et al., 2001; Hudson, 2009). This suggests that the interaction between KpsE and KpsD does not require KpsM. The data presented here on the interaction of KpsD and KpsE_{ΔCterm} confirm this, since KpsE_{ΔCterm} retains the interaction with KpsD in the absence of the C-terminus which abolished, or at least diminished, the strong interaction with KpsM.

A study using cross-complementation assays of KpsDE and CtrAB demonstrated that when co-expressed, CtrA and CtrB can restore polysaccharide export in a KpsED-deletion strain (Larue et al., 2011). However, when either CtrA or CtrB were expressed in the absence of the partner protein they were unable to restore polysaccharide export in a KpsD- or KpsE-deletion strain, respectively (Larue et al., 2011). This strongly suggests that the interaction between CtrA and CtrB is highly specific, while the interaction of CtrB with the ABC transporter appears to be less specific. This interaction-specificity has also been reported between Wza and Wzc involved in colonic acid export of *E. coli* (Reid and Whitfield, 2005). A similar situation is likely to occur in the K5 transporter. The removal of the KpsE C-terminal did not affect the membrane localisation or oligomerisation state of KpsD. This suggests that the previously reported interaction between the two proteins does not require the C-terminal domain of KpsE. This result does not exclude the C-terminus as point of interaction, but it rather points towards other regions in KpsE which must be essential for association with KpsD.
The periplasmic loop of KpsE appears the most likely candidate for providing regions of interaction. In colonic acid and group 1 polysaccharide export, the periplasmic loop of Wzc is closely associated with the periplasmic region of Wza (Reid and Whitfield, 2005; Collins et al., 2007). As briefly mentioned in section 3.6.1, in the haemolysin A exporter computational analysis identified a periplasmic region of around 40 residues in HlyD which is homologous to periplasmic residues in TolC (Schulein et al., 1994). It is predicted that HlyD and TolC interact via this homologous region (Schulein et al., 1994).

In order to investigate whether similar homologous regions exist between KpsE and KpsD, BLAST (Altschul et al., 1990) was used to identify primary sequence alignments of the two proteins. Four aligned regions were identified; in three of the alignments the sequences of both KpsE and KpsD were, according to the topology models of KpsE and KpsD presented in Figures 3.1 and 5.2, located in the periplasm. The identified regions were between 13 and 16 residues long and contained 31%-43% identical and 38%-63% similar amino acids. The first region contains a TXGFX_polarPX_acidic-X_aromaticXLXXXXX_polar-X_aliphatic motif and is located at residues 157-172 in the KpsE periplasmic region and at residues 84-99 of KpsD. The second region comprises a QAQXXXXVXXXX_aliphatic motif, located at residues 228-240 in the KpsE periplasmic region and at residues 164-175 of KpsD. According to the topology model of KpsD presented in Chapter 5 (Figure 5.2), region 1 and 2 are likely to be surface-exposed and buried into the outer membrane, respectively. These regions are thus unlikely to directly interact with KpsE. The third region carries the previously mentioned QGDXXNXXXX_acidicFXE motif (cf. section 3.6.1); this motif is formed by residues 289-302 in the KpsE periplasmic region and by residues 255-268 in the KpsD periplasmic region. The third motif exhibits the highest identity and similarity between KpsE-KpsD residues. To fully elucidate the points of interaction between KpsE and KpsD in the future, these residues could form the starting point of a mutagenesis approach in KpsD using co-purification of mutant variants of KpsD with native hisKpsE.
3.6.4 The impaired function of KpsE\textsubscript{ΔCterm} is due to disruption of the KpsE-KpsM interaction

C-terminal deletion mutagenesis of KpsE resulted in a non-functional KpsE variant, demonstrating that the C-terminal residues are essential for correct function of KpsE as member of the K5 translocon. Analysis showed that the observed loss of function was not due to sub-optimal protein expression or disrupted inner membrane insertion. Loss of function could have also been caused by misfolded KpsE, resulting in disruption of intra- or inter protein interactions due to KpsE misfolding (including disruption of KpsE oligomerisation) rather than loss of KpsE-KpsM interaction between associated residues. However, this is less likely due to the following observations, which suggest that the truncated KpsE variant did not express impaired folding. A) KpsE\textsubscript{ΔCterm} retained its native inner membrane location (section 3.3.2). B) The C-terminal deletion did not affect native KpsD outer membrane location and dimerisation (section 3.4.1), even though the importance of KpsE for KpsD membrane localisation and oligomerisation had been demonstrated (Hudson, 2009). This strongly suggests that KpsE\textsubscript{ΔCterm} folding is sufficient to support the interaction between KpsE and KpsD which is required for native KpsD expression. The observed disrupted function of KpsE and the loss of KpsE-hisKpsM co-purification were thus interpreted as genuinely resulting from the disruption of interacting residues from KpsE and KpsM. To further confirm correct KpsE folding, cysteine-scanning labelling could be utilised to analyse whether the residues located at the new C-terminus of truncated KpsE are still located in the periplasm (cf. Chapter 5 for a discussion about a cysteine-scanning labelling approach in KpsD). Alternatively, trypsin or proteinase K digestion analysis could be used to confirm KpsE folding.

Wild-type KpsE was shown to be co-purified by hisKpsM (Hudson, 2009). The results presented here demonstrated that after the removal of the C-terminal \(\alpha\)-helix of KpsE the ability of hisKpsM to co-purify KpsE was abolished, as KpsE\textsubscript{ΔCterm} was not detected after the co-purification procedure. The absence of observable hisKpsM after co-purification suggests that the interaction between KpsE and KpsM is disrupted in such a way that it affects the ability of the his-tag on KpsM to bind to the nickel column. It is likely that this loss of, or decrease in binding affinity with the column was caused by conformational changes in KpsM due to the absence of the KpsE C-terminus. These conformational changes might have resulted in the his-tag becoming
inaccessible to binding, or might have affected the ability of KpsM to self-associate, resulting in the breakdown of oligomeric KpsM into monomers and availability of only a single his-tag rather than multiple ones for nickel binding. As hisKpsM was detected in the inner membrane prior to co-purification, it is suggested that the stability of hisKpsM is not significantly affected by the absence of the C-terminal helix of KpsE. In conclusion, the data presented here imply that the C-terminus of KpsE is a site of interaction between KpsE and KpsM. This result strengthens the hypothesis KpsE and KpsM directly interact and that this interaction is likely to have an important role in the complex formation of the K5 translocon.

Future work should be directed towards identifying the region(s) of KpsM which are responsible for interacting with the C-terminal domain of KpsE. Previous site-directed mutagenesis of the conserved residues 66, 147, 190 and 213 of KpsM was unable to identify residues essential for KpsM function (Hudson, 2009). The three periplasmic loops as well as the six transmembrane domains of KpsM are potential sites of interaction. A so-called linker region which is located between the fifth and sixth transmembrane domain and is predicted to be only partly buried in the membrane, is a promising candidate for interaction with the C-terminal of KpsE. Site-directed mutagenesis or removal of this region combined with co-purification of mutated KpsM with wild-type hisKpsE could reveal the significance of the linker region in KpsE-KpsM interaction.

Due to the suggested nature of a direct interaction and due to the nature of the proposed K5 translocon complex, it is very likely that KpsE and KpsM interact via further regions in addition to the KpsE C-terminus. To investigate further potential points of interaction, a BLAST alignment of KpsE and KpsM was carried out to identify any homologous regions. The alignment identified a single region, a $X_{\text{acidic}}X_{\text{aromatic}}X_{\text{basic}}E$ motif, located at residues 87-93 of KpsE and at residues 14-20 of KpsM. The sequences of the two proteins share 57% identity and 100% similarity. According to the KpsE model presented in Figure 3.1, this motif is located in the periplasm within the second $\alpha$-helix after the transmembrane domain of KpsE. In KpsM the motif is located at the N-terminal domain and the current topology model of KpsM (Pigeon and Silver, 1994) positions the motif in the cytoplasm. An interaction of KpsE and KpsM via this motif is therefore unlikely. Further mutagenesis work on
larger regions of the periplasmic loop of KpsE combined with a co-purification approach involving hisKpsM should prove useful in identifying further points of interaction of KpsE with KpsM.
Chapter 4

Strategies to purify KpsE and KpsM as higher order complex
4.1 Introduction

Obtaining structural information of the translocon members would greatly enhance our understanding of the group 2 polysaccharide translocation process. As previously discussed (section 1.5.3.4) and reiterated by the results in Chapter 3, protein-protein interactions appear to be crucial for the correct conformation and stability of some members of the K5 translocon. Therefore, obtaining structural information of an isolated translocation protein might either be impossible (e.g., in the case of KpsM, which requires KpsTE for stable expression) or might result in protein conformations not relevant to the export process (e.g., this might be the case for KpsE). For these reasons, the effort of studies on KpsEDMT should ideally be focused towards obtaining the relevant protein in a complex with its translocon partners when aiming to isolate the proteins for structural analysis such as electron microscopy or crystallography.

This chapter focuses on the putative KpsE-KpsM complex. The co-purification study in Chapter 3 (section 3.5.3) confirmed that KpsE and hisKpsM are likely to interact. An earlier study which attempted to obtain a stable form of the KpsE-KpsM complex by nickel-affinity chromatography found that the KpsE-KpsM complex dissociated quickly and could not be used for further structural studies (Hudson, 2009). This lack of stability was attributed to the high concentrations of imidazole used to elude the complex. This chapter investigates different strategies for their feasibility to obtain a stable KpsE-KpsM complex.

4.2 Nickel-affinity chromatography and proteolytic cleavage of KpsE-hisKpsM

To address the possible negative effect of imidazole on the stability of the KpsE-KpsM complex, it was investigated whether nickel-affinity chromatography followed by Tobacco etch virus (TEV) protease cleavage of the his-tag would be a feasible alternative approach to using high concentrations of imidazole to elude the complex. This approach requires construction of a vector which carries a) the target gene(s), b) a his-tag sequence and c) a TEV protease recognition sequence located between the target gene(s) and the tag sequence (cf. Figures 4.1 and 4.3). The fusion protein(s) are bound to the nickel-column by the his-tag as in a standard affinity chromatography
protocol (Figure 4.1). Subsequent incubation of the bound fusion protein with TEV protease results in cleavage at the TEV protease recognition site, releasing the fusion protein(s) from the his-tag (Figure 4.1). Besides eliminating the need for high concentrations of imidazole, removal of his-tags by TEV protease cleavage has the additional advantage of potentially obtaining a protein complex which is closer to the native conformation than its his-tagged variant. Removal of the his-tag from a single protein for the latter purpose has been successfully applied prior to crystallography of Wzb of *E. coli* K30 (Liu and Naismith, 2009) and the ferric uptake regulator Fur of *Vibrio cholerae* (Sheikh and Taylor, 2009). Digestion of TEV protease in these studies was applied after affinity chromatography. In this part of the chapter, a protocol was developed which aimed to remove the his-tag from the KpsE-KpsM complex by on-column protease incubation. This approach aimed to test whether the KpsE-KpsM complex is still formed in the absence of a his-tag and whether a more stable complex can be obtained in the absence of imidazole.

### 4.2.1 Generation of pBAD28 construct carrying kpsE, hiskpsM and TEV cleavage site

For the purpose of purification, a construct carrying *kpsE*, *hiskpsMT* and a TEV cleavage recognition site was created. TEV protease recognises the seven amino acid consensus sequence EXXYXQ▼S/G (where X denotes any amino acid and ▼denotes the cleavage site) (Dougherty et al., 1989). Here, the most commonly used recognition sequence, ENLYFQG, was incorporated into the construct. The fragment *hisTEVkpsMT* was amplified by PCR (Figure 4.2, lane 1) from a previous version of *hisTEVkpsMT* (which contained unsuitable restriction sites) using primers hisTEVKpsMT XbaI F and KpsT HindIII R (section 2.5.2; Table 2.3). The fragment was cloned into pBADE, generating the construct pBADEhisTEVMT (Figure 4.3). pBADEhisTEVMT was transformed into DH5α and positive clones were identified by colony PCR (Figure 4.2, lane 2). The construct was verified by restriction digest analysis (Figure 4.2, lanes 3-5) and DNA sequencing. Correct protein function was assessed using the K5 bacteriophage assay. pBADEhisTEVMT was able to complement MSΔEΔMT (Table 4.1), confirming correct protein function. Interestingly, the hisTEV variant produced a lower number of plaques (Table 4.1) and plaques were less distinct (data not shown) than the his-variant or wild-type. This implies that the addition of the TEV recognition site to the vector might have had a slight altering effect on protein
Figure 4.1 Concept of on-column proteolytic cleavage following nickel-affinity chromatography. 1: Suitable preparations from a strain expressing hisTEV-tagged protein and associated proteins (such as membranes) are solubilised using a suitable detergent and are loaded onto a his-trap column containing a sepharose base matrix charged with nickel ions; 2: the his-tag binds to the nickel ions, retaining the hisTEV-tagged protein and associated, interacting proteins; 3: the column is incubated with TEV protease which cleaves at the TEV protease recognition site to separate the formerly hisTEV-tagged protein and associated proteins from the his-tag; 4: washing the column releases the recombinant protein complex; 5: washing the column with a high concentration of imidazole releases any uncleaved hisTEV-tagged proteins and associated proteins as well as TEV protease. Ni$^{2+}$: nickel ions; 6xhis: hexahistidine (his) tag; TEV: TEV protease recognition site ENLYFQG.
Figure 4.2 Generation of pBADEhisTEVMT construct. The plasmid pBADEhisTEVMT was generated as described in sections 2.5.2 and 2.5.4-2.5.10. Samples were separated on a 1% agarose gel. Lane 1: PCR amplification of hisTEVkpsMT, generating a 1.5 kb product; lane 2: colony PCR of pBADEhisTEVMT using pBAD28 F and R primers, generating a 2.65 kb product; lane 3: pBADEhisTEVMT digested with Sacl and Xmal, generating a 1.15 kb product containing kpsE and a 7.3 kb vector product containing the pBAD28 backbone and hisTEVkpsMT; lane 4: pBADEhisTEVMT digested with Xbal and HindIII, generating a 1.5 kb product containing hisTEVkpsMT and a 6.95 kb vector fragment containing the pBAD28 backbone and kpsE; lane 5: pBADEhisTEVMT digested with Sacl and HindIII, generating a 2.6 kb product containing kpsE and hisTEVkpsMT as well as a 5.8 kb fragment containing the pBAD28 vector backbone.
**Figure 4.3 Illustration of the plasmid pBADEhisTEVMT.** The triangle within the TEV recognition sequence denotes the TEV protease cleavage site. his: 6X histidine-tag sequence; T: TEV recognition sequence; MCS: multiple cloning site; p1: arabinose-inducible pBAD promoter; p2: AmpR promoter; term: transcriptional terminator rrnB T1; araC: arabinose transcriptional regulator; f1 ori and p15A ori: f1 and p15A origin of replication, respectively; AmpR and CmR: ampicillin and chloramphenicol resistant marker, respectively; NsiI: NsiI restriction site.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plaque count at different K5 phage dilutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10(^{-6})</td>
</tr>
<tr>
<td>MS101</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔEΔMT</td>
<td>0</td>
</tr>
<tr>
<td>MSΔEΔMT (pBADEhisMT)</td>
<td>53</td>
</tr>
<tr>
<td>MSΔEΔMT (pBADEhisTEVMT)</td>
<td>39</td>
</tr>
</tbody>
</table>

**Table 4.1** K5 bacteriophage assays of KpsEhisTEVKpsMT construct and controls. TMTC: too many plaques to count.
function, even though K5 export still occurred. It was therefore critical to analyse KpsE and hisTEVKpsM expression and membrane localisation.

