Molecular Characterisation of the CS5 Pilus Encoding Region from Enterotoxigenic \textit{Escherichia coli} O115:H40

Thomas George Duthy
B. Sc. (Hons.) (Adelaide)

A thesis submitted for the degree of Doctor of Philosophy

Department of Microbiology and Immunology

University of Adelaide

September, 2000
I dedicate this thesis to my mother Geisula Duthy
Abstract

Enterotoxigenic *Escherichia coli* (ETEC) possess two major virulence factors: the pilus or fimbriae that are required for attachment of the bacterium to the intestinal epithelium of the host; and the two enterotoxins, heat-labile and heat-stable, which mediate the characteristic watery diarrhoea observed during ETEC-mediated infection. To date, 21 distinct pilus types have been identified from ETEC strains pathogenic for humans. Of these, detailed molecular studies on pilus biogenesis has been carried out on a select few types. Coli surface-associated antigen 5 (CS5) is an important sub-component of the colonisation factor antigen IV (CFA/IV) expressing ETEC strains, which is present on the surface of 10-20% of all ETEC strains. Despite its obvious importance, very little is known at the molecular level regarding the genes involved in CS5 biogenesis and how the final pilus structure is assembled. This thesis specifically identifies the genes involved in biogenesis and their function, as well as providing a suitable model for assembly of the CS5 pilus.

The region required for biosynthesis of CS5 pili consisted of six *coli*-surface-factor-five (*csf*) genes, denoted *csfA-csfF*. The *csfA* gene has been shown by Clark *et al.* (1992) to encode the major subunit. Homology searches showed that *csfC* was similar to a number of outer membrane usher proteins, including CooC of CS1 pili. For the remaining *csf* genes, no other amino acid homologies to other known pilus systems were identified; however, *csfB* was very similar both in sequence and structural terms to *csfF* and likewise *csfD* showed some protein similarity with *csfA*. The region was flanked by a number of distinct insertion sequence (IS) elements, including a complete IS1 and a defective IS30, along with sequences homologous to IS911, IS66 and orfB from Tn21. Protein expression studies of the *csf* genes showed that each was capable of encoding a protein product in approximate agreement with the predicted molecular masses, with the exception of CsfC which migrated faster than predicted. Expression of the *csf* gene cluster without flanking IS sequences from ColE1-based plasmids in *E. coli* K-12 produced large amounts of peritrichous CS5 pili, which served as a
suitable system for mutagenesis studies and functional analysis of each of the newly identified csf genes.

Two minor pilins were identified in this thesis, which were absolutely required for pilus initiation and termination. Firstly, examination of pilus biogenesis in a csfD deletion strain showed that CsfD was required for pilus initiation and did not function as a specific chaperone for the major subunit. Western analysis on purified CS5 pili showed that CsfD was an integral minor subunit, constituting between 1.2-1.8% of the pilus. The association of CsfD within the pilus, composed primarily of CsfA, was a stable interaction, and could only be liberated upon boiling of the pili. Direct localisation of CsfD within the pilus by immunogold electron microscopy was subjective due to bundling of the pili. Ultrastructure analysis of purified CS5 resolved the pilus morphology to be a flexible 2 nm fibrillae, with CsfD predicted to localise at intervals along the shaft and at the distal tip. The association of CsfE and termination of pilus assembly was shown directly by constructing a deletion in the csfE gene, which resulted in extended pilus morphologies. Complementation with a wild type copy of the csfE gene on a number of different expression plasmids was able to shorten the mean length of the pili, although CsfE was not rate-limiting for termination of pilus assembly. Fractionation studies showed that CsfE localised in the outer membranes and was not detected in crude or purified CS5 pili.

The csfC gene was shown to act as an outer membrane usher protein for CS5 biogenesis, since CsfC was localised in the outer membranes by fractionation experiments and a deletion in the csfC gene resulted in the complete absence of pilin subunits in the supernatant of this strain. The specificity of the CsfC usher for the pilin subunits was demonstrated by the inability of homologous ushers from the CS1 and CFA/I pilus systems to translocate and indeed assemble CS5 subunits to form cell surface pili.

Two chaperone-like proteins were identified which possessed specificity for either the major subunit CsfA or the minor subunits CsfD and CsfE. CsfB was shown to specifically stabilise the major subunit CsfA, since a deletion in the csfB gene resulted in the complete
absence of CsfA in periplasmic fractions prepared from the strains. Complementation experiments with a wild type copy of the csfB gene was able to restore the accumulation of CsfA in the periplasm and cell surface CS5 pili expression, as judged by immunogold electron microscopy. Fractionation studies also indicated that CsfB was located in the periplasm of the bacteria and was not associated with the final pilus structure.

An in-frame deletion in the csfF gene was shown to result in an extended pilus morphology by immunogold electron microscopy, and such pili contained no detectable levels of CsfD by western immunoblot analysis. CsfF was shown to stabilise CsfD in the periplasm of the bacteria by constructing a deletion in both csfF and csfB genes and re-introducing a wild type copy of the csfF gene back into the system. CsfF was shown to be rate-limiting in terminating pilus assembly, since modulating CsfF expression resulted in significantly shorter pili versus the wild type strain. Therefore, CsfF is predicted to deliver CsfE to the outer membrane to terminate assembly, and the rate of incorporation of CsfE is dependent on CsfF levels in the periplasm.

A model for CS5 pilus biosynthesis has been proposed in this thesis. Initiation of pilus biogenesis occurs when a CsfD-CsfF complex is delivered to the CsfC outer membrane usher protein, which results in the translocation of the CsfD minor pilin subunit across the outer membrane. This event marks a conformational change in CsfC, thereby allowing CsfA-CsfB complexes to be delivered and CsfA major subunits translocated across the outer membrane and polymerised into the growing pilus. Further CsfD-CsfF complexes are also incorporated into the pilus at a rate determined by the stoichiometric ratio of CsfA and CsfD in the periplasm. Termination of pilus assembly occurs when a CsfE-CsfF complex is delivered to CsfC, resulting in the incorporation of CsfE, which prevents further polymerisation of the pilus.
Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any University or other tertiary institution and to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited into the University library, being made available for loan and photocopying.

Thomas George Duthy, September 2000
Acknowledgements

I would firstly like to extend my sincere gratitude to Professor Paul Manning for giving me the opportunity to undertake this work in his laboratory. I would also like to thank him for maintaining such a vested interest in my work, despite being thousands of kilometres away in Boston for the last 3 years. My deepest thanks to Dr Mike Heuzenroeder who took over primary supervision of this work after the departure of Paul, and who has provided me with the support and reassurance that all PhD students crave. I would also like to thank him for teaching me how to drink beer “in stereo”.

I acknowledge the financial support of Northfield Laboratories Pty. Ltd. over the last 3 years as part of the University of Adelaide-Northfield Laboratories scholarship awarded to me, and in particular to Dr Barry Combs (pronounced Coombs) with whom I have had a close working relationship with over the last 3 ½ years. I also wish to thank Professor John Heuser and Dr Robyn Roth from the Washington University School of Medicine, St Louis, Missouri for the excellent electron micrographs of deep-etch freeze-fractured CS5 pili.

From the Department of Microbiology and Immunology I extend my gratitude to Dr Lothar Staendner who was a post-doctoral researcher on ETEC pilus systems, Drs Steve Attridge, Renato Morona and Tony Focareta for many useful discussions (shares or otherwise !), and fellow PhD’ers Jamie Botten (Mr Byte), Chris V (you know what the V stands for) and Leanne Purins (nee Blakeney)- good luck in the future, and to all past and present members of the MPU, I extend my thanks. I acknowledge Drs Maria Mavris, Monica Ogierman and Uwe Stroeher in the earlier years of my PhD, they are loads of laughs and very good scientists too. Thanks also to Chris Clark, who knows more about pilus biogenesis than anyone, for many useful discussions and advice. To my lab mate for the last 3 years Chris Cursaro: thanks for putting up with Triple J, my complaining, my singing and ‘alternative’ sense of humour.