4.2.2 Expression analysis of KpsE and hisKpsM from DH5α (pBADEhisTEVMT)

For the purpose of co-purification analysis, pBADEhisTEVMT was expressed in the capsule-deficient DH5α. In order to identify the condition which gives the best expression levels, cells were induced at OD_{600}: 0.4-0.5 (mid-log) with either 0.1% arabinose for 3 hours (section 2.7.1) or 0.1% arabinose for 3 hours after growing the cells twice to mid-log density (section 2.7.2). Cells were analysed for KpsE and KpsM expression by Western blot using anti-KpsE and anti-his antibodies (Figure 4.4, lanes 1, 2 and 4). Expression of KpsE as well as hisTEVKpsM was well induced using the standard induction method. Growing the cells to higher densities did not increase expression of KpsE. It was therefore decided not to probe the high cell density cells with anti-KpsM, and to use the standard induction protocol for further studies. To verify expression levels and localisation in the membrane, total membranes were isolated from induced whole cells of DH5α (pBADEhisTEVMT) and probed with anti-KpsE and anti-his antibodies (Figure 4.4, lane 3). Both KpsE and hisTEVKpsM were detected in the total membrane. Co-purification analysis was carried out from total membranes to avoid loss of these strong KpsE and his signals which would occur by isolating inner membranes.

4.2.3 Co-purification of KpsE and KpsM using TEV protease cleavage

It had previously been shown that, unlike some other detergents, DM does not inhibit the enzymatic activity of TEV protease (Mohanty et al., 2003) and it was thus deemed an appropriate choice for use here. As preparation for co-purification analysis, total membranes were solubilised with 0.2% DM as carried out for conventional his-tag purification in Chapter 3 (cf. sections 2.7.8 and 3.5.3). The following protocol for purification and TEV cleavage was developed (section 2.7.9). Solubilised membranes were passed through a conventional his-trap nickel column and the flowthrough was collected. This was followed by a wash containing 20 mM imidazole. For on-column cleavage of his-tagged proteins by TEV protease, his-tagged TEV protease was loaded onto the column. Incubation for 16 hours at 4°C was found to provide best cleavage conditions (data not shown). A protease-substrate ratio of 1:100 was found to produce
the best purification results (data not shown). The column was washed with protease cleavage buffer and proteins released as a result of protease digestion were collected. Any remaining column-bound his-tagged proteins were eluted off the column by a high concentration of imidazole (designated elution fractions). All fractions were analysed by SDS-PAGE and Western blot using anti-KpsE and anti-his antibodies (Figure 4.5).

A large quantity of protein was detected by SDS-PAGE in the flowthrough and the 20 mM imidazole wash. Western blot analysis revealed that the general protein content contained some KpsE but little hisTEVKpsM. After on-column incubation with TEV protease, two faint bands at the sizes of KpsE and KpsM were observed on the PAGE gel. Western blot analysis confirmed that the larger band was KpsE. The lower band was predicted to be KpsM, which cannot be detected by anti-his antibody due to on-column removal of the his-tag. Bands corresponding to the size of KpsM were also detected in elution fractions 1-3 by SDS-PAGE. The anti-KpsE Western blot showed that after elution with imidazole, KpsE was mainly present in elution fraction 2. The PAGE gel and anti-his Western blot showed that the his-tagged TEV protease was present in all elution fractions. Non-cleaved hisKpsM was not detected in elution fraction 1. Due to the strong his-signal from the protease, it was unclear whether any non-cleaved his-tagged KpsM was present in elution fractions 2 and 3. Taken together, these results show that KpsE and KpsM are present in the same fraction after cleavage of the his-tag by TEV protease. This supports the evidence from conventional nickel affinity chromatography that KpsE and KpsM interact in vivo in the absence of other capsule gene products and polysaccharide substrate. However, as the protein yield of the putative complex was very low, the proteins were not used for further studies. Increasing protein expression in order to obtain higher capsule protein yields was addressed at a later stage and will be discussed in section 4.4.

4.3 Size exclusion chromatography of KpsE-KpsM

Section 4.2 has further supported the evidence for a KpsE-KpsM complex. However, it also became apparent that various factors, such as protein yield and protease cleavage efficiency can affect the efficiency of the nickel affinity-protease cleavage purification approach. When considering alternative approaches, it was decided to investigate the feasibility of size exclusion chromatography to isolate a stable KpsE-KpsM complex.
Figure 4.4 Expression analysis of KpsE and hisKpsM from DH5α (pBADEhisTEVMT). Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row) of DH5α (pBADEhisTEVMT). Proteins were separated on a 12% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical densities of cells. Lane 1: uninduced whole cells, lane 2: whole cells induced with 0.1% arabinose for 3h, lane 3: total membranes induced with 0.1% arabinose for 3h, lane 4: whole cells prepared according to high cell density method, induced with 0.1% arabinose for 3 hours.

Figure 4.5 Co-purification of KpsE and KpsM using TEV protease cleavage from total membranes of DH5α (pBADEhisTEVMT). (A) SDS-PAGE analysis (12% gel). Lane 1: flowthrough; lane 2: wash with 20 mM imidazole; lane 3: on-column incubation with TEV protease; lanes 4-8: 1 ml fractions of 250 mM imidazole wash (elution fractions 1-5). (B) Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row). Lane 1: flowthrough; lane 2: wash with 20 mM imidazole; lane 3: on-column incubation with TEV protease; lanes 4-6: 1 ml fractions of 250 mM imidazole wash (elution fractions 1-3).
Size exclusion chromatography potentially allows complex isolation and simultaneously provides information about the molecular weight distribution of single proteins and protein complexes. This additional advantage of the approach lead to a second objective: the analysis of the oligomeric state of KpsE in the absence of other capsular gene products and the K5 polysaccharide and analysis of the effect of co-expression of KpsM(T) on the oligomerisation state of KpsE. Previous work suggested that KpsE can form dimers in the presence of chemical cross-linkers (Arrecubieta et al., 2001). However, it remains unclear whether a dimer is the in vivo state of KpsE. Using size exclusion chromatography to analyse the oligomerisation state of KpsE in the absence of cross-linkers might therefore reveal interesting information.

4.3.1 Size exclusion analysis of KpsE

To allow comparison between chromatography profiles of KpsE in the presence and absence of hisKpsM(T), first size exclusion chromatography of KpsE was carried out and results were then compared to results from chromatography of KpsEM(T). For the purpose of size exclusion chromatography of KpsE, total membranes of DH5α (pBADE) were prepared (section 2.7.5) and solubilised using 0.2% DM. Size exclusion chromatography was carried out using a Superdex 200 column as described in section 2.7.10. The Superdex 200 column was chosen for its protein separation range of 10 – 600 kDa, allowing appropriate separation of a variety of potential translocon complex formations. The chromatogram of the elution profile displaying protein concentrations measured in absorbance at 280 nm was recorded (Figure 4.6). The molecular weights of the observed shoulders and peaks were determined as described in section 2.7.10. The elution profile showed two main shoulders, corresponding to molecular weights of 87 kDa and 186 kDa, respectively (cf. Table 4.2). To confirm KpsE was present in these fractions, every second fraction within the elution volume range of 8 to 17 ml was analysed by Western blot using anti-KpsE antibody. KpsE was detected in all fractions from 8 to 14 ml but not in fractions from 15 to 17 ml. The strongest signal of KpsE was detected in the 12 ml fraction. Western blot analysis of selected fractions is shown in Figure 4.7 (top row). Thus, Western blot analysis confirmed that the 87 kDa and 186 kDa peaks contain KpsE. Taken together, these data suggest that KpsE
can form dimers and tetramers *in vivo* in the absence of other capsule gene products and the polysaccharide substrate.

### 4.3.2 Size exclusion analysis of KpsE-KpsM(T)

Total membranes from DH5α (pBADEhisMT) were prepared (section 2.7.5) to establish the suitability of size exclusion chromatography for the isolation of a stable KpsE-KpsM complex as well as to analyse KpsE oligomerisation in the presence of KpsM(T) and the oligomerisation state of KpsE-KpsM complex. Western blot analysis in Chapter 3 (Figure 3.11) had already confirmed the presence of KpsE and hisKpsM in total membranes. Size exclusion chromatography was carried out as described in section 4.3.1. The elution profile from DH5α (pBADEhisMT) showed three main shoulders and one distinct peak, corresponding to approximate molecular weights of 480 kDa, 295 kDa, 205 kDa and 73 kDa, respectively (Figure 4.6).

Western blot analysis detected KpsE in the same fractions as detected in the fractions from size exclusion chromatography of KpsE only in section 4.3.1 (Figure 4.7, top row). As in the results for DH5α (pBADE), the signal of KpsE from DH5α (pBADEhisMT) was strongest in the 12 ml fraction. Western blot analysis to detect hisKpsM was carried out on the 8, 10, 11, 12 and 14 ml fractions (Figure 4.7, bottom row). A signal at the size of monomeric KpsM (29.5 kDa) could not be detected. However, a signal just below the 50 kDa molecular weight standard was detected in analysed fractions. This signal was absent in the elution fractions from DH5α (pBADE), strongly suggesting that this signal is specific to the presence of hisKpsMT. This hisKpsMT-specific signal was strongest in the 10 ml fraction. Comparison of SDS-PAGE analysis of 8-14 ml elution fractions from DH5α (pBADE) and DH5α (pBADEhisMT) did not reveal obvious bands which could be associated with either hisKpsM or the hisKpsM-specific signal (Figure 4.8). The identity of the hisKpsM signal remains unclear and will be further discussed in section 4.5.4. However, as the signal could only be detected in strains carrying hisKpsMT, and protein expression, membrane preparation and chromatography conditions of DH5α (pBADE) and DH5α (pBADEhisMT) were identical, this signal was interpreted as evidence for the presence of hisKpsM. Assuming the hisKpsM-specific signal corresponded to the presence of hisKpsM, the 8, 10, 11, 12, and 14 ml fractions contained both KpsE and hisKpsM.
Figure 4.6 Chromatogram of size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). Broken trace: DH5α (pBADE); trace line: DH5α (pBADEhisMT). Arrows indicate elution volumes of calibration standards. Note that monomeric KpsE has a size of 43 kDa (Pazzani et al., 1993; Rosenow et al., 1995a), monomeric KpsM a size of 29.5 kDa and monomeric KpsT a size of 25 kDa (Pigeon and Silver, 1994). Elution fractions analysed by Western blot in Figure 4.7 are highlighted by black boxes.

Figure 4.7 Analysis of KpsE and KpsM in elution fractions from size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). Western blot analyses of elution fractions using anti-KpsE (top row) and anti-his antibodies (bottom row). Proteins were separated on a 12% SDS-PAGE gel prior to electrophoretic transfer. Lanes designated EMT are figures of DH5α (pBADEhisMT), while lanes designated E are derived from DH5α (pBADE). Elution fractions given in ml refer to the fractions highlighted in Figure 4.6. Lane 1: 8 ml elution fraction; lane 2: 10 ml elution fraction; lane 3: 11 ml elution fraction; lane 4: 12 ml elution fraction; lane 5: 14 ml elution fraction.
Table 4.2 Predicted homo- and hetero-oligomeric complexes resulting from size exclusion chromatography of DH5α (pBADE) and DH5α (pBADEhisMT). The table illustrates the various different possible complex formations of KpsE and KpsE-KpsMT resulting from size exclusion chromatography. The calculated molecular weights of the major peaks on the chromatogram and the predicted molecular weight of possible complexes are given in kDa.
These fractions approximately correspond to the 480 kDa, 295 kDa and 205 kDa shoulders, and the 73 kDa peak. The presence of the two proteins in the same elution fraction strongly suggests that the two proteins interact and thus confirms previous findings. This suggests that size exclusion chromatography can be used to isolate the KpsE-KpsM complex. However, further investigations into the possible reasons for the absence of a monomeric hisKpsM signal and the identity of the observed hisKpsM-specific signal are required first.

Even though KpsT is a membrane-associated protein rather than an integral inner membrane protein, it is predicted that KpsT integrates into the membrane during K5 polysaccharide export when interacting with KpsM (Bliss and Silver, 1997; Nsahlai and Silver, 2003) and KpsT has previously been detected in total membranes in the presence and absence of the polysaccharide substrate (Pavelka et al., 1994). In order to investigate the potential role of KpsT in KpsE-KpsM complex formation, whole cells and elution fractions from DH5α (pBADEhisMT) were analysed with anti-KpsT antibody (Figure 4.9). Expression of KpsT was induced after induction with arabinose in whole cells. KpsT could be detected in the 10 ml and 10.5 ml elution fractions but not in the 11.5 ml or 14 ml fractions. The KpsT signal was strongest in the 10 ml fraction. This suggests that KpsT, KpsM and KpsE are all present in the elution fraction which corresponds to a size of 480 kDa, suggesting the three proteins might form a higher order complex. These data also suggest that KpsT interacts with KpsM in this elution fraction but not in elution fractions where KpsM and KpsE appear to be present as smaller complexes.

Table 4.2 summarises the different possibilities of KpsE-KpsMT complex formations in the various fractions. The 73 kDa peak is predicted to contain monomeric KpsE and monomeric hisKpsM. The 205 kDa shoulder would most closely correspond to a complex of dimeric KpsE and tetrameric KpsM. The 295 kDa shoulder most closely corresponds to a complex of tetrameric KpsE and tetrameric KpsM, while the 480 kDa shoulder most closely corresponds to a complex of hexameric KpsE, tetrameric KpsM and tetrameric KpsT. This could indicate that KpsM and KpsT interact as a tetramer with a KpsE which can adopt different oligomeric states. Analysis of the chromatogram from DH5α (pBADE) supports the notion that KpsE can change its oligomeric organisation. Interestingly, the strongest KpsT and hisKpsM-specific signals were
Figure 4.8 Analysis of total protein from elution fractions from size exclusion chromatography of total membranes of DH5α (pBADE) and DH5α (pBADEhisMT). SDS-PAGE analysis (12% gel). Lanes designated EMT are figures of DH5α (pBADEhisMT), while lanes designated E are derived from DH5α (pBADE). Lane 1: 10 ml elution fraction; lane 2: 11 ml elution fraction; lane 3: 12 ml elution fraction; lane 4: 14 ml elution fraction.

Figure 4.9 Analysis of KpsT in whole cells and elution fractions from size exclusion chromatography of total membranes of DH5α (pBADEhisMT). Western blot analysis of elution fractions using anti-KpsT antibody. Proteins were separated on an 12% SDS-PAGE gel prior to electrophoretic transfer. Lane 1: uninduced whole cells; lane 2: whole cells induced with 0.1% arabinose for 3 hours; lane 3: 10 ml elution fraction; lane 4: 10.5 ml elution fraction; lane 5: 11.5 ml elution fraction; lane 6: 14 ml elution fraction.
detected in the fraction which corresponds to 480 kDa, the largest complex identified by size exclusion chromatography. Taken together, the data presented here suggest that KpsE and hisKpsM can form higher order complexes larger than a mere monomer-monomer complex. However, the limitations of size exclusion chromatography do not allow specific deductions for the organisation of the complex.

4.4 Optimisation of KpsE and KpsM expression

In section 4.2 it became apparent that the optimisation of expression of KpsE and KpsM is required in order to obtain a higher yield of the protein complex from the TEV cleavage approach, and increased protein expression might also lead to more unequivocal results from size exclusion chromatography (section 4.3).

4.4.1 pBAD28 expression trial

As a first step towards improving protein yield, an expression trial was carried out with the strain DH5α (pBADDehisMT), which was used during nickel-affinity chromatography in Chapter 3 (section 3.5.3). The pBAD28 vector had been chosen as the use of titratable concentrations of arabinose provides accurate control over protein expression levels. Previous protein expression in cells or membranes used in experiments in Chapter 3 and Chapter 4 had been carried out by inducing expression at OD$_{600}$: 0.4-0.5 with 0.1% arabinose for 3 hours. To investigate whether an increase in arabinose concentration during induction or an elongated induction time would increase KpsE and KpsM expression levels, whole cells were induced at OD$_{600}$: 0.4-0.5 using 0.1% and 0.5% arabinose for either 3 or 16 hours. Whole cells were analysed by Western blot using anti-KpsE and anti-his antibodies (Figure 4.10 A). Compared to expression of KpsE and hisKpsM from cells induced with 0.1% arabinose for 3 hours, increasing induction time or arabinose concentration appeared not to enhance KpsE or hisKpsM expression. In order to examine whether changing the time point of induction and the consequent further elongation in induction would affect protein expression, whole cells were induced with 0.1% arabinose for 16 hours either at OD$_{600}$: 0.2 or directly after subculturing. Whole cells were analysed by Western blot using anti-KpsE and anti-his antibodies (Figure 4.10 B). Inducing cells under these conditions resulted in
Figure 4.10 Expression trial of DH5α (pBADEhisMT) using various induction conditions.

(A) Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row). Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical density of cells. Cells were induced at OD$_{600}$: 0.4-0.5. Lane 1: uninduced whole cells; lane 2: whole cells induced with 0.1% arabinose for 3 hours; lane 3: whole cells induced with 0.1% arabinose for 16 hours; lane 4: whole cells induced with 0.5% arabinose for 3 hours; lane 5: whole cells induced with 0.5% arabinose for 16 hours.

(B) Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row). Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical density of cells. Lane 1: uninduced whole cells; lane 2: whole cells induced at OD$_{600}$: 0.4-0.5 with 0.1% arabinose for 3 hours; lane 3: whole cells induced at OD$_{600}$: 0.2 with 0.1% arabinose for 16 hours; lane 4: whole cells induced at stage of subculturing with 0.1% arabinose for 16 hours.
a slight increase in KpsE signal and a strong increase in his signal, compared to cells induced at OD<sub>600</sub> 0.4-0.5 for 3 hours. To test whether this observed increase in signal intensity would translate into a significantly increased protein yield after KpsE-hisKpsM co-purification, whole cells induced with 0.1% arabinose directly after subculturing for 16 hours were used to prepare inner membranes and nickel-affinity chromatography was carried out. However, despite an increase in protein expression in whole cells, no significant difference in protein yield after affinity chromatography was observed (data not shown).

4.4.2 Generation of pET-24(+) construct carrying kpsE and hiskpsMT

The previous section (section 4.4.1) showed that increasing expression levels using the DH5α (pBAD28) host-vector system did not provide sufficient improvement in KpsE and hisKpsM expression levels which can be translated into high yields after co-purification. Therefore, it was decided to investigate whether a different host-vector system would resolve this issue.

pET vectors are T7 promoter-driven systems and can provide high levels of protein expression. The vector pET-24(+) was chosen for its suitable restriction sites and the option of using its C-terminal his-tag sequence. The <i>E. coli</i> BL21-AI™ expression strain is ideal for recombinant protein expression from a pET system as it carries a chromosomal copy of the T7 RNA polymerase gene, which is under the control of a promoter which can be induced by IPTG or auto-induction medium. It should be noted that BL21 (DE3) – from which BL21-AI™ is derived – carries homologues of region 1 and 3 of the group 2 capsular gene cluster on its chromosome (Andreishcheva and Vann, 2006a). Even though these region 1 and 3 homologues have shown to be functional in polysaccharide export if provided with a substrate, BL21 (DE3) does not express a polysaccharide capsule due to an insertion sequence element in region 2 (Andreishcheva and Vann, 2006a). As expression from the engineered pET construct would not result in polysaccharide synthesis in BL21-AI™, the chromosomal presence of part of the capsular gene cluster was deemed a negligible issue.

For the purpose of a KpsE-KpsM expression trial, a pET-24(+) construct carrying KpsE and N-terminally his-tagged KpsM(T) was generated. The fragments <i>kpsE</i> and
hiskpsMT were amplified by PCR (Figure 4.11, lane 1) using primers KpsE SacI F and KpsE R (which carries an XmaI restriction site), and hisKpsMT XbaI F and KpsT HindIII R, respectively (section 2.5.2; Table 2.3). The fragments were cloned into pET-24(+), generating construct pETEhisMT (Figure 4.12). The construct was then transformed into DH5α (Figure 4.11, lane 2) and verified by restriction digest (Figure 4.11, lanes 3-5) and DNA sequencing.

4.4.3 pET-24(+) induction trial

For the purpose of protein expression, the pETEhisMT construct was transformed into BL21-AI™. Five positive clones were obtained and screened for the best KpsE and KpsM expression by analysing induced whole cells by Western blot (data not shown). The clone which produced the best protein expression was selected for subsequent experiments. Whole cells of BL21-AI™ (pETEhisMT) were induced at OD₆₀₀: 0.4-0.5 with 1 mM IPTG and 0.2% arabinose for 3 hours. Protein expression was subsequently analysed by Western blot using anti-KpsE and anti-his antibodies (Figure 4.13, lanes 1-2). Both KpsE and hisKpsM were detected in induced whole cells. KpsE was also found in uninduced cells, suggesting some basal expression in the absence of the IPTG inducer. Induced whole cells were then used to prepare total and inner membranes, and these were analysed for correct KpsE and hisKpsM localisation by Western blot (Figure 4.13, lanes 3-4). The analysis confirmed the localisation of KpsE and hisKpsM to the inner membrane, and both proteins were readily detected. Subsequent nickel-affinity chromatography from total membranes of BL21-AI™ (pETEhisMT) provided good anti-KpsE signals of all chromatography steps (Figure 4.14). However, the anti-his signal was not strong enough to be detected (data not shown).

4.4.4 Autoinduction from pET-24(+) and comparison of KpsEMT expression from DH5α (pBAD28) and BL21-AI™ (pET-24(+)) systems

In order to further enhance protein expression from BL21-AI™ (pETEhisMT), it was tested whether changing induction conditions would increase protein expression to detectable levels of hisKpsM during nickel affinity purification. Instead of inducing with IPTG, whole cells were induced according to the autoinduction method (section 2.7.3), which can provide much higher cell densities and protein yields than standard
Figure 4.11 Generation of plasmid pETEhisMT. The plasmid pETEhisMT was generated as described in sections 2.5.2 and 2.5.4-2.5.10. Samples were separated on a 1% agarose gel. Lane 1: PCR amplification of kpsE, generating a 1.15 kb product; lane 2: colony PCR of pETEhisMT using pET-24(+) T7 promoter and terminator primers, generating a 2.65 kb product; lane 3: pETEhisMT digested with SacI and Xmal, generating a 1.15 kb product containing kpsE and a 6.7 kb vector product containing the pBAD28 backbone and hiskpsMT; lane 4: pETEhisMT digested with Xbal and HindIII, generating a 1.5 kb product containing hiskpsMT and a 6.35 kb vector product containing the pBAD28 backbone and kpsE; lane 5: pETEhisMT digested with SacI and HindIII, generating a 2.6 kb product containing kpsE and hiskpsMT as well as a 5.25 kb product containing the pBAD28 vector backbone.

Figure 4.12 Illustration of the plasmid pETEhisMT. his: 6X histidine-tag sequence; MCS: multiple cloning site; p: IPTG-inducible T7 promoter; lacO: Lac operator; term: T7 transcriptional terminator; lacI: transcriptional regulator LacI; f1 ori and pBR 322 ori: f1 and pBR 322 origin of replication, respectively; KanR: kanamycin resistant marker.
Figure 4.13 Expression and membrane localisation of KpsE and hisKpsM from BL21-Ai™ (pETEhisMT). Western blot analyses using anti-KpsE (top row) and anti-his antibodies (bottom row). Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Loadings of whole cells were standardised according to optical density. Lane 1: uninduced whole cells; lane 2: whole cells induced with 1 mM IPTG and 0.2% arabinose for 3 hours; lane 3: induced total membranes; lane 4: induced inner membrane.

Figure 4.14 Nickel-affinity chromatography of total membranes from BL21-Ai™ (pETEhisMT). Western blot analysis using anti-KpsE. Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Lane 1: flowthrough; lane 2: 10 mM imidazole wash; lane 3: 50 mM imidazole wash; lane 4: elution fraction 1. hisKpsM visualisation is not shown as no anti-his signal could be detected.
IPTG induction (Studier, 2005). Autoinduced whole cells of BL21-AI™ (pETEhisMT) were analysed by Western blot using anti-KpsE and anti-his antibodies. Compared to the uninduced control, the anti-KpsE signal from autoinduced cells was increased. However, compared to IPTG-induced cells, the signal was not enhanced (Figure 4.15, lanes 6-8, top row). The anti-his signal was slightly stronger in autoinduced cells than in the uninduced control, but signal was decreased compared to IPTG-induced cells (Figure 4.15, lanes 6-8, middle row). It was concluded that the autoinduction method had induced KpsE expression well but not to higher levels than the IPTG-induction. The expression of hisKpsM had not been well induced and autoinduction produced lower levels of hisKpsM than IPTG-induction. This suggests that autoinduction of BL21-AI™ (pETEhisMT) is not advantageous with respect to KpsE and hisKpsM expression levels.

To be able to directly compare protein expression levels from the various vectors, hosts, and induction conditions analysed so far, whole cells of BL21-AI™ (pETEhisMT), DH5α (pBADEhisMT) and DH5α (pBADEhisTEVMT) induced according to previously described induction methods (sections 4.2.2, 4.4.1, 4.4.3 and 4.4.4) were analysed by Western blot alongside each other (Figure 4.15). KpsE expression levels were constant throughout, regardless of type of vector, host cell or induction conditions (Figure 4.15, top row). Expression of hisKpsM was slightly increased in the IPTG-induced BL21-AI™ (pET-24(+)) system compared to DH5α (pBAD28); but KpsM expression from the autoinduced BL21-AI™ (pET-24(+)) was drastically decreased compared to both IPTG-induced BL21-AI™ (pET-24(+)) and DH5α (pBAD28) (Figure 4.15, middle row). KpsT expression did not differ significantly between DH5α (pBAD28) and IPTG-induced BL21-AI™ (pET-24(+)) (Figure 4.15, bottom row). As autoinduction decreased KpsM expression, expression of KpsT was not tested in autoinduced cells, as KpsT expression levels are predicted to be similar to KpsM levels.
Figure 4.15 Comparison of expression profiles of KpsE, hisKpsM and KpsT from different vectors and hosts using various induction conditions. Western blot analyses of whole cells using anti-KpsE (top row), anti-his (middle row) and anti-KpsT (bottom row) antibodies. Proteins were separated on a 15% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical densities of cells within the same antibody used but not between different antibodies. Lane 1: uninduced DH5α (pBADEhisTEVMT); lane 2: DH5α (pBADEhisTEVMT) induced with 0.1% arabinose for 3 hours; lane 3: induced high cell-density cells of DH5α (pBADEhisTEVMT); lane 4: uninduced DH5α (pBADEhisMT); lane 5: DH5α (pBADEhisMT) induced with 0.1% arabinose for 3 hours; lane 6: uninduced BL21-AI™ (pETEhisMT); lane 7: BL21-AI™ (pETEhisMT) induced with 1 mM IPTG and 0.2% arabinose for 3 hours; lane 8: BL21-AI™ (pETEhisMT) induced with autoinduction method.
4.5 Discussion

4.5.1 KpsE forms homo-oligomeric complexes independent of other capsule proteins and K5 polysaccharide

The data from size exclusion chromatography of DH5α (pBADE) presented here suggest that KpsE self-associates as at least dimers and tetramers in the absence of other capsular gene products and the K5 polysaccharide. The finding that KpsE can form dimers is in support of previous work on KpsE using chemical cross-linkers in the presence as well as absence of other Kps proteins and the K5 polysaccharide (Arrecubieta et al., 2001). These cross-linking experiments, however, failed to identify any higher oligomers. The experiments presented here are the first to identify KpsE as a potential tetramer.

The literature on the oligomeric states of other PCP proteins reveals diversity of oligomeric organisation within this protein family. PCP-2 protein Wzc has been identified as tetramer in the presence of its binding partner Wza and the remaining capsular gene products (Collins et al., 2006). The chain-length regulators WzzE of E. coli and WzzB of S. Typhimurium have been shown to self-associate as octamer and pentamer, respectively (Tocilj et al., 2008). FepE (WzzFepE) of E. coli can assemble oligomers consisting of 3, 9, and 18 protomers (Tocilj et al., 2008), while WzzSF of Shigella flexneri can form dimers and higher order structures (Papadopoulos and Morona, 2010). Mutational analysis in WzzSF demonstrated a direct relationship between protein function (chain-length regulation) and oligomeric state (Papadopoulos and Morona, 2010). Mutants which were able to form the native stable dimers produced O-antigen chain lengths similar to the wild-type or slightly larger. Mutants which were unable to assemble as dimers produced shorter O-antigens. These data, taken together with the observation of the variety of oligomeric organisation present in PCP proteins, suggest that the oligomerisation of PCP proteins is important for protein function.