Finally I wish to thank my mother and father for all their years of support and belief in me and this thing called a PhD- thankyou so much, it will be worth all your sacrifices I’m sure ! Last but not least I wish to thank Rachel Kohler, who is the kindest, most sincere person I know (and who happens to be doing a PhD herself !)- thanks for EVERYTHING Rache (Cokey).
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>µCi</td>
<td>microCurie</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µl</td>
<td>microlitre</td>
</tr>
<tr>
<td>µm</td>
<td>micrometre</td>
</tr>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</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>cytosine</td>
</tr>
<tr>
<td>CFA</td>
<td>colonisation factor antigen</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming unit</td>
</tr>
<tr>
<td>Ci</td>
<td>Curie</td>
</tr>
<tr>
<td>CS</td>
<td>coli surface associated antigen</td>
</tr>
<tr>
<td>csf</td>
<td>coli surface factor five</td>
</tr>
<tr>
<td>cyto</td>
<td>cytoplasm</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxy adenosine nucleoside triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxy cytosine nucleoside triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxy guanosine nucleoside triphosphate</td>
</tr>
<tr>
<td>dH₂O</td>
<td>distilled water</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded DNA</td>
</tr>
<tr>
<td>DTT</td>
<td>diothiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxy thymidine nucleoside triphosphate</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>ECL</td>
<td>enhanced chemiluminescence</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene-diamine-tetra-acetic-acid</td>
</tr>
<tr>
<td>ETEC</td>
<td>enterotoxigenic Escherichia coli</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>g/l</td>
<td>gram per litre</td>
</tr>
<tr>
<td>hr</td>
<td>hour(s)</td>
</tr>
<tr>
<td>i/b</td>
<td>inclusion bodies</td>
</tr>
<tr>
<td>IM, i/m</td>
<td>inner (cytoplasmic) membrane</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalacto-pyranoside</td>
</tr>
<tr>
<td>Kan</td>
<td>kanomycin</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>Klenow</td>
<td>Klenow fragment of <em>E. coli</em> DNA polymerase I</td>
</tr>
<tr>
<td>kV</td>
<td>kilovolts</td>
</tr>
<tr>
<td>LA</td>
<td>luria agar</td>
</tr>
<tr>
<td>LacZ</td>
<td>β-galactosidase</td>
</tr>
<tr>
<td>LT</td>
<td>heat-labile enterotoxin</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mA</td>
<td>milliamps</td>
</tr>
<tr>
<td>MBP</td>
<td>maltose binding protein</td>
</tr>
<tr>
<td>MDa</td>
<td>megadalton</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>ml</td>
<td>millilitre</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>MQ</td>
<td>milli-Q water</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NA</td>
<td>nutrient agar</td>
</tr>
<tr>
<td>NB</td>
<td>nutrient broth</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
</tr>
<tr>
<td>nm</td>
<td>nanometre</td>
</tr>
<tr>
<td>no.</td>
<td>number</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>°C</td>
<td>degrees celsius</td>
</tr>
<tr>
<td>OD_{600nm}</td>
<td>optical density at 600 nm</td>
</tr>
<tr>
<td>oligo</td>
<td>oligonucleotide</td>
</tr>
<tr>
<td>OM, o/m</td>
<td>outer membrane</td>
</tr>
<tr>
<td>OMP</td>
<td>outer membrane protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>P</td>
<td>probability</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Pap</td>
<td>pili associated with pyelonephritis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol-8000</td>
</tr>
<tr>
<td>peri</td>
<td>periplasm</td>
</tr>
<tr>
<td>PHD</td>
<td>profile network prediction Heidelberg</td>
</tr>
<tr>
<td>pmol</td>
<td>picomole</td>
</tr>
<tr>
<td>R</td>
<td>resistant</td>
</tr>
<tr>
<td>Rif</td>
<td>rifampicin</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SAP</td>
<td>shrimp alkaline phosphatase</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>sec</td>
<td>second(s)</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of the mean</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single stranded DNA</td>
</tr>
<tr>
<td>ST</td>
<td>heat-stable enterotoxin</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>TBS</td>
<td>tris-buffered saline</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N',N'-Tetramethyl-ethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris (hydroxymethyl) amino methane</td>
</tr>
<tr>
<td>TTBS</td>
<td>Tween-20 plus tris-buffered saline</td>
</tr>
<tr>
<td>U</td>
<td>unit(s)</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>W</td>
<td>watt</td>
</tr>
<tr>
<td>w/c</td>
<td>whole cells</td>
</tr>
<tr>
<td>w/m</td>
<td>whole membranes</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
## Table of Contents

Chapter 1: General Introduction

1.1 Introduction ...........................................................................................................1

1.2 Enterotoxigenic *Escherichia coli* (ETEC) ..........................................................2

1.3 ETEC Enterotoxins ...............................................................................................3
   1.3.1 LT Enterotoxin ...............................................................................................4
      1.3.1.1 Classification of LT Enterotoxins .........................................................4
      1.3.1.2 The Receptor and Structure of LT .....................................................4
      1.3.1.3 Release and Mechanism of Action ....................................................6
      1.3.1.4 Adjuvant Properties of LT and CT ....................................................7
   1.3.2 ST Enterotoxins ...............................................................................................9
      1.3.2.1 Classification of ST Enterotoxins .........................................................9
      1.3.2.2 Receptor and Structure of ST Enterotoxins .......................................10
      1.3.2.3 Mechanism of Action ........................................................................11
   1.4 *E. coli* Pili .........................................................................................................12
      1.4.1 Type I Pili .................................................................................................13
      1.4.2 Curli ..........................................................................................................14

1.5 Pili of Enterotoxigenic *Escherichia coli* ............................................................14

1.6 Animal ETEC Colonisation Factors ....................................................................15
   1.6.1 K88 ..............................................................................................................16
   1.6.2 K99 ..............................................................................................................17
   1.6.3 987P ............................................................................................................18

1.7 Human ETEC Colonisation Factors ....................................................................18
   1.7.1 Identification, Morphology and Genetics ..................................................19
Chapter 1: Regulation of Expression

1.8.2 Regulation of Expression ................................................................. 41
1.8.3 The PapG Adhesin ............................................................. 42
1.8.4 The Pap Pilins- PapA, PapE, PapK, PapF and PapH .................................. 43
1.8.5 PapD ................................................................. 46
1.8.6 PapC ................................................................. 49

Chapter 1: Vaccines Against Human ETEC Infection

1.9 Vaccines Against Human ETEC Infection .......................................................... 51
1.9.1 Inactivated Whole Cell and Toxoid Vaccines ........................................ 52
1.9.2 Purified Pili Vaccines ........................................................................... 53
1.9.3 Live Oral Vaccines ............................................................................. 54
1.9.4 Passive Vaccines ................................................................................ 55

Chapter 1: Aims of this Research ........................................................................ 56

Chapter 2: Materials and Methods

2.1 Chemicals and Reagents ........................................................................ 57
2.2 Enzymes ................................................................................................. 58
2.3 Bacterial Strains and Plasmids ................................................................. 58
2.4 Growth Media .......................................................................................... 58
2.5 Maintenance of Bacterial Strains ............................................................... 59
2.6 Synthesis of Oligonucleotides .................................................................. 60
2.7 Transformation of E. coli K-12 ................................................................. 60

2.7.1 Preparation of Super Competent Cells ................................................ 60
2.7.2 Transformation Procedure .................................................................. 61

2.8 Plasmid Preparation Procedures .................................................................. 61

2.8.1 Small Scale Preparation ...................................................................... 61
2.8.2 Medium Scale Preparation for Sequence Analysis .................................. 62
2.9 Analysis and Manipulation of DNA

2.9.1 DNA Quantitation

2.9.2 Restriction Endonuclease Digestion of DNA

2.9.3 Analytical and Preparative Separation of Restriction Fragments

2.9.4 Calculation of Restriction Fragment Size

2.9.5 Purification of DNA Fragments

2.9.6 Dephosphorylation of DNA

2.9.7 Ligation of DNA Fragments

2.9.8 DNA Cloning Procedures

2.9.8.1 In vitro Cloning

2.9.8.2 PCR Screening

2.9.8.3 Generation of Step-Wise Deletions

2.10 Polymerase Chain Reaction (PCR) Protocol

2.10.1 Standard PCR Reaction

2.10.2 Cloning of PCR Products

2.11 DNA Sequencing Protocols

2.11.1 Sequencing Using Dye-Labelled Primers

2.11.2 Sequencing with Dye-Labelled Terminators

2.11.3 Electrophoresis and Analysis of DNA Sequences

2.12 Protein Analysis

2.12.1 SDS-Polyacrylamide Gel Electrophoresis

2.12.2 Western Immunoblot Analysis

2.12.3 Protein Expression Systems

2.12.3.1 Standard T7 Polymerase-Mediated Expression

2.12.3.2 L-[³⁵S]-Methionine Labelling of Over-expressed Proteins

2.12.3.3 (Histidine)₆ Tagged Protein Expression
2.13 Autoradiography .................................................. 73
2.14 Cell Fractionation ................................................. 73
2.15 N-Terminal Sequence Analysis ................................ 74
2.16 Polyclonal Antisera Preparation ................................ 74
   2.16.1 Protein Preparation and Immunisation Schedule ...... 74
   2.16.2 Absorption .................................................. 75
   2.16.3 Affinity Purification of Antisera .......................... 75
2.17 Pili Preparation .................................................. 76
   2.17.1 Crude Pili Preparation Method 1 ......................... 76
   2.17.2 Crude Pili Preparation Method 2 ......................... 76
   2.17.3 Caesium Chloride (CsCl) Purification of CS5 Pili ....... 77
2.18 Slide Agglutination .............................................. 78
2.19 Mannose Resistant Haemagglutination (MRHA) .......... 78
2.20 Immunogold Electron Microscopy .............................. 78

Chapter 3: Sequence Analysis and the Initial Characterisation of the CS5 Pilus Encoding Region from Enterotoxigenic E. coli O115:H40

3.1 Introduction ...................................................... 80
3.2 Results ............................................................. 80
   3.2.1 Sequencing of the CS5 Region from Enterotoxigenic E. coli O115:H40 .... 81
      3.2.1.1 Sub-cloning of the CS5 Region ......................... 81
      3.2.1.2 Generation of Nested Deletions from pPM5303 and pPM5302 and DNA Sequencing .............. 81
   3.2.2 Analysis of the DNA Sequence Flanking the csf Cluster .......... 82
   3.2.3 Analysis of the DNA Sequence Constituting the csf Gene Cluster .... 84
3.2.3.1 Transcriptional Terminators......................................................... 86