It is plausible to suggest that KpsE forms oligomeric structures higher than a dimer, as KpsE is likely to share the channel-like structure observed in other PCP proteins (Collins et al., 2006; Tocilj et al., 2008; Larue et al., 2011), which is thought to provide an export pathway for the substrate. Being able to accommodate the polysaccharide substrate in such a channel is structurally more plausible with a number of protomers.
higher than two. However, it cannot be excluded that KpsE might occur in vivo in several oligomeric states, and the data presented here would be supportive of this. This phenomenon is, for instance, seen in the pig heart CoA transferase, which can be active as dimer as well as tetramer (Rochet et al., 2000). The oligomeric state of KpsE in the presence of KpsM will be discussed in section 4.5.3.

The results from the size exclusion chromatography presented here clearly show that KpsE can form dimers and tetramers. However, as the analysis was carried out in a capsule-deficient host strain, it remains unclear whether KpsE changes its oligomeric organisation in the presence of other capsular gene products and/or during K5 export. Therefore, a further step towards elucidating the mechanisms behind K5 export would be to analyse the oligomeric state of KpsE in a capsule-producing host using size exclusion chromatography. As previous work on KpsE did not observe a difference in KpsE dimerisation in a capsule-deficient versus capsule-producing host (Arrecubieta et al., 2001), it is predicted that size exclusion chromatography of KpsE in the presence of the complete capsular gene cluster will detect higher order structures of KpsE. However, these oligomers might differ from the ones observed in the absence of other Kps proteins and the substrate, as it is not unlikely that the oligomeric state of KpsE switches between a ‘resting’ state in the absence of the substrate and an ‘engaged’ stage in the presence of the substrate. Indeed, one step towards answering the question of whether KpsE oligomerisation is affected by other Kps proteins was taken here by comparing the oligomeric state of KpsE in the absence and presence of its putative binding partner KpsM (cf. section 4.5.3).

To further elucidate the native oligomerisation state(s) of KpsE in the future, it would be interesting to identify regions of KpsE which are important for correct oligomerisation. In Wzc, the cytoplasmic C-terminal domain of the protein was found to be crucial for a tetrameric assembly (Doublet et al., 2002). Interestingly, in Wzc the second transmembrane domain was not required for correct oligomerisation. In a similar approach, truncation mutants of KpsE could be applied to size exclusion chromatography, comparing the oligomeric organisation of wild-type and mutants. The C-terminal deletion mutant generated in Chapter 3 (section 3.3.2) could be used as a starting point in this analysis.
α-Helical coiled-coil domains are known to play important roles as folding and oligomerisation motifs in proteins (Burkhard et al., 2001). In the WzzB protomer, two α-helices form an anti-parallel coiled-coil interaction with a third α-helix in the periplasmic domain of the protein (Tocilj et al., 2008). These helices are also involved in interactions with the helices from adjacent protomers in the pentameric state of the protein (Tocilj et al., 2008). According to computational analysis of the KpsE primary structure using the PairCoil2 programme (McDonnell et al., 2006), KpsE is predicted to form coiled-coil interactions in at least one (residues ~184-220), and possibly in three (residues ~226-255 and ~264-292), α-helical regions. These regions are within the three long α-helical domain of KpsE and are located in the periplasm (cf. Figure 3.1). Comparing the predicted secondary structure of KpsE with the experimentally confirmed tertiary structure of WzzB indicates that the three helical domains in KpsE might correspond to the three helical domains of WzzB which are involved in important protomer and protomer-protomer interactions. Closer analysis of these residues in KpsE might reveal interesting information about intra- and inter-protomer interactions of KpsE.

4.5.2 KpsM and KpsE form protein-protein complex after proteolytic release from nickel-column and after size exclusion chromatography

Previous work has shown that KpsE is co-purified by his-tagged KpsM using standard affinity chromatography (Hudson, 2009). Here, one objective of using nickel-affinity chromatography in combination with proteolytic cleavage as an alternative approach was to investigate whether the interaction between KpsM and KpsE remains when the affinity tag on KpsM is cleaved off. The results presented here show that KpsE and KpsM are present in the same fraction after release from the nickel column through proteolytic cleavage. This suggests that the two proteins interact strongly with each other and support the evidence from previous results using standard nickel-affinity chromatography on its own. The presented data from size exclusion chromatography of KpsEMT support this observation. The co-presence of KpsE and KpsM in several size exclusion fractions confirms the data of the affinity chromatography approach. The observed KpsE-KpsM interactions occur in the absence of other capsule proteins (except KpsT, which is co-expressed from the plasmid) and the K5 polysaccharide. Taken together, the data presented here further support the hypothesis that KpsE and
KpsM might form a pre-complex prior to engagement with the polysaccharide substrate.

4.5.3 KpsE and KpsM(T) form hetero-oligomeric complexes in the absence of the K5 polysaccharide

Information about the oligomeric organisation of the putative KpsE-KpsM complex would greatly enhance the understanding of the mechanisms behind K5 polysaccharide export. Two of the objectives of the analysis of KpsE and KpsM by size exclusion chromatography were to obtain information about the oligomerisation of KpsE-KpsM and to investigate whether KpsE changes its oligomeric state in the presence of KpsM. As size exclusion chromatography cannot provide information about the oligomeric state of individual components of a hetero-oligomeric protein complex within a shoulder, it remains unclear into which oligomeric states the two proteins assemble when in a putative complex. Table 4.2 illustrates that, except for the smallest peak which can only be a monomer-monomer KpsEM complex due to its predicted molecular weight, there were always at least 2 different oligomeric states of the KpsE-KpsM complex possible within each observed shoulder.

The results from size exclusion chromatography suggest that KpsE, KpsM and KpsT are present in a single fraction from the shoulder with the highest molecular weight. Assuming KpsM and KpsT are presented in equal amounts, this size of the shoulder corresponds to either a potential hexamer-tetramer-tetramer or a trimer-hexamer-hexamer KpsEMT complex. The former complex appears to fit the predicted size of the shoulder more accurately than the latter. The absence of hexameric KpsE from size exclusion chromatography of KpsE in the absence of KpsM might suggest that KpsE can only form hexamers in the presence of KpsM and otherwise forms dimers and tetramers. The hexamer of KpsE might consist of three sets of dimers. It could be hypothesised that KpsE is more stable as dimer subunits than as hexamer. This would explain the reoccurring observation of dimeric KpsE in the literature. However, a complex consisting of trimeric KpsE and hexameric KpsM is also possible. The predicted diversity of oligomeric states in the putative KpsEM complex might be interpreted as intermediate oligomeric stages and might suggest that the equilibrium of the KpsEM monomers and oligomers is dynamic. As discussed for the results from size
exclusion analysis of KpsE (section 4.5.1), analysing the fractions of size exclusion chromatography from a capsule-producing strain for the presence of KpsE, KpsM, and KpsT would be very useful for establishing oligomerisation dynamics during polysaccharide export.

KpsMT had previously been predicted to be active as a dimer during ATP hydrolysis and inner membrane transport of the K5 polysaccharide due to its membership in the ABC transporter family (Pigeon and Silver, 1994). Many ABC transporters have been identified as dimers, but the focus of analysis was more directed towards the oligomeric state of the NBD rather than towards the whole complex (Smith et al., 2002; Zaitseva et al., 2005; Guo et al., 2006). As these studies were carried out using isolated ABC domains they might not reflect the situation in the complete ABC transporter. Some ATP-binding proteins have been shown to form higher order homo-oligomers. The inner membrane ATPase EscN, which is part of the E. coli type 3 secretion system, can assemble as monomers, hexamers and dodecamers (Andrade et al., 2007) and forms a centre pore when in a hexameric state (Zarivach et al., 2007). So it appears not unlikely that KpsMT might acts as higher order oligomer.

It was a surprising observation that KpsT was detected in a single elution fraction with KpsM (and KpsE) in the absence of K5 polysaccharide. The current model of ABC transport of the polysaccharide across the inner membrane suggests that association of KpsT with the substrate leads to an ATP-induced conformational change of KpsT. The subsequent interaction of KpsT with KpsM is predicted to result in membrane insertion of KpsT and ATP hydrolysis (Bliss and Silver, 1997; Nsahlai and Silver, 2003). If the co-presence of KpsT with KpsEM in the highest molecular weight fraction is a genuine reflection of the association of KpsT with KpsM, this would suggest that the two proteins can also interact in the absence of the polysaccharide. KpsT is only found in one fraction; this might either indicate that the association with KpsM is due to non-specific interactions (i.e. detection of KpsT as an artefact of unspecific interactions), or that KpsT only interacts with KpsM when forming a specific oligomeric state (such as tetramer or hexamer). In the latter case, this oligomeric state might involve a pre-complex of KpsT and KpsM in the absence of the polysaccharide; the oligomeric organisation of KpsMT might change upon K5 binding. In contrast, it is also possible that KpsT associates with KpsM in lower molecular weight oligomers but that the
interactions are not strong enough and the two proteins dissociate during chromatography.

As it becomes evident from the discussion of the data obtained through size exclusion chromatography of KpsEM, many unanswered questions remain and the identity of the functional form(s) or the hetero-oligomeric complex of KpsEM(T) is still to be revealed. Possible future directions towards establishing the oligomeric state(s) of KpsEM(T) will be discussed in section 4.5.6.

4.5.4 The absence of a hisKpsM Western blot signal at the monomeric size of KpsM after size exclusion chromatography

Western blot analysis of the size exclusion fractions from DH5α (pBADEhisMT) could not detect a his-signal at 29.5 kDa, the molecular weight KpsM is usually observed at, but revealed a hisKpsM-specific signal just below the 50 kDa molecular weight standard. This signal was not observed in size exclusion fractions from DH5α (pBADE) and was equally absent from whole cell or total membrane preparations of DH5α (pBADEhisMT). The reasons for this observation remain unclear. Abnormal SDS-PAGE migration (gel shift) is a common occurrence in hydrophobic membrane proteins and might result in proteins locating at a size from -46% to +45% of their predicted molecular weight on the gel (Rath et al., 2009). This observation has been reported for inner membrane proteins such as BtuC (Deveaux et al., 1986) and AcrA (Pos and Diederichs, 2002) and has been attributed to the altered binding affinity of the protein to SDS (i.e., binding a higher or lower amount of SDS than globular proteins of identical size (Rath et al., 2009). However, in the case of KpsM described here, the his-signal of hisKpsM is usually found at the predicted size for KpsM and the phenomenon of a higher signal is only seen after size exclusion chromatography. It can only be hypothesised that the size exclusion procedure is able to isolate oligomeric (most likely dimeric) KpsM which binds different to SDS than the form purified from affinity chromatography or the form existent in whole cells and membranes. Possible future analyses which are discussed in section 4.5.6 might be able to fully elucidate the observed phenomenon.
4.5.5 Suitability of affinity chromatography combined with proteolytic cleavage to isolate KpsE-KpsM complex

The results presented here show that nickel-affinity chromatography followed by on-column cleavage by TEV protease was in principle successful in isolating KpsE and KpsM in the same purification fraction. However, the putative protein complex could not be used for further structural studies as the protein quantity was very low. It was therefore impossible to assess the stability of the protein-protein complex in comparison to the stability of complexes isolated by standard nickel-affinity chromatography. There are several factors that could have affected the efficiency of the approach, most notably low protein expression, non-optimal conditions during the purification procedure, and poor protease cleavage activity.

Buffer and incubation conditions were carefully selected based on successful previous use of TEV protease (Liu and Naismith, 2009) and own experimental trials. TEV protease is active in a wide range of conditions (pH 4-9, NaCl 0-400 mM, and temperatures 4-30°C) and is thus suitable for affinity-tag cleavage. TEV protease can be inhibited by some detergents, but the detergent DM used in the experiments presented here had previously been shown not to have a negative effect on the activity of TEV protease (Mohanty et al., 2003). However, it cannot be excluded that the protease activity was inhibited by other factors during the purification protocol. On-column cleavage is expected to be less efficient compared to cleavage in solution due to unstable protease and possible un-specific column binding of the protease. A 1:100 protease-substrate ratio was found to be most efficient. However, a further increase in protease quantity during incubation might produce increased cleavage. Furthermore, selecting a different, more stable variant of TEV protease, which exhibits higher proteolytic efficiency and reduced autolytic activity (such as the S219V mutant generated by Kapust et al. (2001)) might increase cleavage levels. For the purpose of future work with TEV protease, it might also need to be investigated whether TEV protease efficiency is inhibited by inaccessibility of the protease cleavage site. Additional amino acid residues might need to be added between the his-tag sequence and the protease recognition sequence on the construct to provide better access to TEV protease.
One logical conclusion from the lack of clearly visible protein bands after on-column cleavage would be to suggest that the majority of the target protein did not bind to the nickel column during previous steps of the purification protocol. However, Western blot analysis showed that only small amounts of KpsE and hisKpsM were present in the flowthrough and low-concentration imidazole wash (Figure 4.5), indicating that the majority of protein was retained in the column for the proteolytic cleavage step. Nevertheless, improving protein expression levels might still have a positive effect on obtaining a higher final protein yield. Even though KpsE and hisKpsM were detected in total membranes prior to purification, the protein quantity appears not to be adequate for obtaining larger amounts of purified protein. Once high expression levels can be obtained, purification from inner membranes rather than total membranes will provide purer purification products.

In conclusion, taken into considerations improvements with respect to protein expression and TEV protease efficiency, nickel-affinity chromatography combined with TEV proteolytic cleavage is suitable to isolate the putative KpsE-KpsM complex in higher quantities.

4.5.6 Suitability of size exclusion chromatography to isolate the KpsE-KpsM complex and to study KpsE-KpsM oligomerisation

If basing interpretations on the assumption that the observed his-specific signal represents the presence of hisKpsM, the data presented here suggest that size exclusion chromatography is suitable to isolate the putative KpsE-KpsM complex. However, as a sufficient yield of the complex after purification could not be obtained, it was impossible analyse whether a complex obtained through size exclusion chromatography is more stable than a complex isolated through other techniques. This question can be addressed in future work after successful improvement of protein yield.

Size exclusion chromatography on KpsE was successful in supplying interesting information about the oligomeric state of KpsE. As expected, size exclusion chromatography is a useful technique to obtain data about the oligomeric self-association of a single protein. However, size exclusion has limitations with respect to the analysis of the hetero-oligomeric state of a protein complex. Even though
information obtained through size exclusion analysis of the KpsE-KpsM complex provides a starting point for further analysis, size exclusion chromatography was unable to provide unambiguous results. Interpretation of the result from size exclusion chromatography of KpsE-KpsM(T) was additionally impeded by the fact that KpsM cannot be stably expressed without the presence of KpsE. This made comparison of these results to data from size exclusion chromatography of KpsM impossible. The first step towards gaining a more accurate picture of the oligomeric organisation of the complex would be an analysis of the protein complex by Blue Native PAGE. Analysis of the size exclusion fractions in the absence of detergent is predicted to provide information about the molecular weight of KpsE and KpsM in the different fractions. Using inner membranes, Blue Native PAGE analysis might also reveal additional oligomeric states to the ones observed by size exclusion chromatography. The type 3 secretion system ATPase EscN was identified as monomer and dodecamer by size exclusion chromatography; analysis by Blue Native PAGE revealed additional hexamers and intermediate oligomeric states (Andrade et al., 2007). If the complex is still found to dissociate too quickly, analysis by perfluoro-octanoic acid (PFO)-PAGE might be worth considering. PFO is a relatively mild detergent which has been shown to preserve the high affinity interactions found within oligomeric complexes of membrane proteins (Ramjeesingh et al., 1999). In this technique, proteins are run on a standard Tris-glycine gel (in the absence of SDS) with PFO present in the loading buffer. Preparation for PFO-PAGE is thus less technically demanding than preparation of a Blue Native gel. PFO-PAGE has been shown to provide accurate information about the oligomeric state of homo-oligomers (Ramjeesingh et al., 1999) and has more recently been used to identify the oligomeric organisation of Wzc after size exclusion chromatography (Collins et al., 2006). In the future it could be investigated whether this approach, complemented with Western blot analysis, provided further information about the oligomerisation of KpsE-KpsM.

In order to obtain more accurate data about the molecular weight of the oligomers obtained by size exclusion chromatography, proteins could be analysed by multi-angle light scattering. Unlike size exclusion chromatography, light scattering determines molecular weight independent of the shape of the protein. The technique can also be used to determine the stoichiometry of protein-protein interactions, providing information about the ratio of molecule A to molecule B.
A further technique to gain a better understanding of the oligomeric state(s) of KpsE-KpsM as well as of the nature of their protein-protein interaction is analytical ultracentrifugation. Analytical ultracentrifugation measures concentration distribution of the sample by absorbance, interference or fluorescence. The main advantage of analytical ultracentrifugation is that it analyses proteins in their native conditions and in the absence of any matrixes or surfaces. Samples can also be used for further analysis after analytical ultracentrifugation. The technique is very versatile and can be applied as two principle methods: sedimentation equilibrium and sedimentation velocity (cf. Cole et al. (2008) for a review). Sedimentation velocity can provide hydrodynamic data about the size and shape, including folding stage and oligomeric state, of proteins. Sedimentation equilibrium can be used to determine the molecular weight of monomers and complexes, as well as the stoichiometry of hetero-oligomers. In heterologous protein complexes, sedimentation velocity can determine the ratio of free and complexed protein and provide information about the binding stoichiometries of the two proteins (as multi-angle light scattering) (Rivas et al., 1999). One limitation of analytical ultracentrifugation applicable in the context of the K5 translocon proteins is that the technique requires around 95% pure protein to provide reliable data. Results from analytical ultracentrifugation of the K5 translocon proteins would make an extremely valuable contribution to the model of K5 polysaccharide export with respect to the in vivo state of the K5 translocon complex. Analytical ultracentrifugation, and possibly multi-angle light scattering, could determine the oligomeric states of KpsE and KpsM and elucidate in which ratio their oligomers are present and whether this changes in the presence of other Kps protein and the substrate.

4.5.7 Future directions for obtaining higher KpsE and KpsM yields

Difficulties with obtaining high yields of membrane proteins due to expression issues are a well known problem. In a case like the one of the K5 translocon complex the dependency of the proteins on each other for stability is a further challenge. Low expression issues have also been reported for the HlyA secretion system members HlyB and HlyD (Wang et al., 1991). In the over-expression trials of KpsE and hisKpsM presented here, the system which provided the best over-expression results for both proteins simultaneously was the BL21-AI (pET-24(+)) host-vector system induced by 1 mM IPTG and 0.2% arabinose for 3 hours. However, compared to the other systems
and induction conditions studied, expression was not improved sufficiently to result in significantly higher protein yields during affinity chromatography. There are several improvements which might further improve expression levels and protein yields after purification.

One reason for obtaining low protein yields might be that the over-expressed proteins aggregate inside inclusion bodies. Lowering the temperature from 37°C to 30°C or lower during induction might decrease the accumulation of proteins in inclusion bodies and thus increase the yield of soluble protein (Schein, 1989). Geertsma et al. (2008) have developed a fast fluorescence assay based on C-terminal green fluorescent protein-fusions to determine the quantity of correctly folded and non-folded protein after over-expression. This technique could be applied to test whether aggregated proteins are a problem in the current expression systems, and if this was found to be the case the assay could also determine the ratio of folded to non-folded proteins. Although the incorporation of a C-terminal GTP-tag into the translocon proteins requires careful consideration, this folding assay would also provide concurrent information about the suitability of such a tag on the test protein.

In order to increase protein yield after nickel-affinity purification, it might be advisable to investigate strategies to enhance the binding affinity of the his-tag to the resin during chromatography. Increasing the length of the his-tag might be one option. Deca-his-tags have been shown to produce better results compared to hexa-his-tags for immobilising his-tagged proteins to nickel-coated surfaces (Fischer et al., 2011). Additionally, deca-his-tags have successfully been used to purify proteins such as CtrB of N. meningitidis (Larue et al., 2011) and the O-antigen ligase WaaL of Helicobacter pylori (Hug et al., 2010) by nickel-affinity chromatography. However, increasing his-tag length might not always result in higher protein yield after purification as a study on the aquaporin AqpZ demonstrated. Expressing the aquaporin AqpZ as deca-his-tagged variant rather than hexa-his-tagged variant reduced protein yields after affinity chromatography by three-fold (Mohanty and Wiener, 2004). It would need to be investigated whether increasing the length of the his-tag on KpsM resulted in an improvement of protein yield after nickel-affinity chromatography.
Chapter 5

Strategies for the topology analysis of KpsD
5.1 Introduction

The outer membrane protein KpsD is the least studied component of the group 2 polysaccharide translocon. It is anticipated that elucidating the structure, including topology, of KpsD will provide valuable insights into KpsD function. Approaches such as epitope tagging and protease cleavage have been used to map the topology of various outer membrane proteins (Harms et al., 1999; Guedin et al., 2000; Kulkarni et al., 2003; Henderson et al., 2004; Surana et al., 2006). Previous attempts to characterise the topology of KpsD involved enterokinase cleavage site-insertions with subsequent protease cleavage analysis (McNulty, 2005). This approach identified one surface-exposed and three periplasmic sites within KpsD. However, as a large number of enterokinase cleavage site-fusions were non-functional, reliable mapping of further residues could not be carried out. Attempts to obtain crystals of KpsD have proven unsuccessful to date (Thompson, J. and Roberts, I., unpublished data). This chapter discusses structural and topological predictions of KpsD and investigates two strategies – cysteine-scanning mutagenesis and c-myc-epitope insertion analysis – for their suitability in mapping the topology of KpsD.

5.2 Topological and structural predictions of KpsD

The transmembrane domain of most outer membrane proteins are made by $\beta$-strands which form a $\beta$-barrel (Koebnik et al., 2000; Schulz, 2002). Typical examples include OmpA (Pautsch and Schulz, 1998; Arora et al., 2001), TolC (Koronakis et al., 2000), PapC (Henderson et al., 2004) and FhaC (Guedin et al., 2000). However, notable exceptions of this rule are Wza (Dong et al., 2006) and the VirB10 protein TraF (part of a three-protein outer membrane complex of the type 4 secretion system) (Chandran et al., 2009), which cross the outer membrane via an $\alpha$-helical channel. To date, Wza is the only outer membrane polysaccharide export protein whose crystal structure has been resolved (cf. Figure 1.4), and it remains unclear whether the helical channel structure is common among outer membrane polysaccharide exporters. However, theoretical predictions of the secondary structure of CtrA (Frosch et al., 1992) as well as KpsD (see below) suggest that these two proteins form a $\beta$-barrel across the outer membrane.

In contrast to the polysaccharide co-polymerases involved in polysaccharide inner membrane and periplasmic export, the outer membrane components of polysaccharide
export systems are more divergent with respect to their amino acid sequence (cf. Figure 1.3 and Table 1.2). ClustalW alignment (Thompson et al., 1994) of full-length Wza and KpsD revealed that the two proteins share 13% identical and 28% similar residues. This is not surprising, as various properties of KpsD have been known to differ from those of Wza. KpsD is 179 residues longer than Wza and is not a lipoprotein. The N-terminus of Wza is acylated (Nesper et al., 2003), a characteristic absent in KpsD. Both proteins locate their N-terminus to the periplasmic face of the outer membrane (McNulty, 2005; Dong et al., 2006). However, the observation that the location of the C-terminus in Wza is the extracellular environment and in KpsD is in the periplasm (McNulty, 2005; Dong et al., 2006) suggests a difference in structure of the two proteins.

Interestingly, BLAST alignment (Altschul et al., 1990) identified two regions within KpsD which show higher number of similar residues compared to sequence similarity across the full sequence. ClustalW analysis of these regions showed that KpsD90-166 shares 29% identical and 47% similar residues with Wza96-162, and KpsD445-528 shares 31% identical and 48% similar residues with Wza240-336 (Figure 5.1). Wza96-162 and KpsD90-166 carry the PES domain, which is conserved among outer membrane polysaccharide export proteins to various degrees (Dong et al., 2006). In Wza, the PES region forms domain 1, which is the periplasmic domain furthest away from the outer membrane and closest to, and likely interacting with, Wzc (cf. Figures 1.4 and 1.5).

In order to construct a model of the topology of KpsD which could be tested experimentally, the secondary structure and topology of KpsD were predicted using theoretical prediction algorithms and the available experimental results obtained by McNulty (2005). Firstly, the overall secondary structure of KpsD was predicted using I-TASSER (Zhang, 2007; Roy et al., 2010). The SWISS-MODEL programme (Schwede et al., 2003) was then used on regions 96-325 and 460-529, which show some extended homology to the core homology regions 90-166 and 445-525, in order to obtain a more accurate secondary structure prediction based on the known structure of Wza. The secondary structure model of KpsD presented here was constructed by combining the information from the two prediction software (Figure 5.2, top line). The N-terminal region preceding the PES domain contains 2 small α-helices. In Wza, the region preceding the PES domain contains a β-α-β-α motif, of which the latter two helices
<table>
<thead>
<tr>
<th>Wza</th>
<th>90</th>
<th>GDVLMOVTVWDHELTTPAGQYRSASDTGNW</th>
<th>119</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpsD</td>
<td>96</td>
<td>GDSIQVRLW----------GAFT--FDGALQ</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wza</td>
<td>120</td>
<td>VNSDGTIFYPYIGKVQVGKTVSQVRQDIT</td>
<td>149</td>
</tr>
<tr>
<td>KpsD</td>
<td>115</td>
<td>VDPKGNIFLPNVGPVKIAGVSNSQNLV</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wza</td>
<td>150</td>
<td>SRLTT--YIESPOQDVSIA</td>
<td>166</td>
</tr>
<tr>
<td>KpsD</td>
<td>145</td>
<td>SKVEVYQSNVNVYASL</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wza</td>
<td>240</td>
<td>LLYHGDIFIPSNDDLKVFVMGEVGKQSTLKMDR</td>
<td>273</td>
</tr>
<tr>
<td>KpsD</td>
<td>445</td>
<td>LLEDGDVINPEKTSL--VMVHEVLPNAVSWQK</td>
<td>477</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wza</td>
<td>274</td>
<td>SGMTEALGAEGISQEMSDATGIFVQRQLGD</td>
<td>307</td>
</tr>
<tr>
<td>KpsD</td>
<td>478</td>
<td>-GMITEDYIEKCGGLTQKSQNR--IVIRQ--NGA</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wza</td>
<td>308</td>
<td>RTGKIAIYQLNASDASVMLGTEFQLQP</td>
<td>336</td>
</tr>
<tr>
<td>KpsD</td>
<td>509</td>
<td>R----------VNAEDVDSLKFGDEIMVLPR</td>
<td>528</td>
</tr>
</tbody>
</table>

**Figure 5.1 Clustal alignments of homologous regions between KpsD and Wza.** Sequence alignments of the PES region of Wza (residues 90-166) and KpsD (residues 96-162) as well as of residues 240-336 of Wza and residues 445-528 of KpsD were generated using ClustalW (Thompson et al., 1994). Identical residues are highlighted in red. Wza of *E. coli* K30 (gene bank accession number Q9X4B7) and KpsD of *E. coli* K5 (P42213).
Figure 5.2 Model of the secondary structure and topology of KpsD. The secondary structure was constructed on the basis of predictions using I-TASSER (Zhang, 2007; Roy et al., 2010) and SWISS-MODEL (Schwede et al., 2003); the predictions by SWISS-MODEL used Wza as template. SS denotes the signal sequence of KpsD (residues 1-18). α-Helical regions are depicted in blue and β-sheets are depicted in orange. α-helices and β-sheets which are predicted by I-TASSER but not by SWISS-MODEL are shown in light blue and light orange, respectively. The location of cysteine 488 is indicated in black. The symbols ° and * depict the location of surface-exposed and periplasmic residues according to previous experimental data (McNulty, 2005), namely residues 94-95, 126-127, 149-150, and 368-369, respectively. Topology information is given in green. The predicted topology of KpsD is based on theoretical predictions by SWISS-MODEL and previous experimental data (McNulty, 2005). The theoretical domains of KpsD relating to the three-dimensional model in Figure 5.3 are illustrated in black.
are located on the sites of domain 2 and 3, respectively (cf. Figure 5.3). As in Wza, the PES domain of KpsD is made of β-sheets and a single helix. Similarly, the region following the PES domain up to the helical C-terminus is primarily made β-sheets and interspersed helices. The only exception within these β-β-α-β motifs is a helical region between residues 380 and 440, which is predicted to form coiled coils.

Fluorescence activated cell sorting of truncated variants of KpsD previously determined that residues 19-179 are associated with the outer membrane, while residues 180-558 are located in the periplasm (McNulty, 2005). Enterokinase digestions of fusion-variants of KpsD identified residues 94-95 as surface-exposed and residues 126-127, 149-150, and 368-369 as periplasmic (McNulty, 2005). I-TASSER analysis of KpsD predicted an 8-stranded β-barrel transmembrane structure. This fits with the model proposed by Christian Anderson, University of Würzburg, Germany (2005), who developed an algorithm for topology predictions of KpsD. According to this model the transmembrane region of KpsD has 8 transmembrane β-strands with four surface-exposed loops and three periplasmic loops. Combining the information from the experimental data and secondary structure and topological predictions, a topology model of KpsD was created (Figure 5.2, second and bottom row). In the model presented here, the outer membrane is crossed eight times, twice via α-helices and six times via β-sheets. The major part of the transmembrane region is located in the PES domain, while the first and the last transmembrane region lie just before and after the PES domain.