3.2.4 Homologies................................................................................... 86

3.2.4.1 CsfC is Homologous to the CooC Membrane Usher Protein of CS1 Pili...... 86

3.2.4.2 CsfB is Homologous to CsfF.......................................................... 87

3.2.4.3 CsfA Shares Weak Homology with CsfD......................................... 88

3.2.5 Construction of a Suitable CS5 Pili Expression System......................... 88

3.2.5.1 Re-Cloning of the csf Region without Flanking IS Elements............... 88

3.2.5.2 Expression of pPM5631 in E. coli K-12......................................... 88

3.2.6 Protein Expression Studies............................................................. 90

3.2.6.1 Cloning of Individual csf Genes Behind the T7 Promoter.................. 90

3.2.6.2 Specific Expression of CsfB, CsfC, CsfD, CsfE and CsfF Proteins......... 91

3.3 Summary and Conclusions.................................................................. 92

Chapter 4: The Role of CsfD in CS5 Pilus Biogenesis

4.1 Introduction....................................................................................... 94

4.2 Results.............................................................................................. 95

4.2.1 Construction of a csfD Deletion Mutant.......................................... 95

4.2.2 Construction of a csfD Complementing Plasmid.................................. 95

4.2.3 The Role of CsfD in CS5 Pili Biogenesis......................................... 96

4.2.3.1 Analysis of Pilus Assembly.......................................................... 96

4.2.3.2 Stability of CsfA Subunits in the Periplasm.................................... 98

4.2.3.3 Cell Fractionation......................................................................... 98

4.2.3.4 Association of CsfD with the CS5 Pilus....................................... 99

4.2.3.5 The Ratio of CsfD to CsfA in the CS5 Pilus................................... 99

4.2.4 Generation of Specific Antiserum to CsfD and N-terminal Sequence Analysis... 100
4.2.4.1 Specificity of Anti-CsfD Antibodies in Immunogold Electron Microscopy...101

4.2.5 Ultrastructure Analysis of CS5 Pili by Quick-Freeze Deep-Etch Electron Microscopy .................................................................102

4.2.6 Secretion of CsfD in the Absence of CsfA........................................103

4.3 Summary and Conclusions.................................................................104

Chapter 5: The Role of CsfE in CS5 Pilus Biogenesis

5.1 Introduction.......................................................................................107

5.2 Results..............................................................................................108

5.2.1 Construction of an In-Frame Deletion in the csfE Gene..................108

5.2.2 Construction of a csfE Complementing Plasmid............................109

5.2.3 Effect of a Deletion in csfE on CS5 Pilus Biosynthesis...................109

5.2.3.1 Examination of Pilus Morphology and Function........................109

5.2.3.2 Periplasmic and Whole Cell Analysis........................................111

5.2.4 Specific Modulation of CsfE with PBAD Promoters.......................111

5.2.4.1 Construction of csfE in pBAD18-Kan........................................112

5.2.4.2 The Effect on CS5 Pilus Biogenesis from the Specific Modulation of CsfE .................................................................112

5.2.4.3 Immunogold Electron Microscopy and Quantitative Analysis of Pili Numbers and Lengths..........................................................113

5.2.5 Over-Expression of CsfE...............................................................115

5.2.6 Generation of Specific Antibodies Against CsfE and Cellular Location ..........................................................115

5.2.7 The Association of CsfE with CS5 Pili...........................................116

5.3 Summary and Conclusions...............................................................117
Chapter 6: The Role of CsfC in the Assembly of CS5 Pili

6.1 Introduction ........................................................................................................................................... 119

6.2 Results .................................................................................................................................................... 120

6.2.1 Construction of an In-Frame Deletion in the csfC Gene .............................................................. 120

6.2.2 Construction of a csfC Complementing Plasmid ........................................................................... 121

6.2.3 The Effect of a csfC Deletion on CS5 Pilus Biosynthesis ............................................................ 121

6.2.3.1 Examination of CS5 Pilus Assembly ....................................................................................... 122

6.2.3.2 Periplasmic Analysis ............................................................................................................. 123

6.2.4 Localisation of CsfC in E. coli K-12 ............................................................................................. 123

6.2.5 Complementation Experiments with Homologous Outer Membrane Assembly Proteins .... 124

8.7 Summary and Conclusions .................................................................................................................. 126

Chapter 7: Characterisation of the Homologous Proteins Csfb and CsfF

7.1 Introduction ............................................................................................................................................. 128

7.2 Results .................................................................................................................................................... 129

7.2.1 Construction of an In-Frame Deletion in the csfF Gene .............................................................. 129

7.2.2 Construction of a csfF Complementing Plasmid ......................................................................... 130

7.2.3 The Effect of a csfF Deletion on Biogenesis of CS5 pili ............................................................. 130

7.2.3.1 Initial Phenotypic Characterisation ....................................................................................... 130

7.2.3.2 Periplasmic, Heat Extract and Cell Supernatant Analysis .................................................. 131

7.2.4 Specific Modulation of csfF Expression ........................................................................................ 133

7.2.5 Construction of an In-Frame Deletion in the csfB Gene .............................................................. 134

7.2.6 Construction of a csfB Complementing Plasmid ......................................................................... 135

7.2.7 The Role of Csfb in CS5 Pilus Biogenesis ...................................................................................... 135
References ........................................................................................................ 166

Appendix I

Appendix II
1.1 Introduction

*Escherichia coli* (*E. coli*) is a normal inhabitant of the intestinal tract of humans and animals, with colonisation occurring shortly after birth. *E. coli* then remains as part of the normal faecal flora. Since its discovery by Theodor Escherich in 1885, *E. coli* has been extensively studied, culminating in the complete genome sequence of the common *E. coli* strain K-12 being reported in 1997 (Blattner *et al.*, 1997). It has been said that we know more about the biology of *E. coli* than we do about ourselves.

Pathogenic *E. coli* differ from other *E. coli* by the ability to colonise host tissues where non-pathogenic strains usually do not survive (Finlay and Falkow, 1989), and are limited to a certain number of O:H serotypes (Ørskov *et al.*, 1977).

There are eight categories of pathogenic *E. coli* which cause intestinal, urinary, septicaemia and meningitis type infections:

Enterotoxigenic *E. coli* (ETEC) which is a major cause of infantile diarrhoea in developing countries and traveller’s diarrhoea.

Enterohaemorrhagic *E. coli* (EHEC) are a cause of haemorrhagic colitis and haemolytic uraemic syndrome (HUS).

Enteropathogenic *E. coli* (EPEC) is involved in infantile diarrhoea, although EPEC have declined in the developed world as major causes of infantile diarrhoea, while still remaining very important in the developing world.

Enteroinvasive *E. coli* (EIEC) comprise only a small proportion of the diarrhoeagenic organisms found in non-tropical countries; however, they are an important cause of morbidity.
and mortality, being a common cause of medical consultation for young children. They are also important causes of dysentery-like diarrhoea in tropical countries.

Enteroaggregative \textit{E. coli} (EAggEC) exhibit enteroaggregative adherence on tissue culture cells and are mainly associated with persistent childhood diarrhoea, with diarrhoeal episodes that can last more than 14 days.

Localised adherent \textit{E. coli} (LAEC) which form adherent microcolonies on HEp-2 cells, have been associated with acute non-bloody diarrhoea in children.

Uropathogenic \textit{E. coli} (UPEC) is by far the most common cause of urinary tract infections (UTI) and can range in severity from asymptomatic through to cystitis and pyelonephritis. Women are far more frequently affected than men.

Neonatal meningitis \textit{E. coli} (NMEC) are a major cause (up to 80%) of neonatal meningitis and around 80% of the isolates possess the K1 capsular antigen, which is chemically and immunologically identical to the group B acidic polysaccharide of \textit{Neisseria meningitidis}.

This thesis involves the sequence analysis, characterisation and proposed mechanism of biogenesis of the CS5 (coli surface-associated antigen 5) pili sub-component from CFA/IV expressing human enterotoxigenic \textit{Escherichia coli} (ETEC) serotype O115:H40. An overall description of ETEC toxins and pili is given, along with a model system representing the biogenesis of \textit{E. coli} pili based on studies of Pap pili of uropathogenic \textit{E. coli}. Furthermore, a discussion is presented on receptor identification, distribution and antigenicity of human ETEC pili, including the current development of vaccines against ETEC.

1.2 Enterotoxigenic \textit{Escherichia coli} (ETEC)

ETEC is the most common cause of diarrhoea in children under five years of age in developing countries and in travellers to these areas (Black, 1990; Black, 1993). Worldwide incidence of ETEC infection is upwards of 650 million cases of diarrhoea and 800,000 deaths of children under the age of five (Gaastra and Svennerholm, 1996). In patients who
experience diarrhoea whilst travelling to developing countries, over half of all cases is caused by ETEC, while it has been estimated that of the 16 million people travelling to these regions annually, 40-70% contract ETEC-related diarrhoea (Black, 1990; Gorbach and Edelman, 1986). Traveller's diarrhoea has gone under a variety of regional and descriptive titles, such as gypsy tummy, Aden gut, Bali belly, Delhi belly, Montezuma's revenge, Hong Kong dog, Matlab monsoon and the Aztec two-step. Besides a major concern for human health, ETEC cause severe diarrhoea in neonatal and young herd animals such as calves and pigs. However, the serotypes and pili types found on such ETEC are distinct from human ETEC strains, and are unable to mediate specific colonisation or diarrhoea in human subjects, indicating species-specific characteristics (DuPont et al., 1971).