In order to gain a better insight into the possible tertiary structure of KpsD, a ribbon diagram of KpsD_{56-558} was created using SWISS-MODEL by utilising the three-dimensional data from Wza as structural template (Figure 5.3). As homology between KpsD and Wza is highest at residues 96-325 and 460-529 (KpsD numbering), the accuracy of the structural model is highest within these two regions. No tertiary information was provided for residues 19-85 as homology between the two proteins is too low within this region. The tertiary model of KpsD proposes 6 domains (Figure 5.3). Even though the motifs of mixed α-helix-β-sheet structure of domains 1-3 in Wza and domains 2-5 in KpsD suggests relatively high structural similarity between these domains of the two proteins, the structural orientation of KpsD appears to be turned ‘bottom up’ compared to Wza (Figure 5.3). In contrast to Wza, the PES region in
Figure 5.3 Three-dimensional model of KpsD_{86-555} based on KpsD-Wza alignment. The illustration shows two selected rotational angles of a three-dimensional model of the KpsD_{86-555} monomer (right) and the established structure of the Wza monomer (Protein Data Bank ID: 2J58) (Dong et al., 2006) for comparison (boxed on left). To aid comparison of domains between Wza and KpsD, the Wza monomer is depicted as its vertical mirror image. The ribbon diagram of selected domains of monomeric KpsD was constructed using SWISS-MODEL (Schwede et al., 2003), based on the structural data of Wza as an alignment template. Residues immediately preceding and succeeding an α-helix or β-sheet motif are highlighted. Domain 2: 98-162; domain 3: 167-246; domain 4: 250-323; domain 5: 332-527; domain 6: 530-555). Note that the model is most accurate within the regions of KpsD which show homology with Wza (residues 96-325 and 460-529). Due to low homology with Wza, no model of residues 19-80 (domain 1) could be constructed.

Please see next page for figure.
Figure 5.3 Three-dimensional model of KpsD86-555 based on KpsD-Wza alignment.
KpsD forms the top most part of the structure and anchors the protein into the outer membrane. While the α-helical C-terminus of KpsD is located in the periplasm, in Wza it forms the helical barrel in the outer membrane. Wza is proposed to act as octamer, with a helical membrane barrel periplasmic made by domain 4 and three rings with a central cavity formed by domains 1-3 (Dong et al., 2006). If the quaternary structure of domains 2-5 in KpsD is similar to that of Wza, domains 2-5 of oligomeric KpsD would form 4 rings, with the top region integrated into the membrane and the bottom part in the periplasm. The proposed β-sheets would be facing the inside of the ring and the α-helices would be facing the outside. Future analysis of KpsD using techniques such as transmission electron microscopy would provide important insights into the quaternary structure of KpsD. The first step towards elucidating the quaternary structure of KpsD would be an in-depth study of the in vivo oligomeric state of KpsD.

5.3 Substituted-cysteine accessibility method

Having established a topology model of KpsD which can be tested, it was decided to investigate whether the substituted-cysteine accessibility method (also known as cysteine-scanning mutagenesis) is suitable for topology analysis of KpsD. Substituted-cysteine accessibility method uses site-directed mutagenesis of the target protein and utilises the reactivity of the sulfhydryl group in cysteines with sulfhydryl-specific chemical agents to elucidate the topology of membrane proteins. Membrane-permeable sulfhydryl-reactive reagents can distinguish between cysteines accessible (i.e. in the aqueous environment) and those inaccessible (i.e. buried in the membrane) to the reagent. The combined use of membrane-permeable and membrane-impermeable sulfhydryl-specific reagents enables to map the location of the cysteine residue more specifically to the surface or the cellular/periplasmic environment. Treatment of the target cell or membrane with the membrane-impermeable reagent labels cysteines exposed to the cell surface. This pre-treatment is followed by treatment with the membrane-permeable reagent, which labels cysteines on the other side of the membrane. In a Gram-negative organism, a residue with periplasmic location shows biotin labelling unaffected by pre-treatment with the membrane-impermeable reagent, while a residue with extracellular location shows a reduction in biotin label by pre-treatment with the membrane-impermeable reagent. A residue with transmembrane
location shows little or no biotinylation unaffected by pre-treatment with the impermeable reagent.

The major advantage of cysteine-scanning mutagenesis over epitope-tagging or protease cleavage approaches for topology analysis is that cysteine mutagenesis minimally interferes with the native sequence of the target protein and thus analyses the location of residues under near-native conditions. Using variants of the technique and a range of sulfhydryl-reactive reagents, cysteine-scanning mutagenesis has been applied to study the topology of a wide variety of eukaryotic as well as prokaryotic membrane proteins in order to circumvent problems with crystallography approaches encountered with many membrane proteins. Application of cysteine-scanning mutagenesis in molecular bacteriology has revealed the topology of membrane proteins such as the lactose permease of *E. coli* (Frillingos et al., 1998), HasA of group A streptococci (Kumari et al., 2002), the outer membrane usher protein PapC (Henderson et al., 2004), the twin-arginine translocation pathway inner membrane protein TatC (Punginelli et al., 2007), and the inner membrane polysaccharide co-polymerase CtrB of *N. meningitidis* (Larue et al., 2011).

It was envisaged to first create a cysteine-less variant of KpsD which could be used as PCR template for site-directed mutagenesis and to then create a library of single cysteine mutants which could be used for cysteine-mapping. On the basis of findings from previous reports using different sulfhydryl-reactive reagents, biotin maleimide and qBBr were selected as the membrane-permeable and membrane-impermeable reagents, respectively. It was decided to use the pBADED<sub>cys-variant</sub> constructs carrying KpsE and KpsD cys-variant for mutagenesis and functional analysis of the KpsD cys-variants by K5 bacteriophage assay, and subsequently clone the *hiskpsMT* fragment into pBADED<sub>cys-variant</sub> (to obtain pBADED<sub>cys-variant,hisMT</sub>) for biotin labelling and KpsD cys-variant analysis in order to mimic conditions closer to the native translocation conditions.

### 5.3.1 Expression trial of DH5α (pBADED)

To determine the optimal protein expression conditions from DH5α (pBADED), whole cells were induced at OD<sub>600</sub> 0.4-0.5 using either 0.1%, 0.2%, 0.5%, or 1%
arabinose for either 3, 6, or 22 hours. Whole cells were analysed by Western blot using anti-KpsD antibody (Figure 5.4). As expected, induction with 0.1% arabinose for 3 hours resulted in the lowest expression of KpsD. Induction with 0.1% arabinose for 6 hours and 1% arabinose for 3 hours increased expression. Induction with 0.1% arabinose for 22 hours, or 0.2% and 0.5% arabinose for 3, 6, or 22 hours further increased KpsD expression. Induction conditions of 0.1% arabinose for 22 hours were selected for cysteine-scanning mutagenesis experiments.

5.3.2 Generation of construct carrying cysteine-less kpsD

The wild-type KpsD sequence contains two cysteines. As the cysteine at position 13 is located in the signal sequence which is cleaved off post-translationally, the only cysteine in processed KpsD is a cysteine at position 488. In order to create a cysteine-less KpsD variant which could be used as negative control and background for cysteine substitutions, a single substitution from cysteine 488 to serine was introduced into KpsD in the plasmid pBADED. Serine only differs from cysteine by a single atom; in serine the sulphur atom is replaced by a oxygen, resulting in the formation of a hydroxyl rather than a thiol group. Site-directed mutagenesis was carried out by PCR-based QuickChange® (section 2.5.3) using primers KpsD C488S F and R (Table 2.3). The QuickChange® reaction was transformed into DH5α, and mutant plasmids were isolated (section 2.5.5). Colony PCR (Figure 5.5) and DNA sequencing were used to identify and verify the mutations, respectively. Successful mutagenesis of cysteine 488 to serine resulted in the plasmid pBADEDΔcys (Figure 5.6).

To determine whether cysteine-less KpsD is functional, pBADEDΔcys was transformed into MSΔD and a K5 bacteriophage assay was carried out. MS101 and MSΔD (pBADED) served as positive controls; MSΔD served as negative control. The plasmid pBADEDΔcys was able to complement MSΔD, producing a plaque count higher than the wild-type (Table 5.1). This suggests that KpsDΔcys is functional as part of the K5 translocon and the cysteine at position 488 is not essential for KpsD function. The cysteine-less variant can thus be used as negative control and background for cysteine substitutions in topology analysis of KpsD by cysteine-scanning mutagenesis.
Figure 5.4 Expression trial of KpsD from DH5α (pBADED). Western blot analysis using anti-KpsD antibody. Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical density of cells. Lane 1: whole cells induced with 0.1% arabinose for 3 hours; lane 2: whole cells induced with 0.1% arabinose for 6 hours; lane 3: whole cells induced with 0.1% arabinose for 22 hours; lane 4: whole cells induced with 0.2% arabinose for 3 hours; lane 5: whole cells induced with 0.2% arabinose for 6 hours; lane 6: whole cells induced with 0.5% arabinose for 3 hours; lane 7: whole cells induced with 0.5% arabinose for 6 hours; lane 8: whole cells induced with 0.5% arabinose for 22 hours; lane 9: whole cells induced with 1% arabinose for 3 hours.

Figure 5.5 Generation of plasmid pBADEDΔcys. Plasmid pBADEDΔcys was generated as described in sections 2.5.3-2.5.10. Colony PCR of the plasmids was carried out as described in section 2.5.4. Samples were separated on a 1% agarose gel. Lane 1: colony PCR of pBADED using KpsD Start F and pBAD28 R primers, generating a 1.6 kb product; lane 2: colony PCR of pBADEDΔcys using KpsD Start F and pBAD28 R primers, generating a 1.6 kb product; lane 3: colony PCR of pBADED using pBAD28 F and R primers, generating a 2.7 kb product; lane 4: colony PCR of pBADEDΔcys using pBAD28 F and R primers, generating a 2.7 kb product.
Figure 5.6 Illustration of the plasmid pBADEDΔcys. MCS: multiple cloning site; p1: arabinose-inducible pBAD promoter; p2: AmpR promoter; term: transcriptional terminator rrnB T1; araC: arabinose transcriptional regulator; f1 ori and p15A ori: f1 and p15A origin of replication, respectively; AmpR and CmR: ampicillin and chloramphenicol resistant marker, respectively; NsiI: NsiI restriction site.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plaque count at different K5 phage dilutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10^-6</td>
</tr>
<tr>
<td>MS101</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔD</td>
<td>0</td>
</tr>
<tr>
<td>MSΔD (pBADED)</td>
<td>TMTC</td>
</tr>
<tr>
<td>MS101 (pBADED)</td>
<td>TMTC</td>
</tr>
<tr>
<td>MSΔD (pBADED Δcys)</td>
<td>TMTC</td>
</tr>
<tr>
<td>[MSΔD (pBADED C488S)]</td>
<td></td>
</tr>
<tr>
<td>MSΔD (pBADEDCTmyc)</td>
<td>0</td>
</tr>
<tr>
<td>MS101 (pBADEDCTmyc)</td>
<td>TMTC</td>
</tr>
</tbody>
</table>

Table 5.1 K5 bacteriophage assays of KpsD mutants and controls. TMTC: too many plaques to count.
5.3.3 Development of a protocol for substituted-cysteine accessibility method for KpsD topology analysis

The first step towards developing a protocol for cysteine-scanning topology analysis of KpsD was to determine the optimal biotin incubation conditions. As processed KpsD only contains a single cysteine, wild-type KpsD served as the first cysteine variant. Whole cells of DH5α (pBADEDhisMT) were induced with 0.1% arabinose for 22 hours. For the purpose of biotin-labelling (cf. section 2.7.7 for details), cells were resuspended in PBS and divided into one unlabelled control sample and three biotin-labelling samples. Cells of the biotin-labelling samples were incubated with either 0.2 mM biotin for 10 minutes, 0.2 mM biotin for 30 minutes, or 1 mM biotin for 30 minutes, respectively. The control cells were incubated in the absence of biotin for 30 minutes. Controls and biotin-labelled cells were incubated with 2% 2-mercaptoethanol for 10 minutes to stop the biotin-labelling reaction. Samples were washed with PBS and used to prepare total membranes (section 2.7.5). KpsD and biotin-labelling were visualised by Western blot using anti-KpsD antibody and by streptavidin which detects biotin (Figure 5.7 A). As expected, KpsD was detected in all four samples. Biotin detection of the unlabelled control showed some biotin-labelling. This confirmed that DH5α (pBADEDhisMT) contains several cysteines which are labelled by biotin and need to be taken into consideration when analysing the results of the labelled membranes. The biotin-signal of labelled cells was stronger than the signal from the control. As an increase in incubation time or concentration did not appear to significantly enhance the biotin signal, it was decided to use 0.2 mM biotin for 10 minutes in future experiments.

In order to identify the KpsD-specific biotin-signal in total membranes, biotin-labelled membranes of DH5α (pBADEDhisMT) were compared to biotin-labelled membranes of DH5α (pBADEhisMT) (Figure 5.7 B). The area of the KpsD-specific biotin signal (denoted by an arrow on Figure 5.7 B) could be identified by simultaneously probing the streptavidin blot with anti-KpsD antibody. However, additional biotin signals around the area of the KpsD signal obscured clear identification. In a further attempt to identify the KpsD-specific biotin signal, biotin-labelled total membranes of DH5α (pBADED) were compared to biotin-labelled total membranes of DH5α (pBADEDΔcys) (Figure 5.8). As with the DH5 α (pBADEDhisMT) strain, the area of the KpsD-specific biotin signal could be identified by simultaneously probing the streptavidin blot with anti-
KpsD antibody, but signals immediately above and below obscured to identification of a KpsD-specific signal.

In order to reduce the number of biotinylated background signals, outer membranes of biotin-labelled DH5α (pBADED) and DH5α (pBADEDΔcys) were isolated (section 2.7.6) and compared to each other by biotin detection with streptavidin and Western blot using anti-KpsD antibody (Figure 5.9). This approach provided a much clearer visualisation of the bands surrounding the area where the KpsD-specific biotin signal was predicted to locate to. However, no biotin signal specific to the DH5α (pBADED) strain which was absent in the DH5α (pBADEDΔcys) strain was detected. There are two possible explanations for the absence of a KpsD-specific biotin signal in DH5α (pBADED): the cysteine residue at position 488 is not accessible to biotin-labelling due to its location in or close to the outer membrane, or cysteine 488 is accessible to biotin but binds biotin too weakly to result in a strong signal which could be detected by the assay used here.