There are two virulence factors involved in the pathogenesis of ETEC, namely pili (also known as fimbriae) and the two enterotoxins, heat-labile (LT) and heat-stable (ST). Bacterial colonisation is mediated by the specific interaction of the pili with host cell receptors, which brings the ETEC bacteria into close proximity to the brush border membranes of the intestinal epithelium with the enterotoxins then released. The ETEC enterotoxins mediate the fluid accumulation in the lumen, resulting in the watery diarrhoea observed in cases of ETEC mediated disease.

1.3 ETEC Enterotoxins

ETEC strains from both human and animal origin are capable of producing either ST or LT enterotoxins, or both. Collections of human ETEC strains from diverse geographical areas indicate that, in general, 40% of the strains are LT⁺ST⁺, about 40% are LT⁻ST⁺ and the remaining 20% are LT⁺ ST⁻ (Gross and Rowe, 1985; Levine, 1987).
1.3.1 LT Enterotoxin

LT is an immunoactive, multimeric protein consisting of a single enzymatic A subunit of 28 kDa (LT-A) and a pentameric B subunit (LT-B) consisting of five identical monomers of 11.6 kDa each (Gill et al., 1981). The LT holotoxin is structurally, immunologically and physiologically related to the cholera enterotoxin (CT) produced by Vibrio cholerae (Clements et al., 1980; Dallas and Falkow, 1980).

1.3.1.1 Classification of LT Enterotoxins

LT toxins are classified into two serogroups designated LT-I and LT-II, with several antigenic variants within each group. Serogroup one consists of type I enterotoxin (LT-I) including LTh-I and LTp-I produced by human and porcine strains respectively, CT and related enterotoxins from other Gram negative bacteria (Calva et al., 1989; Holmes and Twiddy, 1983; Prasad et al., 1992; Prasad et al., 1990). LTh-I and LTp-I are plasmid-encoded and share 95% sequence identity, while the codon usage and G+C content is suggestive that these LT genes were derived from Vibrio cholerae (Gyles, 1974; Yamamoto et al., 1984; Yamamoto and Yokota, 1983).

LT-II was initially identified on the basis of altered antigenicity to LT-I, since it is not neutralised by anti-CT or anti-LT-I antisera (Green et al., 1983). Two antigenic types of LT-II have since been identified and are known as LT-IIa and LT-IIb (Guth et al., 1986). LT-IIa and LT-IIb are structurally similar to LT-I, consisting of A and B polypeptides similar in size to LT-I (Guth et al., 1986). Unlike LT-I, the structural genes for LT-II are chromosomally encoded and only 56% nucleotide sequence identity is shared between the LT-IIa and LTh-I A subunit, with no significant homology between the B subunits (Pickett et al., 1989).

1.3.1.2 The Receptor and Structure of LT

The receptor for both LT-I and CT has been identified as galactose residues on galactosyl-N-acetylgalactosaminyl-[N-acetylneuraminyl]-galactosylglucosylceramide, denoted
as ganglioside GM₁, which is present across the brush border membranes of the intestinal epithelium, but also on many other eukaryotic cells (King and Van Heyningen, 1973; Merritt et al., 1994; Wisnieski and Bramhall, 1981). LT-I is also able to interact with receptors not recognised by CT. These include weak binding to GM₂ and asialo-GM₁ (Fukuta et al., 1988). LT-II does not bind efficiently to GM₁, but binds best to ganglioside GD₁b (LT-IIa) or ganglioside GD₁a (LT-IIb) (Fukuta et al., 1988). The pentameric B-subunits mediate the specific binding of the LT-I holotoxin to GM₁.

Pre-LT-A and Pre-LT-B are formed in the cytoplasm and transported into the periplasm to form A and B subunits, which are then assembled into holotoxin molecules (Hofstra and Witholt, 1985). The process of holotoxin formation and stability is dependent on a specific interaction between the A and B subunits (Streatfield et al., 1992).

The crystal structure of the LTp-I enterotoxin revealed an exposed flexible loop comprising residues 50-64 of the B subunit (Sixma et al., 1991). The importance of this loop in immune recognition and neutralisation was shown from studies involving the generation of a synthetic peptide encompassing this region being able to elicit protective antibodies against both LT and the closely related CT (Jacob et al., 1986; Jacob et al., 1985; Jacob et al., 1984). Essentially, the LT enterotoxin consists of a doughnut shaped B pentamer bound to the enzymatic A subunit (Sixma et al., 1991). The A polypeptide has two domains, a globular A₁ domain that extends outwards from one face of the pentamer and an A₂ domain with an extended carboxy-terminal tail that penetrates through the highly charged central pore of the B pentamer (Sixma et al., 1991).

The crystal structure of an LT:galactose complex at 2.2 Å resolution showed that residues Glu-51, Gln-56, His-57 and Gln-61 of the B subunit are involved in the specific binding of galactose, which forms part of the oligosaccharide portion of GM₁ (Merritt et al., 1994). The exposed loop exhibited flexibility in solution, allowing entry of the receptor into the binding site. The loop then becomes tightly ordered to secure the saccharide during binding (Merritt et al., 1994). Mutation of Ala-64 to Val-64 was found to interfere with
pentamer formation (Iida et al., 1989). This was confirmed further by the elucidation of the crystal structure of LT, where residue Ala-64 was found to be in contact with the completely buried Met-31 of the next subunit (Sixma et al., 1991). Structural analysis also revealed the importance of Trp-88 and Gly-33 in forming the receptor binding domain for LT. This is supported by mutational analysis of these two residues that resulted in the abolition of receptor binding (Ludwig et al., 1985; Sixma et al., 1991; Tsuji et al., 1985).

The A subunit is synthesised as a single polypeptide in E. coli. In contrast, CT-A is proteolytically nicked during secretion from Vibrio cholerae, by the soluble haemagglutinin/protease which gives rise to two polypeptides, A1 and A2 which are covalently linked by a single disulphide bond (Booth et al., 1984; Gill and Rappaport, 1979). However, in vitro treatment of LT with trypsin results in the cleavage of the surface exposed loop between A1 and A2 to produce a nicked toxin with the A1 and A2 fragments joined by a single disulphide bond (Clements et al., 1980). The inability in E. coli to nick and secrete the LT toxin could explain the reduced severity of disease when compared to cholera (Pearson and Mekalanos, 1982).

1.3.1.3 Release and Mechanism of Action

Unlike CT from Vibrio cholerae in which the toxin is actively secreted, LT remains in the periplasm in ETEC strains (Hirst et al., 1984a). Leakage of LT from the periplasm of bacterial cells can be induced by polymyxin B (Evans et al., 1974), urea (Qu et al., 1991) and host intestinal factors such as bile salts at physiological concentrations (Hunt and Hardy, 1991). It has been shown that physiologically active LT is associated with outer membrane vesicles released from ETEC strains (Horstman and Kuehn, 2000). The LT holotoxin is located both inside and on the exterior of the vesicles, and can be removed under dissociating conditions (Horstman and Kuehn, 2000). The LT holotoxin and the B subunit can be efficiently secreted from Vibrio strains. However, when a plasmid-encoding for CT is introduced into E. coli, the CT holotoxin is sequestered in the periplasm of these strains (Hirst
et al., 1984b; Michel et al., 1995; Pearson and Mekalanos, 1982). A cluster of genes in *Vibrio cholerae* has been identified that is required for extracellular secretion of this enterotoxin (Overbye et al., 1993; Sandkvist et al., 1993).

The binding of the LT holotoxin to the GM₁ receptor on the intestinal epithelial cell facilitates the membrane translocation of the A subunit through the cell membrane and a further translocation from the apical to the basolateral surface of the epithelial cell. Studies suggest that translocation of the A subunit occurs from an endosomal compartment after endocytosis (Lencer et al., 1995a; Lencer et al., 1995b). In the eukaryotic cell, the enzymatic activity of the A₁ fragment is dependent on the reduction of the disulphide bond between Cys-187 and Cys-189 linking A₁ to A₂ (Gill and Rappaport, 1979; Mekalanos et al., 1979; Tomasi and Montecucco, 1981). The identity of the endogenous reducing agent and mechanism of reduction is not yet known, but reduction of the disulphide bond is required for the enzymatic activity of A₁. Proteolytic nicking of the A polypeptide can enhance, but is not absolutely required for, enzymatic activity and toxicity (Grant et al., 1994). Crystallographic data indicates that cleavage does not result in a conformational change in A₁ (Merritt et al., 1994).

A₁ catalyses the ADP-ribosylation of the stimulatory GTP-binding protein in the adenylate cyclase enzyme complex on the basolateral surface of the epithelial cells, resulting in an increase in cyclic AMP (cAMP) production (Cassel and Selinger, 1977). This leads to the inhibition of absorption of sodium chloride and increased water efflux from intestinal crypt cells, thereby producing watery diarrhoea.

### 1.3.1.4 Adjuvant Properties of LT and CT

The mucosal adjuvant activities of the enterotoxins LT and CT are well known (Holmgren et al., 1994; Rappuoli et al., 1999; Snider, 1995). LT is a powerful mucosal immunogen and hence has been used as an oral adjuvant for the enhancement of mucosal and serum responses to co-administered antigens (Clements et al., 1988). Epitope mapping has
revealed that the pentameric LT-B contains the majority of the immunodominant epitopes resulting in a specific LT antibody response (Belisle et al., 1984).

Receptor binding is an essential part of the adjuvant properties of both LT and LT-B subunits (Guidry et al., 1997; Nashar et al., 1996). A non-toxic mutant which lacks ADP-ribosylating activity but retains holotoxin formation, due to a single amino acid substitution from Arg-7 to Lys-7 of the A subunit, was able to bind eukaryotic cells and act as a mucosal adjuvant for co-administered proteins (Douce et al., 1995), indicating that toxicity does not correlate with adjuvanticity. The importance in receptor binding for immunogenicity and toxicity of LT was examined by comparing native LT and a B-subunit receptor binding mutant of LT, due to a single amino acid substitution from Gly to Asp at residue 33 of the B-subunit (Guidry et al., 1997). In contrast to the native LT holotoxin, the receptor binding mutant lacked GM₁ binding activity and had no enterotoxicity in a mouse model. Furthermore, the mutant was not able to elicit an antibody response against itself or co-administered antigens, thereby indicating that binding is not only essential for immunogenicity, but also toxicity (Guidry et al., 1997).

Studies examining the immunogenicity of B-subunits of LT only showed a Gly to Asp mutant at residue 33 of the B-subunit gave greatly reduced immunogenicity following subcutaneous immunisation when compared to the native LT-B (Nashar et al., 1996). Furthermore, immunogenicity was completely eliminated following oral immunisation (Nashar et al., 1996). Lymphocyte examination from mesenteric lymph nodes showed native LT-B and not the binding mutant, was specifically able to increase the proportion of B cells, deplete CD₈⁺ T cells, and increase activation of CD₄⁺ T cells (Nashar et al., 1996).

The mucosal adjuvant properties of LT and LT-B have led to the development of hybrid fusion proteins in which non- or poorly-immunogenic proteins such as ST (Aitken and Hirst, 1993; Cardenas and Clements, 1993a; Clements, 1990; Guzman-Verduzco and Kupersztoch, 1990; Sanchez et al., 1988; Sanchez et al., 1986) or other important bacterial or
viral epitopes (Dale and Chiang, 1995; Lipscombe et al., 1991; Loregian et al., 1999; Ricci et al., 2000) can be rendered immunogenic.

1.3.2 ST Enterotoxins

There are two types of unrelated ST toxins, although they do share common properties such as low molecular weight, heat stability, and interfere with water and electrolyte movement across the intestinal epithelium.

1.3.2.1 Classification of ST Enterotoxins

ST is classified into two unrelated groups, namely STa (ST-I) or STb (ST-II). These are structurally, functionally, immunologically and genetically distinct from one another (Burgess et al., 1978; Burgess et al., 1980; Gyles, 1992). STa is a small, non-immunogenic, low molecular weight peptide toxin with a molecular mass of < 2 kDa, which is methanol soluble and active in the infant mouse gut (Whipp et al., 1981). Furthermore, STa is resistant to temperatures of 100°C for up to 10 minutes, is soluble in organic solvents and water, and is resistant to proteolytic enzymes such as pronase, trypsin and chymotrypsin (Alderete and Robertson, 1978). However, STa can be inactivated by reducing and oxidising agents which disrupt disulfide bonds (Robertson et al., 1983).

STa can be further divided into two classes, namely ST-Ia (or STp) which is produced by animal and human isolates and ST-Ib (STh) which is produced by human isolates only (Moseley et al., 1983a; Sekizaki et al., 1985; Sekizaki et al., 1984; So and McCarthy, 1980; Takao et al., 1983). Genes for both ST-Ia and ST-Ib may be carried by a single strain of human ETEC (Moseley et al., 1983b).

The genes encoding STb are usually plasmid encoded and form part of a 9 kb transposon Tn4521 (Hu and Lee, 1988; Leece and Hirst, 1992; Picken et al., 1983). Synthesis of STb is mostly limited to porcine ETEC strains, but STb-specific genes have been detected in the faeces of humans on farms in Thailand (Echeverria et al., 1985).
1.3.2.2 Receptor and Structure of ST Enterotoxins

The receptor for STa in both rat and human intestinal cells has been identified as guanylyl cyclase C (GC-C or STaR), which is located on the brush border membranes of intestinal epithelia, with a monomeric molecular mass of 140 kDa and contains an extracellular ligand binding site and a cytoplasmic guanylyl cyclase catalytic domain, which is characteristic of the natriuretic peptide receptor family (de Sauvage et al., 1991; de Sauvage et al., 1992; Schulz et al., 1990). Guanylyl cyclase C exists as an oligomer both in the presence and absence of the ST toxin, and consists of functionally distinct subunits which are non-covalently linked in the intestine (Rudner et al., 1995; Vaandrager et al., 1994). The active GC-C complex has a molecular mass of 393 kDa, suggesting that GC-C is a trimer under native conditions (Vaandrager et al., 1994). The binding site for ST is located within extracellular domain of guanylyl cyclase C consisting of residue 387 to residue 393, which is close to the trans-membrane portion of guanylyl cyclase C on the external cellular surface (Hasegawa et al., 1999). The specific receptor for STb has yet to be identified.

STp (ST-Ia) is an 18 amino acid toxin (So and McCarthy, 1980; Takao et al., 1983) and STh (ST-Ib) is a 19 amino acid toxin (Aimoto et al., 1982; Chan and Giannella, 1981; Staples et al., 1980). Both are synthesised as 72 amino acid precursors consisting of pre- (residues 1-19), pro- (residues 20-54 for STp and residues 20-53 for STh) and mature-(residues 55-72 for STp and residues 54-72 for STh) regions. The precursor is translocated across the inner membrane by the general export pathway, consisting of Sec proteins, which specifically cleaves the 19 amino acid leader peptide (Rasheed et al., 1990). The signal sequence (Pre) is essential in the delivery of Pro-STh into the periplasm and the absence of the Pre region leads to increased susceptibility of STh to intracellular proteases (Yang et al., 1992).

The Pro region of STp is thought to guide the mature region of ST into the periplasm, but its presence is not required for efficient extracellular secretion of mature STp (Yamanaka et al., 1993; Yang et al., 1992). The Pro region may also determine the correct pattern of
disulphide formation, since a mutant Cys residue within the Pro region results in a significant decrease of ST activity in the cell supernatant, the accumulation of inactive STp in the periplasm, and an alteration to the cleavage site of the intermediate mature STp (Yamanaka et al., 1994). The formation of intramolecular disulphide bonds mediated by the disulphide bond epimerase DsbA, with the TolC outer membrane protein required for secretion of STp into the extracellular environment (Yamanaka et al., 1994; Yamanaka et al., 1997; Yamanaka et al., 1998a). A Glu residue on STp has been shown to be crucial for recognition of ST by DsbA in the periplasm to catalyse disulphide bond formation (Yamanaka et al., 1998b).

The mature STa protein consists of three intramolecular disulphide bonds which are essential for toxicity, although the contribution of each disulphide bond to toxicity is not equal (Okamoto et al., 1987). Epitope mapping of STh using a panel of monoclonal antibodies revealed three distinct epitopes in ST (Takeda et al., 1993). The three distinct antigenic sites were located in the N-terminal region, the core region and the C-terminal region respectively. A monoclonal antibody was developed that showed potent neutralisation activity against the ST toxin by recognising the N-terminal epitope Asn-Ser-Ser-Asn-Thr, which has been shown to play no role in the toxic activity of ST (Takeda et al., 1993).

STb (ST-II) consists of 48 amino acids with two intramolecular disulphide bonds between Cys-10 and Cys-48 and on Cys-21 and Cys-36 (Fujii et al., 1991). These bonds are absolutely required for the biological action of STb, but also help stabilise the protein and allow the toxin to be efficiently secreted (Arriaga et al., 1995; Okamoto et al., 1995). The specific proteins involved in disulphide formation and secretion are DsbA and TolC respectively (Foreman et al., 1995).

1.3.2.3 Mechanism of Action

The initial step in the biological action of STa involves the binding of the toxin to guanylyl cyclase C receptors on the brush border membranes of the intestinal epithelium, which leads to the activation of particulate guanylate cyclase (Schulz et al., 1990; Waldman et
The activated guanylate cyclase converts GTP to cyclic GMP (cGMP) resulting in the accumulation of cGMP in the intestinal cells (Field et al., 1978; Huott et al., 1988; Waldman et al., 1986). Elevated cGMP causes fluid secretion by the inhibition of sodium coupled chloride absorption in the villus tips and stimulation of electrogenic chloride secretion in crypt cells, leading to excessive fluid in the intestinal lumen (Gyles, 1992; Waldman et al., 1986).

In contrast to the mechanism of action of STα, STβ does not cause an increase in the intracellular concentration of cyclic AMP or cyclic GMP and the intestinal secretory response is mediated by the release of bicarbonate rather than chloride (Gyles, 1992). STβ specifically mediates calcium influx by opening a GTP-binding regulatory protein-linked receptor-operated calcium channel in the plasma membrane of epithelial cells (Dreyfus et al., 1993).