5.4 Myc-fusion approach: C-terminal myc-epitope insertion into KpsD

As an alternative approach to cysteine-scanning mutagenesis for the topology analysis of KpsD, a c-myc-epitope insertion approach was chosen. Accessibility of the myc-epitope from the cell surface of intact cells can be analysed by Western blot using anti-myc antibody and by flow cytometry. Comparing these results with Western blot and flow cytometry data of outer membranes can identify myc-epitopes located in the periplasm. Myc-epitope insertion analysis has been previously used to elucidate the topology of outer membrane proteins such as the fimbrial usher FaeD of *E. coli* (Harms *et al.*, 1999) and two outer membrane members of the two-partner secretion system, HMW1 adhesion secreter HMW1B of *H. influenzae* (Surana *et al.*, 2006) and filamentous haemagglutinin secreter FhaC of *Bordetella pertussis* (Guedin *et al.*, 2000). For the purpose of KpsD topology mapping, myc-epitopes were planned to be inserted between known residues of the predicted transmembrane region.
Figure 5.7 Localisation of KpsD in total membranes and biotin-labelling trial using DH5α (pBADEDhisMT). D and E denote KpsEDhisMT- and KpsEhisMT-expressing strains, respectively. – and + refer to unlabelled and biotin-labelled samples, respectively. Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to total protein quantities determined by Bradford assay. (A) Western blot analyses and biotin-labelling of total membranes from DH5α (pBADEDhisMT) using anti-KpsD antibody (top row) and streptavidin to detect biotin (bottom row), which were simultaneously used. Lane 1: unlabelled total membranes; lane 2: total membranes labelled with 0.2 mM biotin for 10 minutes; lane 3: total membranes labelled with 0.2 mM biotin for 30 minutes; lane 4: total membranes labelled with 1 mM biotin for 30 minutes. (B) Western blot analysis and biotin-labelling of total membranes from DH5α (pBADEDhisMT) and DH5α (pBADEhisMT) using streptavidin to detect biotin (lanes 1 and 2) and anti-KpsD antibody (lanes 3 and 4). The arrow on the biotin blot donates the predicted position of KpsD on the blot according to the results from simultaneous probing with anti-KpsD antibody. Lanes 1 and 3: total membrane of DH5α (pBADEDhisMT) labelled with 0.2 mM biotin for 10 minutes; lanes 2 and 4: total membrane of DH5α (pBADEhisMT) labelled with 0.2 mM biotin for 10 minutes.
Figure 5.8 Biotin-labelling of total membranes from DH5α (pBADED) and DH5α (pBADEDΔcys). D and DΔcys denote KpsED- and KpsEDΔcys-expressing strains, respectively. – and + refer to unlabelled and biotin-labelled samples, respectively. Western blot analyses and biotin-labelling of total membranes using anti-KpsD antibody (panels 1 and 3) and streptavidin to detect biotin (panels 2 and 3). Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to total protein quantities determined by Bradford assay. The blot was incubated with anti-KpsD antibody and streptavidin simultaneously. Lane 1: unlabelled DH5α (pBADED); lane 2: DH5α (pBADED) labelled with 0.2 mM biotin for 10 minutes; lane 3: DH5α (pBADEDΔcys) labelled with 0.2 mM biotin for 10 minutes; lane 4: unlabelled DH5α (pBADEDΔcys). Panel 1: anti-KpsD antibody; panel 2: streptavidin to detect biotin; panel 3: anti-KpsD antibody and streptavidin.
Figure 5.9 Biotin-labelling of outer membranes from DH5α (pBADED) and DH5α (pBADEDΔcys). D and DΔcys denote KpsED- and KpsDΔcys-expressing strains, respectively. − and + refer to unlabelled and biotin-labelled samples, respectively. Western blot analyses and biotin-labelling of outer membranes using anti-KpsD antibody (panels 1 and 3) and streptavidin to detect biotin (panels 2 and 3). Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to total protein quantities determined by Bradford assay. The blot was incubated with anti-KpsD antibody and streptavidin simultaneously. Lane 1: unlabelled DH5α (pBADED); lane 2: DH5α (pBADED) labelled with 0.2 mM biotin for 10 minutes; lane 3: DH5α (pBADEDΔcys) labelled with 0.2 mM biotin for 10 minutes; lane 4: unlabelled DH5α (pBADEDΔcys). Panel 1: anti-KpsD antibody; panel 2: streptavidin to detect biotin; panel 3: anti-KpsD antibody and streptavidin.
5.4.1 Generation of construct carrying C-terminally myc-tagged kpsD

Previous analyses suggested that the C-terminus of KpsD is located in the periplasm (McNulty, 2005). In order to obtain a myc-variant with a known cellular location which can be used as positive labelling control, the first myc-KpsD variant was chosen to be tagged at its C-terminus. In order to create the C-terminally myc-tagged KpsD variant, \(kpsD_{CTmyc}\) was PCR amplified using template pBADED and primers KpsD XbaI v2 F and KpsD-CT myc HindIII R (Table 2.3), tagging the myc sequence EQKLISEEDL at the C-terminus of KpsD (Figure 5.10; section 2.5.1). \(kpsD_{CTmyc}\) was cloned into pBADE, utilising restrictions sites XbaI and HindIII. The resulting plasmid was transformed into DH5\(\alpha\) and plasmids were isolated (section 2.5.5). Colony PCR, restriction digestion analysis (Figure 5.10) and DNA sequencing were used to verify the resulting plasmid pBADED\(_{CTmyc}\) (Figure 5.11).

To determine whether the C-terminally myc-tagged KpsD variant is functional, pBADED\(_{CTmyc}\) was transformed into MS\(\Delta D\) and a K5 bacteriophage assay was carried out. MS101 and MS\(\Delta D\) (pBADED) served as positive controls; MS\(\Delta D\) served as negative control. Surprisingly, pBADED\(_{CTmyc}\) was unable to complement MS\(\Delta D\), producing a plaque-negative profile identical to MS\(\Delta D\) which lacks wild-type KpsD (Table 5.1). This suggested that the C-terminal tag abolishes KpsD function as part of the K5 translocon. However, KpsD\(_{CTmyc}\) might still be used as periplasmic control for myc location in future myc-fusion analysis for KpsD topology purposes.

5.4.2 Expression and location analysis of C-terminally myc-tagged KpsD

To investigate possible reasons for the loss of function observed in C-terminally myc-tagged KpsD, whole cells of MS\(\Delta D\) (pBADED) and MS\(\Delta D\) (pBADED\(_{CTmyc}\)) were induced with 0.01% arabinose for 3 hours and analysed by Western blot using anti-KpsD antibody (Figure 5.12 A, top row, lanes 1 and 2). The signal for wild-type KpsD was slightly stronger than for KpsD\(_{CTmyc}\). This result suggested that the C-terminal myc-tag affects KpsD expression. Expression of the myc-tag in induced whole cells of MS\(\Delta D\) (pBADED\(_{CTmyc}\)) was verified by Western blot using anti-myc antibody (Figure 5.12 A, bottom row). Increasing arabinose concentration from 0.01% to 0.03% during induction resulted in KpsD\(_{CTmyc}\) expression at the level of KpsD from MS\(\Delta D\) (pBADED) (Figure 5.12 A, lane 3). Even though KpsD expression levels appear
Figure 5.10 Generation of plasmid pBADED<sub>CTmyc</sub>. Plasmid pBADED<sub>CTmyc</sub> was generated as described in sections 2.5.3-2.5.10. Samples were separated on a 1% agarose gel. Lane 1: PCR amplification of kpsD<sub>CTmyc</sub>, generating a 1.6 kb product; lane 2: colony PCR of pBADED<sub>CTmyc</sub> using pBAD28 F and R primer, generating a 2.7 kb product; lane 3: pBADED<sub>CTmyc</sub> digested with XbaI and HindIII, generating a 1.6 kb product containing kpsD<sub>CTmyc</sub>.

Figure 5.11 Illustration of the plasmid pBADED<sub>CTmyc</sub>. myc: c-myc sequence; MCS: multiple cloning site; p1: arabinose-inducible pBAD promoter; p2: AmpR promoter; term: transcriptional terminator rnmB T1; araC: arabinose transcriptional regulator; f1 ori and p15A ori: f1 and p15A origin of replication, respectively; AmpR and CmR: ampicillin and chloramphenicol resistant marker, respectively; NsiI: NsiI restriction site.
impaired by the C-terminal myc-tag, lower expression is unlikely to be the cause for the observed negative result in the K5 bacteriophage assay.

In order to determine whether KpsD\textsubscript{CTmyc} is a dominant negative mutant, which would not only cause non-functional K5 export in MS\textDelta{D} but also in the MS101 wild-type, pBADED\textsubscript{CTmyc} was transformed into MS101. The K5 bacteriophage sensitivity of MS101 (pBADED\textsubscript{CTmyc}), along with controls MS101 and MS101 (pBADED), was tested (Table 5.1). The plaque count of MS101 (pBADED\textsubscript{CTmyc}) was wild-type like, suggesting that the C-terminal myc-tag does not confer dominant negativity upon KpsD. Western blot analysis of MS101 (pBADED\textsubscript{CTmyc}) and control MS101 (pBADED) using anti-KpsD antibody showed the same difference in expression levels between KpsD and KpsD\textsubscript{CTmyc} as observed from MS\textDelta{D} host strains (Figure 5.12 B).

To investigate whether the C-terminal myc-tag affects KpsD membrane location, total and outer membranes as well as periplasmic fractions were isolated from MS\textDelta{D} (pBADED\textsubscript{CTmyc}) and control MS\textDelta{D} (pBADED) (section 2.7.4-2.7.6). Membrane and periplasmic preparations were analysed by Western blot using anti-KpsD antibody (Figure 5.13). The quality of the isolated periplasmic fractions was routinely verified by testing for the presence of the periplasmic \textit{E. coli} enzyme β-lactamase using Western blot and anti-β-lactamase antibody (data not shown). As expected, the signal of KpsD\textsubscript{CTmyc} in total and outer membranes was slightly weaker than of the KpsD wild-type, as expression levels had been observed to be lower. The KpsD\textsubscript{CTmyc} signal was enhanced above wild-type levels when total membranes were prepared from cells induced with 0.03% arabinose. The level of periplasmic KpsD and KpsD\textsubscript{CTmyc} appeared similar. These results suggest that the myc-tag does not affect membrane location of KpsD in a significant manner.
**Figure 5.12 Expression of KpsD<sub>CTmyc</sub> in whole cells.** (A) Western blot analyses of whole cells from MSΔD (pBADED) and MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.01% or 0.03% arabinose for 3 hours using anti-KpsD and anti-myc antibodies. Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised according to optical density of cells. Lane 1: MSΔD (pBADED) induced with 0.01% arabinose; lane 2: MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.01% arabinose; lane 3: MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.03% arabinose. (B) Western blot of whole cells from MS101 (pBADED) and MS101 (pBADED<sub>CTmyc</sub>) induced with 0.01% arabinose. Loadings were standardised according to optical density of cells. Lane 1: MS101 (pBADED); lane 2: MS101 (pBADED<sub>CTmyc</sub>).

**Figure 5.13 Location of KpsD<sub>CTmyc</sub> in total membranes, outer membranes and periplasm.** Western blot analyses of total membranes (top row), outer membranes (middle row) and periplasmic fraction (bottom row) from MSΔD (pBADED) and MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.01% or 0.03% arabinose for 3 hours using anti-KpsD antibody. Proteins were separated on a 10% SDS-PAGE gel prior to electrophoretic transfer. Loadings were standardised between lane 1 and 2 (according to total protein quantities determined by Bradford assay). Lane 1: MSΔD (pBADED) induced with 0.01% arabinose; lane 2: MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.01% arabinose; lane 3: MSΔD (pBADED<sub>CTmyc</sub>) induced with 0.03% arabinose.
5.5 Discussion

5.5.1 Cysteine residue 488 of KpsD is not essential for native KpsD function and K5 export

The generation of a KpsD variant lacking its single cysteine in the processed protein was a pre-requisite for cysteine-scanning mutagenesis for the topology analysis of KpsD. As expected, changing the cysteine at position 488 to a serine did not impair KpsD function within the K5 translocon. Interestingly, the C488S mutation increased the plaque count in the K5 bacteriophage assay compared to the wild-type. In the structural model of KpsD presented here (Figure 5.2), cysteine 488 is located just after the α-helix of the second sub-domain of domain 5. According to three-dimensional model of KpsD (Figure 5.3), this helix is positioned on the outside of the β-sheet motif of this sub-domain. The change from cysteine to serine might have altered the position of the helix slightly, possibly affecting the interaction of the helix and surrounding residues with the same region in adjacent protomers or with interacting proteins such as KpsE.

5.5.2 Suitability of substituted-cysteine accessibility method for topology analysis of KpsD

Substituted-cysteine accessibility method has been widely used to determine the topology of inner membrane proteins, but its application on outer membrane proteins has been less frequent. Part of this study was aimed at investigating the suitability of this method for topology analysis of KpsD. Using wild-type KpsD carrying the native cysteine at position 488 as first single cysteine variant, a protocol which aimed to label cysteines with the membrane-permeable sulphydryl-reactive reagent biotin maleimide was developed. However, it remains unclear whether cysteine-scanning mutagenesis can be used for topology mapping of KpsD, as the lack of observable biotin-labelled KpsD leaves room for interpretation towards two opposing hypotheses: cysteine 488 was labelled by biotin but the biotin-specific signal was too weak to be detected, or cysteine 488 might have not been labelled due to being inaccessible to biotin.

The second hypothesis suggests that cysteine 488 is either buried in the outer membrane or closely associated with it. According to the model presented here (Figure 5.2) and the previous experimental evidence which suggests that the region between
residues 180-588 is periplasmic (McNulty, 2005), it is very unlikely that cysteine 488 is integrated into the outer membrane. It is possible that the bottom sub-domain of domain 5, in which cysteine 488 is located, is bend back towards to the outer membrane and that the potential close proximity with the membrane results in the lack of biotin-reaction. However, if assuming that domains 3 and 4, and possibly the top sub-domain of domain 5, protrude through the periplasm towards the inner membrane in a channel-like structure, the size of these domains make it impossible for the bottom sub-domain of domain 5 to reach the outer membrane when bending backwards.

The possibility that cysteine 488 was biotin-labelled but that labelling could not be detected appears likely. However, this might mean that cysteine-scanning mutagenesis of KpsD is not a suitable method for KpsD topology mapping. During protocol development, various variations of labelling and membrane extraction conditions were trialled, while the focus was always on developing a straightforward procedure which avoids complicated steps, which might later hamper the comparability of the results when applying the protocol to a large library of cysteine mutants. Protein purification of the cysteine-variants after biotin-labelling might lead to a more unequivocal labelling-result. However, this approach would be laborious if applied to a large number of cysteine-variants and might provide non-consistent intensities of biotin signals, which would make assessing cysteine location more difficult.