1.4 E. coli Pili

The terms pili or fimbriae are used interchangeably to describe the hair-like proteinaceous surface appendages produced by a plethora of bacterial species. The function of pili is to mediate bacteria-host attachment, which is necessary to overcome host factors such as peristalsis or mucus secretion. The pili are usually classified by the ability to cause agglutination of erythrocytes from different species. In non-pathogenic E. coli strains, type I or common pili are produced which exhibit mannose-sensitive haemagglutination (MSHA) of erythrocytes (Collier and DeMiranda, 1955). All members of the family Enterobacteriaceae appear to be capable of expressing type I pili. In addition to producing type I pili, pathogenic E. coli are characterised by the ability to produce specific colonisation factors which are, in most cases, pili responsible for attachment of the bacteria to the intestinal epithelium (Levine, 1987). These pili are distinguishable from type I pili because in almost all cases they exhibit mannose-resistant haemagglutination (MRHA) of erythrocytes (de Graaf and Gaastra, 1994).
1.4.1 Type I Pili

Type I pili are approximately 7 nm in diameter and form a hollowed-core helical array approximately 1-2 μm long (Brinton, 1965). The role of type I pili in infection is somewhat controversial, given that 70% of all normal faecal isolates express type I pili (Denke et al., 1979; Duguid et al., 1979). It is known that type I pili are able to bind mucosal epithelial cells (Isaacson et al., 1978) and that they play a role in promoting bacterial colonisation of the urinary tract (Keith et al., 1986), but not in the large or small bowel (McCormick et al., 1989).

The genes encoding type I pili are located at 98 min on the E. coli chromosome and have been cloned and sequenced (Bachmann, 1990; Klemm, 1984; Klemm and Christiansen, 1987; Klemm et al., 1985; Orndorff and Falkow, 1985). There are eight genes required for biosynthesis of type I pili, designated fimA-H (Figure 1.1). The fimA gene encodes the 17 kDa major structural subunit which forms a tight right-handed helical structure (Brinton, 1965; Klemm, 1984). FimH has been identified as the pilus tip adhesin responsible for receptor binding to mannose residues (Hanson and Brinton, 1988; Klemm and Christiansen, 1987; Krogfelt et al., 1990; Maurer and Orndorff, 1987; Orndorff and Falkow, 1985). The bulk of FimH is assembled as a distinct fibrillar tip structure, approximately 16 nm long, together with FimG and possibly FimF (Jones et al., 1995). FimG also regulates pilus length (Orndorff, 1994; Russell and Orndorff, 1992). FimC is the chaperone-like molecule which is required for pilus assembly, and is highly homologous to the PapD chaperone from Pap pili of uropathogenic E. coli (Klemm and Christiansen, 1990). PapD can complement a mutant fimC to produce type I pili (Jones et al., 1993). FimD is an outer membrane assembly protein which acts as a polymerisation channel for pilin monomers before surface assembly (Klemm and Christiansen, 1990).

The products of fimB and fimE are site-specific tyrosine recombinases required for the phase variation, whereby the cells switch between the ON and OFF phase of type I pili expression (McClain et al., 1993; McClain et al., 1991). The molecular basis for the observed phase variation due to FimB and FimE has now been established, and involves the orientation
Figure 1.1 Genetic organisation of the *fim* gene locus required for biosynthesis of type I pili and mechanism of phase variation.

The specific *fim* genes are indicated and their function defined. The invertible promoter element is represented by a “p” and is orientated according to the arrow (adapted from Orndorff, 1994 and Thanassi *et al.*, 1998a).
Type I pilus (\textit{fim}) gene cluster

250 bp

ON \rightarrow P \rightarrow OFF

1 kb

Regulation Regulation Major pilin subunit Periplasmic chaperone Outer membrane assembly protein Minor pilin subunit Minor pilin subunit Adhesin

Tip fibrillum components
of a 314 bp DNA element (the switch) that undergoes DNA inversion (Kulasekara and Blomfield, 1999; Stentebjerg-Olesen et al., 2000)(Figure 1.1). FimB promotes recombination with little orientational bias, while FimE promotes recombination in the ON-to-OFF direction exclusively, by showing a strong preference for the switch in the ON orientation as a substrate for specific recombination (Kulasekara and Blomfield, 1999).

1.4.2 Curli

Curli are a novel class of surface organelles which have been identified in E. coli and Salmonella, and are morphologically distinct from pili (Collinson et al., 1991; Olsen et al., 1989). Unlike pilus fibres which are assembled from the base, curli fibers form from outside the bacteria by an extracellular nucleation pathway and have been shown to specifically bind the host cell proteins fibronectin, laminin and plasminogen, although the role of curli in the biology and pathogenesis of E. coli or Salmonella is unknown (Arnqvist et al., 1992; Olsen et al., 1989; Sjobring et al., 1994). In certain E. coli K-12 strains such as MC4100 curli expression is induced during stationary phase when grown at temperatures below 37°C, and under conditions of low osmolarity (Olsen et al., 1989). Given the temperatures and conditions in which curli are expressed, they may be specific features for survival of enteric bacteria outside a eukaryotic host. The csgA gene encodes the 15 kDa major subunit protein of the curli fibres and is located on the chromosome at 23.1 minutes (Arnqvist et al., 1994). CsgA bears no homology to any of the characterised pilins from other systems, while very little is known about the biogenesis of curli formation.

1.5 Pili of Enterotoxigenic Escherichia coli

The pili of ETEC strains are remarkably host specific. Colonisation factors are found exclusively on either human, porcine, or bovine strains. No colonisation factors have been detected which occur in both human and animal strains of ETEC. Therefore, ETEC host tropism is attributable to the diversity of pili produced by the different strains, such that
different pili of the human ETEC family do not mediate binding of the bacterium to the intestinal epithelium of other mammals (Edwards and Puente, 1998). For example, human volunteers did not develop diarrhoea when administered a large oral inoculum of a porcine ETEC strain, which is responsible for severe diarrhoea in piglets (DuPont et al., 1971). However, some colonisation factors found on pigs have also been identified in calves, or vice-versa, suggesting that some of these adhesins are less host-specific than those identified on human strains. In this thesis, the colonisation factors of animal ETEC strains will be discussed; however; the various colonisation factors identified from human ETEC strains will be addressed in greater detail.

1.6 Animal ETEC Colonisation Factors

ETEC infections cause significant problems in the agricultural industry, particularly with neonatal calves and pigs. Due to the economic significance of these animals in agriculture, a great deal of research has been undertaken in identifying and characterising colonisation factors important for adhesion and subsequent infection by these ETEC strains. Thus far, 12 different colonisation factors have been identified, and include K88 (F4) (Moseley et al., 1986; Ørskov and Ørskov, 1966), CS31A (Girardeau et al., 1988), K99 (F5)(Smith and Linggood, 1972), 987P (F6) (Isaacson and Richter, 1981; Nagy et al., 1976), CS1541 (Broes et al., 1988), F17 (Fy)(Lintermans et al., 1988), F41 (Moseley et al., 1986), F42 (Yano et al., 1986), F141 (Kennan and Monckton, 1990), F165 (Fairbrother et al., 1986), 2134P (Casey et al., 1992) and 8813 (Salajka et al., 1992). All of the colonisation factors are thin, flexible fibrillar type pili structures approximately 2-4 nm in diameter, with the exception of 987P, F141 and F165 which are rigid pili. No structure has yet been identified for 2134P and 8813 colonisation factors. The best studied colonisation factors are K88, 987P and K99, which are found exclusively on ETEC strains infecting pigs (K88 and 987P), or both pigs and calves (K99). The molecular genetics of biogenesis of these pili is well established. The genetic organisation of the K88 and K99 fimbrial operons show striking structural
similarities to each other, and the genes show a high degree of homology (Figure 1.2). A brief description of the genes involved in the biogenesis of K88, K99 and 987P is outlined below.

1.6.1 K88

The genes required for biosynthesis of K88 fibrillae are located on large virulence plasmids, and have been cloned and sequenced (Bakker et al., 1992; Mooi et al., 1979; Shipley et al., 1978). Ten genes, designated faeA-J are required for regulation and biosynthesis of K88 (Figure 1.2). FaeA and FaeB are specifically involved in the regulation of expression of the assembly genes consisting of faeC-J (Bakker et al., 1992; Huisman et al., 1994; Huisman and de Graaf, 1995; Mooi et al., 1979; Shipley et al., 1978).

FaeG is the 27 kDa major subunit which encodes the adhesin responsible for binding Gal\(\alpha(1-3)\)Gal residues on unknown host cell receptors (Shipley et al., 1981; Willemsen and de Graaf, 1992). The genes for faeC, faeF, faeH, faeI and faeJ all encode minor pilin subunits (Bakker et al., 1992; Oudega et al., 1989). Immunoelectron microscopy showed that FaeC is located at the tip of the K88 fibrilla, while mutants in faeC do not produce pili, thereby indicating FaeC is likely to be the first pilin assembled (Mooi et al., 1984; Mooi et al., 1982; Oudega et al., 1989). FaeF, FaeH and FaeI were detected in or along K88 at more or less regular intervals (Oudega et al., 1989). Mutants in faeF did not produce pili, which suggests that FaeF is also assembled early during biogenesis (Mooi et al., 1984; Mooi et al., 1982). Mutations in faeH and faeI still produce K88 indicating these pilins may be involved in initiation and elongation of the pili, since a double mutant in both produced only 10% of the wild type level of pili (Mol and Oudega, 1996).

The faeD and faeE genes encode the 81.7 kDa outer membrane assembly protein and the 24.8 kDa chaperone protein, respectively (Bakker et al., 1991; Mol et al., 1994; Mooi et al., 1986). The topology and subcellular localisation of FaeD showed it contained a large central domain of 24 membrane spanning segments and two periplasmic regions located at the N- and C-terminal ends of the protein (Valent et al., 1995). The FaeE periplasmic chaperone
Figure 1.2 Genetic organisation of the K88 and K99 fimbrial operons.

The genes are indicated by boxes, and the function of each gene is indicated according to the different shading patterns used. Arrows indicate the direction of transcription of the operons (adapted from Mol and Oudega, 1996).
stabilises the major subunit FaeG, but also FaeH, FaeI and FaeF (Mooi et al., 1983). FaeE exists as a homodimer, but also forms a heterotrimeric complex with FaeG (Mol et al., 1994) but also with FaeH and FaeI proteins (Mol et al., 1996b). The N-terminal β-barrel domain of FaeE is required for both subunit recognition and dimerisation (Mol et al., 1996).

1.6.2 K99

The genes required for biosynthesis of K99 fibrillae are located on a 52 MDa conjugative plasmid (van Embden et al., 1980). The biosynthesis of K99 fibrillae is dependent on two regulatory genes, fanA and fanB (Roosendaal et al., 1987) and six genes which encode the structural proteins, namely fanC, fanD, fanE, fanF, fanG and fanH (Roosendaal et al., 1989; Roosendaal and de Graaf, 1989; Roosendaal et al., 1984; Simons et al., 1990; Simons et al., 1991)(Figure 1.2). The 16.5 kDa major subunit FanC is the adhesin responsible for binding the host cell receptor, identified as the glycolipid hematoside or Neu5Gcα(2-3)Galp-β(1-4)-Glcp-β(1-1) ceramide (Kyogashima et al., 1989; Smit et al., 1984). Mutagenesis studies on FanC have shown that Lys-132 and Arg-136 have a crucial role in receptor binding activity and probably form part of the receptor binding domain, while Trp-67 has an essential structural role (Jacobs et al., 1987b). The tip located FanF, along with FanG and FanH are all minor subunits which are specifically involved in the initiation and elongation, but not receptor binding of K99 (Simons et al., 1990; Simons et al., 1991). The fanD gene encodes the large outer membrane assembly protein (Roosendaal and de Graaf, 1989) and FanE is a periplasmic chaperone protein responsible for protecting FanC from degradation (Bakker et al., 1991). Unlike the FaeE chaperone from K88 which forms a homodimer, FanE is a monomeric protein (Mol et al., 1996a). The chaperones FanE, FaeE and PapD are highly homologous; however, they cannot complement each other, which suggests that the interactions between the chaperones and the subunits in these systems is very specific (Bakker et al., 1991).
1.6.3 987P

The genes required for biosynthesis of 987P pili from porcine ETEC are plasmid-encoded (Schifferli et al., 1990). Eight genes are required for biosynthesis of 987P pili, and are designated *fasA* to *fasH* (Figure 1.3). The *fasA* gene encodes the major structural subunit protein (de Graaf and Klaasen, 1987), FasG is the adhesin (Khan and Schifferli, 1994) and FasF encodes minor pilin subunit thought to function as an adaptor protein between the FasG adhesin located at the tip of the pilus and the FasA structural subunits (Cao et al., 1995). The FasG adhesin is also essential for fimbrial biogenesis (Khan and Schifferli, 1994). The *fasD* gene encodes the outer membrane assembly protein which is specifically responsible for the ordered translocation of the subunits across the outer membrane (Cao et al., 1995). The *fasB* gene encodes the major subunit (FasA)-specific chaperone with strong homology to PapD, *fasC* encodes a second chaperone that is specific for the adhesin FasG and *fasE* encodes a chaperone-like protein which is located in the periplasm and is responsible for optimal translocation of FasG and possibly other subunits (Edwards et al., 1996). The length of the 987P pilus is determined by the levels of expression of the different *fas* genes (Schifferli et al., 1991b). Finally, the *fasH* gene, which is orientated in the opposite orientation to the other *fas* genes (Figure 1.3) functions as transcriptional activator of *fasA* and possibly other *fas* genes (Klaasen and de Graaf, 1990).

1.7 Human ETEC Colonisation Factors

In ETEC strains isolated from humans, the colonisation factors are referred to as colonisation factor antigens (CFAs), coli surface-associated antigens (CS) or putative colonisation factors (PCF), and have been implicated in the establishment of infection (Levine, 1987). Human ETEC pili range in size from 3-7 nm and are composed of hundreds of copies of major subunits varying from 14-27 kDa. The pili may also contain one or more minor pilin subunits incorporated into the pilus structure. Some pili are also referred to as
Figure 1.3 Genetic organisation of the 987P fimbrial operon.

The genes are indicated by boxes, and the function of each gene is indicated by a downward pointing thick arrow from each box. Fine arrows indicate the direction of transcription of the assembly genes and the regulatory gene fasH (adapted from Mol and Oudega, 1996).
987P (fas) OPERON

- fasA: Major subunit
- fasB: Chaperone for adhesin
- fasC: Chaperone for major subunit
- fasD: Outer membrane assembly protein
- fasE: Minor subunit
- fasF: Chaperone-like protein
- fasG: Adhesin
- fasH: Regulation
fibrillae as they are thin (2-3 nm in diameter) and more flexible than normal pili. At present, 21 different colonisation factors have been described for human ETEC, which are discussed in detail below and described in Table 1.1.

1.7.1 Identification, Morphology and Genetics

1.7.1.1 CFA/I

Colonisation factor antigen I (CFA/I) pili was the first identified human ETEC colonisation factor (Evans et al., 1975). CFA/I is one of the more common colonisation factors and is associated with ETEC strains which express ST and LT or ST alone and are of the O4, O7, O15, O20, O25, O63, O78, O110, O126, O128, O136, O153 and O159 serogroups (Blanco et al., 1989; Changchawalit et al., 1984; Cravioto et al., 1982; Darfeuille-Michaud et al., 1987; Levine et al., 1983; Murray et al., 1983). CFA/I is a 7 nm rigid pili which is composed almost entirely of a 15 kDa subunit pilin protein (Evans et al., 1975). CFA/I expressing ETEC strains are able to haemagglutinate bovine and human erythrocytes and mediate adherence to both cultured human enterocytes and the human colonic carcinoma cell line CaCo-2 (Evans et al., 1975; Viboud et al., 1996b).

Monoclonal antibodies against CFA/I cross-react with a number of heterologous colonisation factors including CS1, CS2, CS4, CS14 (PCFO166) and CS17 (McConnell et al., 1989b; Rudin and Svennerholm, 1994). Furthermore, monoclonal antibodies raised against CFA/I have been shown to inhibit binding of CS4 expressing ETEC to human enterocytes and could protect against the same strain of CS4 in rabbit ileal loops (Viboud et al., 1996b). This suggests that CFA/I and CS4 are closely related in evolutionary terms.

All the genes required for CFA/I biosynthesis are located on the same plasmid but within two regions which are approximately 40 kb apart (Smith et al., 1982). The genetic organisation of the CFA/I-encoding region shows striking homology with that of CS1 (section 1.7.1.2.1) and CS2 pili (section 1.7.1.2.2)(Figure 1.4). Region 1 contains four genes required for production of CFA/I pili, designated cfaA, cfaB, cfaC and cfaE (Hamers et al., 1989; Jordi
### Table 1.1 Characteristics of Human ETEC Colonisation Factors
(Adapted from Svennerholm and Gaastra, 1996)