Further structural analysis of KpsD, including three-dimensional analysis using approaches such as transmission electron microscopy, might shed light onto the location of cysteine 488. Testing several other single cysteine mutants with known periplasmic, surface-exposed and transmembrane locations should provide a more unequivocal answer about the suitability of cysteine-scanning mutagenesis for KpsD topology study.

5.5.3 C-terminal myc-tag affects KpsD expression levels and abolishes KpsD function and K5 export

The C-terminally myc-tagged KpsD variant was generated as a positive control for periplasmic epitope location for myc-epitope insertion analysis of KpsD. Surprisingly, tagging the C-terminus with a myc epitope negatively affected KpsD expression levels and resulted in loss of KpsD function during K5 export. This suggests that the C-
terminus of KpsD plays an important role in KpsD function and that its native state is important in order to function correctly. The myc-tag does not appear to significantly affect the outer membrane localisation of KpsD. The observation is in line with previous data which showed that a truncated variant of KpsD carrying residues 1-179 remained membrane-associated (McNulty, 2005). The C-terminal myc-tag might interfere with the folding of the periplasmic region of the protein. The tag might also affect the interaction of the C-terminus with other KpsD protomers or with other members of the K5 translocon. It would be interesting to further analyse whether the C-terminally myc-tagged variant still interacts with KpsE. This could be investigated using a co-purification experiment of his-tagged KpsE and myc-tagged KpsD, similar to the one carried out with KpsE and his-tagged KpsM in Chapter 3.

5.5.4 Future directions towards establishing the KpsD topology

The myc-epitope insertion approach for topology analysis of KpsD remains a promising approach, and future analysis of myc-variants will not only provide insights into KpsD topology but also determine further regions essential for KpsD function and possible protein-protein interaction. Future topology analysis should primarily focus on the proposed transmembrane region in order to elucidate the number of transmembrane domains as well as the number of periplasmic and surface-exposed loops. Thus, suitable myc-insertions would be carried out at residues proposed to locate to the periplasmic and surface-exposed loops (cf. Figure 5.1).

Combining the myc-epitope insertion approach with a second strategy for topology analysis would decrease the likelihood of obtaining a large number of non-functional variants unsuitable for topology study. The use of other epitope insertions such as a TEV cleavage site might be suitable approaches. TEV cleavage analysis as complementary approach for residue mapping has previously been successfully used for outer membrane proteins PapC (Kulkarni et al., 2003) and FhaC (Guedin et al., 2000). Even though TEV cleavage activity was not found to be ideal during experiments in Chapter 4, protease activity is likely to be enhanced under simpler conditions (such as the presence of a high concentration of protein and the absence of detergents).
Chapter 6

Conclusions and future directions
6.1 Conclusions

The expression of a polysaccharide capsule is common among bacteria and is associated with virulence in some pathogenic strains. To overcome the challenge of translocating high-molecular weight polysaccharides across the Gram-negative bacterial membranes, *E. coli* have developed two major export systems, the Wzy-dependent- (group 1 and 4) and the ABC transporter-dependent- (group 2 and 3) polysaccharide exporters. An expanding field within molecular microbiology research is currently concerned with investigating the mechanisms underlying these two translocation systems. Recent structural advances in group 1 research have provided valuable insights into the mechanism of the Wzy-dependent transporter systems. However, limited structural data on group 2 export complexes hamper the development of a solid model of ABC transporter-dependent translocation based on structure-function relationships of the transport proteins. This thesis aimed to advance the model of ABC-dependent polysaccharide transport in group 2 *E. coli* by using the K5 KpsM TED translocation complex as a model system.

KpsE function and oligomerisation were studied using mutagenesis analysis and size exclusion chromatography, respectively (sections 3.2, 3.3 and 4.3.1). Previous work on KpsE identified alanine 287 in a conserved APQ motif within the predicted coiled coil region as crucial for KpsE function and interaction with KpsM (Hudson, 2009). Here, site-directed mutagenesis of glutamine 289 determined that this residue is not required for KpsE function as part of the K5 translocon and is unlikely to be involved in essential interaction with other translocon proteins (section 3.2). This result highlights the need for wider mutagenesis analysis of the residues within the coiled coil region, as this region might be acting as region of interaction with translocon members KpsM and/or KpsD.

KpsE lacking the periplasmic, membrane-associated C-terminus was unable to complement wild-type KpsE as part of the K5 translocon and thus abolished polysaccharide transport (section 3.3). These data revealed the C-terminus as an essential region of KpsE. In combination with previous work which identified the importance of the N-terminal transmembrane domain (Hudson, 2009), this demonstrates that the N- and C-terminal regions are crucial for protein function. Thus, the membrane-associated C-terminus (this study), the N-terminal transmembrane
domain (Hudson, 2009), and alanine residue 287 (Hudson, 2009) are the only sites analysed so far which have been found to be essential for KpsE function.

Size exclusion chromatography of KpsE demonstrated for the first time that KpsE is able to self-associate as dimers, tetramers and possibly higher order oligomers in the absence of chemical cross-linkers, and in the absence of other capsule gene products and the K5 substrate (section 4.3.1). It is plausible to expect that KpsE acts in oligomeric states higher than a dimer if the hypothesis that KpsE forms a structure with a central cavity as observed in CtrB (Larue et al., 2011) is correct.

Previous work suggested a direct in vivo interaction between KpsE and KpsM (Hudson, 2009). The results presented in this thesis confirm this interaction. Nickel-affinity chromatography of KpsEhisMT combined with TEV proteolytic cleavage of the his-tag detected KpsM and KpsE in the same elution fraction, which indicated complex formation (section 4.2). The two proteins were also found in the same fractions from size exclusion chromatography of KpsEhisMT (section 4.3.2). Thus, results from size exclusion analysis also support the theory of a KpsE-KpsM complex. The data from size exclusion chromatography hint that KpsE and KpsM form higher order homooligomers while in a complex which each other. Western blot analysis of high molecular weight size exclusion fractions containing KpsM and KpsE revealed the co-presence of KpsT. Unless this observation is due to an artefactual association of KpsT with KpsME, data also imply that KpsMT can interact in the absence of ATP hydrolysis and K5 substrate. The latter observation would be contrary to the current model of K5 export and requires further verification.

Data from mutagenesis and co-purification analysis suggest that KpsE and KpsM are likely to interact via the C-terminal domain of KpsE (section 3.5). The absence of the C-terminus of KpsE appeared to affect the ability of hisKpsM to bind to the nickel column, suggesting a change in conformation within KpsM. Taken together, these results confirm previous work which demonstrated the in vivo interaction between KpsE and KpsM in the absence of KpsD and the K5 substrate, and supports the idea of a KpsE-KpsM pre-complex prior to K5 export.
Previous work suggested that KpsE and KpsD interact directly \textit{in vivo} and that KpsD requires the co-expression of KpsE for correct outer membrane insertion and oligomerisation (Hudson, 2009). Data from the analysis of C-terminally truncated KpsE revealed that interaction between KpsE and KpsD remains sufficient for KpsD to localise to its native membrane location and to dimerise (section 3.4). Although this does not exclude the interaction between the two proteins via the C-terminal region of KpsE, these observations suggest that the C-terminal domain is not a crucial site for KpsE-KpsD interaction. This result is in agreement with transmission electron microscopy analysis of the group 1 KpsE and KpsD equivalents Wzc and Wza which propose the large periplasmic domains of the two proteins act as interface for interaction (Collins \textit{et al.}, 2006; Collins \textit{et al.}, 2007).

The final part of this thesis attempted to construct a theoretical model of the secondary structure and topology of KpsD and investigated two strategies, substituted-cysteine accessibility method and myc-epitope insertion analysis, for their suitability for topology analysis. It was predicted that KpsD is primarily made of $\beta$-sheets with interspersed $\alpha$-helices, including a larger coiled coil region (section 5.2). In an oligomeric state, KpsD was predicted to form a N-terminal transmembrane barrel made of two $\alpha$-helices and six $\beta$-sheets per protomer in the outer membrane and reach into the periplasm via its large periplasmic domain. A protocol aimed at analysing KpsD topology by the substituted-cysteine accessibility method was developed. However, the experiments presented in this thesis provide insufficient information in order to decide whether this approach is suitable for KpsD topology analysis (section 5.3) and future tests are required. The use of myc-epitope fusions remains promising (section 5.4) and is the recommended approach for future topology studies. A C-terminal myc-epitope fusion created as a positive periplasmic control for topology analysis revealed that a tag at the C-terminus of KpsD abolishes the ability of the fusion protein to complement wild-type KpsD during K5 export (section 5.4.2). This suggests that the native integrity of the C-terminus is crucial for correct KpsD function. The reason for this observation remains unclear. However, it is possible that the tag disrupts native folding of the periplasmic domain of KpsD and might even interfere with interaction of KpsE.
6.1.1 A revised model for K5 translocation

Incorporating data from this thesis into previously established evidence, a revised theoretical model for K5 transport can be formulated. However, as various steps during transport are only hypothetical, in some cases several scenarios of the predicted steps during K5 export have been proposed.

Prior to substrate engagement and ATP hydrolysis, KpsT and KpsM might either be two separate non-interacting units (as previously proposed (Pigeon and Silver, 1997; Silver et al., 2001; Nsahlai and Silver, 2003)), or already interact as a passive complex (as size exclusion chromatography data presented here might suggest). However, it is very likely that, upon polysaccharide engagement, ATP binding by KpsT leads to a conformational change which, in turn, leads to the interaction, or to a change in interaction, with KpsM to form an active ABC transporter complex. It is predicted that the change in conformation of KpsT and KpsM is ‘transmitted’ to the resting KpsE, which subsequently changes its conformation into an active state. Evidence presented here supports the hypothesis that KpsE forms an oligomeric structure during transport and it might be suggested that KpsE forms a channel-like complex to accommodate periplasmic translocation of the polymer. As this thesis has demonstrated, KpsE-KpsM interaction involves the C-terminus of KpsE. Previous evidence and data from this thesis support the idea of a KpsM-KpsE pre-complex which exists even during the non-active state of the translocon. This assumption leads to two possible pre-complexes during transport: in the first scenario the KpsEMT complex already exists prior to substrate engagement, while in the second scenario KpsT would interact with the KpsEM pre-complex upon transport initiation. Evidence presented here suggests that KpsEM(T) are present as higher order oligomers when interacting with each other. It is predicted that a change in conformation of the KpsMTE proteins leads to the interaction of KpsE with KpsD. As previous work suggested that direct interaction of KpsE and KpsD requires expression of KpsM, the KpsE-KpsD interaction might only occur in the presence of a KpsE-KpsM complex. KpsE-KpsD interaction is likely to cause KpsD to subsequently switch into an open state which allows OM translocation of the polymer.
6.2 Future directions

Despite a growing understanding of the group 2 polysaccharide translocation process, many unanswered questions remain. Some future work directly leading on from the studies presented in this thesis has already been highlighted in the discussion sections of the relevant chapters (sections 3.6, 4.5 and 5.5). In addition, future efforts towards establishing a better picture of this complex cellular event could address some of the following areas.

Although structural similarities with other PCP proteins are likely, limited experimental data is available about the exact topology of KpsE. Cysteine-scanning mutagenesis has been successfully applied to study the topology of KpsE-homologue CtrB (Larue et al., 2011). Cysteine-scanning mutagenesis of KpsE might reveal similarities with and highlight further differences in protein structure to closely related proteins such as CtrB. Mutagenesis studies to elucidate further regions essential to KpsE function should focus on the predicted coiled coil region located in the periplasmic domain. This region is likely to be an interface during self-association and/or interaction with KpsD. A mutagenesis approach could be combined with size exclusion chromatography using KpsE and KpsEM/KpsED to establish regions essential for homo-oligomer and hetero-oligomer formation, respectively. As previously discussed (section 4.5), comparing these results from size exclusion chromatography to results from a capsule-producing strain would elucidate any changes in oligomerisation states during K5 export.

Understanding the structure and function of KpsD as the outer membrane component of the group 2 translocation is crucial. Mutagenesis studies combined with size exclusion chromatography identical to the ones described for KpsE would provide useful insight into the significance of the coiled coil region of KpsD and into KpsD oligomerisation in the absence and presence of KpsMT and the K5 polysaccharide. As correct membrane insertion and dimerisation of KpsD depends on KpsE, co-expression of KpsE is required for all experiments on KpsD.

Previous work demonstrated that ATP-binding and hydrolysis is crucial for KpsT function and K5 export, and that KpsT interacts with KpsM during export to form the ABC transporter (Pavelka et al., 1991; Pavelka et al., 1994). Data on KpsT from size
exclusion chromatography presented in this thesis hint that KpsT might also be able to interact with KpsM in the absence of ATP hydrolysis or the K5 substrate (section 4.3.2). It would be interesting to investigate whether KpsT data from size exclusion chromatography from a capsule-producing MS101 strain provide a similar or completely different KpsT profile. This analysis would therefore offer clues on whether ATP hydrolysis or substrate engagement is required for KpsT-KpsM interaction.

The work presented in this thesis highlights the challenges of over-expressing membrane proteins and how obtaining insufficient protein quantities can hamper structure-function analysis. Provided an efficient system for obtaining high quantities of protein can be found and the protein complex to be tested retains sufficient stability, three-dimensional structural analysis as well as kinetic studies of the KpsMTED complex might be possible. A first step towards establishing a three-dimensional model could attempt electron microscopy of lipid reconstituted KpsEMT. As briefly discussed earlier (section 4.5.6), kinetic information about the stoichiometry of the KpsMTED complex would greatly advance the model of group 2 polysaccharide export. Khursigara *et al.* (2005) analysed stoichiometry of two binding partners, the outer membrane protein FhuA and the inner membrane protein TonB, which are part of a siderophore transporter of *E. coli*. Using surface plasmon resonance and analytical ultracentrifugation, they were able to identify various dynamic stages of complex formations, and the use of mutants allowed them to find domains which contributed to the interactions at the various stages. A similar approach using members of the KpsMTED translocon might be feasible.

In summary, the data presented in this thesis have shed new light onto the KpsMTED transport proteins and their interaction with each other as part of the group 2 *E. coli* polysaccharide export system. This knowledge helps to develop a better understanding of the mechanisms underlying ABC transporter-dependent polysaccharide transport in Gram-negative bacteria.
References


193


Tlapak-Simmons, V. L., Kempner, E. S., Baggenstoss, B. A. and Weigel, P. H. (1998). The active streptococcal hyaluronan synthases (HASs) contain a single HAS monomer