<table>
<thead>
<tr>
<th>CF</th>
<th>CS</th>
<th>Morphology</th>
<th>Size (kDa)</th>
<th>Regulation</th>
<th>Toxins</th>
<th>Serogroup</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CFA/I-like group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFA/I</td>
<td>CFA/I</td>
<td>P</td>
<td>7 nm</td>
<td>15.0</td>
<td>CfaR</td>
<td>ST+LT</td>
<td>O4, O7, O15, O20, O25, O63, O78, O110, O126, O128, O136, O153, O159</td>
</tr>
<tr>
<td>CS1</td>
<td>CS1</td>
<td>P</td>
<td>7 nm</td>
<td>16.5</td>
<td>Rns</td>
<td>ST+LT</td>
<td>O6, O139</td>
</tr>
<tr>
<td>CS2</td>
<td>CS2</td>
<td>P</td>
<td>7 nm</td>
<td>15.3</td>
<td>Rns</td>
<td>ST+LT</td>
<td>O6</td>
</tr>
<tr>
<td>CS4</td>
<td>CS4</td>
<td>P</td>
<td>6 nm</td>
<td>17.0</td>
<td>CfaR-like</td>
<td>ST+LT</td>
<td>O25</td>
</tr>
<tr>
<td>PCFO166</td>
<td>CS14</td>
<td>P</td>
<td>7 nm</td>
<td>15.5/17.0</td>
<td>CfaR-like</td>
<td>ST</td>
<td>O20, O71, O98, O166</td>
</tr>
<tr>
<td>CS17</td>
<td>CS17</td>
<td>P</td>
<td>7 nm</td>
<td>17.5</td>
<td>CfaR-like</td>
<td>LT</td>
<td>O8, O15, O48, O114, O146</td>
</tr>
<tr>
<td>CS19</td>
<td>CS19</td>
<td>P</td>
<td>7 nm</td>
<td>16.0</td>
<td>CfaR-like</td>
<td>ST+LT</td>
<td>O8</td>
</tr>
<tr>
<td><strong>'CS5' Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS5</td>
<td>CS5</td>
<td>H</td>
<td>5 nm</td>
<td>23.0</td>
<td>ST</td>
<td>ST</td>
<td>O6, O29, O92, O114, O115, O167</td>
</tr>
<tr>
<td>CS7</td>
<td>CS7</td>
<td>H</td>
<td>3-6 nm</td>
<td>21.5</td>
<td>ST+LT</td>
<td>LT</td>
<td>O15, O103, O114</td>
</tr>
<tr>
<td>PCFO9</td>
<td>CS13</td>
<td>f</td>
<td>27.0</td>
<td></td>
<td>ST+LT</td>
<td>LT</td>
<td>O9, O143</td>
</tr>
<tr>
<td>CS20</td>
<td>CS20</td>
<td>P</td>
<td>7 nm</td>
<td>20.8</td>
<td>CfaR-like</td>
<td>ST+LT</td>
<td>O17</td>
</tr>
<tr>
<td>PCFO20</td>
<td>CS18</td>
<td>P</td>
<td>7 nm</td>
<td>25.0</td>
<td>CfaR-like</td>
<td>ST+LT</td>
<td>O20</td>
</tr>
<tr>
<td><strong>Type-IV-like</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFA/III</td>
<td>CS8</td>
<td>P</td>
<td>7 nm</td>
<td>18.0</td>
<td>LT</td>
<td>ST</td>
<td>O25</td>
</tr>
<tr>
<td>LONGUS</td>
<td>CS21</td>
<td>P</td>
<td>7 nm</td>
<td>22.0</td>
<td>ST+LT</td>
<td>ST, LT</td>
<td>O2, O6, O8, O20, O25, O49, O128, O139, O148</td>
</tr>
<tr>
<td>8786</td>
<td>CS15</td>
<td>nP</td>
<td></td>
<td>16.3</td>
<td>ST</td>
<td>ST+LT</td>
<td>O117</td>
</tr>
<tr>
<td>CS22</td>
<td>CS22</td>
<td>f</td>
<td></td>
<td>15.7</td>
<td>ST+LT</td>
<td>ST+LT</td>
<td>O20</td>
</tr>
<tr>
<td><strong>Distinct</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS3</td>
<td>CS3</td>
<td>f</td>
<td>2-3 nm</td>
<td>15.1</td>
<td>ST+LT</td>
<td>ST+LT</td>
<td>O8, O9, O78, O80, O115, O128, O139, O168</td>
</tr>
<tr>
<td>CS6</td>
<td>CS6</td>
<td>nP</td>
<td></td>
<td>14.5/16.0</td>
<td>ST+LT</td>
<td>ST, LT</td>
<td>O25, O27, O92, O148, O153, O159</td>
</tr>
<tr>
<td>PCFO148</td>
<td>CS11</td>
<td>f</td>
<td>3 nm</td>
<td>19.0</td>
<td>ST+LT</td>
<td>ST+LT</td>
<td>O148</td>
</tr>
<tr>
<td>PCFO159</td>
<td>CS12</td>
<td>P</td>
<td>7 nm</td>
<td></td>
<td>ST+LT</td>
<td>ST+LT</td>
<td>O159</td>
</tr>
<tr>
<td>2230</td>
<td>CS10</td>
<td>nP</td>
<td></td>
<td></td>
<td>ST</td>
<td>ST</td>
<td>O25</td>
</tr>
</tbody>
</table>

**Abbreviations:** P- pilus; H- helical with two fimbrils arranged in a helix; f- fibrillae; nP- non-pilus antigen; CF- colonisation factor; CS- coli surface associated antigen; LT- heat labile enterotoxin; ST- heat stable enterotoxin
Figure 1.4 Genetic organisation of the CFA/I, CS1 and CS2 fimbrial operons.

The boxes represent the individual genes, with the functions indicated by the differential shading patterns used for each box. The operons are orientated according to the arrow (adapted from Sakellaris and Scott, 1998).
et al., 1992b) (Figure 1.4). The overall % G+C content of the region is only 28% versus the 50% overall % G+C content for E. coli (Jordi et al., 1992b). The cfaB gene encodes the 15 kDa subunit protein. The mRNA encoding for the CsfB protein is the most stable and abundant mRNA species present. This may be a factor in gene expression since CfaB protein levels are at much greater amounts in the cell (Jordi et al., 1993). The cfaA gene probably encodes a periplasmic chaperone, since it bears strong homology with the CooB chaperone from CS1 pili (Voegele et al., 1997). Likewise, the cfaC gene shows homology with the CooC outer membrane assembly protein of CS1 pili (Froehlich et al., 1994; Sakellaris et al., 1996) and the cfaE gene with the minor pilin subunit CooD of CS1 (Froehlich et al., 1994). This similarity between the coo and cfa operons has been utilised in trans complementation experiments in which the cfaB gene was introduced into a cooA mutant strain to express cell surface CFA/I pili (Marron and Smyth, 1995). CfaE has been shown to be a tip protein of CFA/I pili essential in CFA/I-mediated adherence, since specific residue changes in CfaE were able to abolish haemagglutination without abrogating CFA/I pili assembly (Sakellaris et al., 1999a).

A single gene designated cfaD or cfaR is contained in region 2 and is required for CFA/I production (Savelkoul et al., 1990). CfaD/CfaR is very basic (pI = 9.5), with a C-terminal helix-turn-helix motif characteristic of DNA binding proteins (Savelkoul et al., 1990). It is highly homologous to Rns, a member of the AraC family of positive regulators, which is required for expression of CS1 and CS2 pili (Caron et al., 1989). CfaD/CfaR can functionally substitute for Rns to induce expression of CS1 and CS2 and likewise Rns can induce expression of CFA/I pili (Caron and Scott, 1990; Savelkoul et al., 1990). CfaD/CfaR specifically mediates a positive regulatory effect on the CFA/I operon by overcoming the repression mediated by the global regulatory protein H-NS (histone-like protein), which acts by sequestering the DNA across a single promoter upstream of the cfaA gene (Jordi et al., 1992a; Jordi et al., 1993).
1.7.1.2 CFA/II

Colonisation factor antigen II (CFA/II) was initially identified as a single antigen on a diarrhoea-causing strain of ETEC serotype O78:H11 (Evans et al., 1975). Subsequent experiments have shown that CFA/II actually consists of three serologically and morphologically distinct antigens, namely CS1 and CS2 pili and CS3 fibrillae (Cravioto et al., 1982; Evans et al., 1975; Smyth, 1982). CFA/II expressing ETEC strains produce CS3 alone or in combination with CS1 or CS2. ETEC strains expressing CS3 only are associated with ST- and LT-producing O8, O9, O78, O80, O85, O115, O128 and O168 serogroups. CS1 and CS3 are produced in O6:H16 or O6:H- strains of biotype A, whereas CS2 and CS3 are expressed in O6:H16 strains of biotype B, C and F. CS1 and CS3 are also found on LT- and ST-producing O139 serogroups (Cravioto et al., 1982; Evans and Evans, 1978; Smyth, 1982; Thomas and Rowe, 1982).

1.7.1.2.1 CS1

CS1 is a rigid, rod-like pilus composed of a 16 kDa subunit with a diameter of 6 nm (Levine et al., 1984). CS1 has been shown to be a colonisation factor, since monoclonal antibodies against CS1 were able to inhibit ETEC-mediated experimental disease in rabbits (Svennerholm et al., 1990), and significant anti-CS1 immunity was observed in patients infected with ETEC strain O139:H28 which expresses CS1 and CS3 (Levine et al., 1984).

1.7.1.2.1.1 Regulation

The genes required for CS1 biosynthesis were initially assumed to be localised on the same plasmid which also encodes ST and LT (Mullany et al., 1983; Smith et al., 1982). It has since been shown that the genes required for CS1 production are chromosomally-located and designated cooA, cooB, cooC and cooD (Figure 1.4). A specific plasmid-borne positive regulator, rms is absolutely required for CS1 expression (Caron et al., 1989). Transcriptional
control of the **coo** genes is from a transcription silencer region which extends over **cooA** and into **cooB**, and is dependent on upstream regions for its effect (Murphree *et al.*, 1997). The DNA-binding protein H-NS acts in the same region as the silencer, therefore H-NS may play a role in the repression of **coo** transcription (Murphree *et al.*, 1997). The silencing is overcome by Rns. The Rns binding site in CS1 has been localised to two distinct regions within the **coo** promoter and a further 60 bp upstream (Munson and Scott, 1999). Rns recognises asymmetric nucleotide sequences in two regions of the major groove and binds along one face of the DNA helix (Munson and Scott, 1999).

### 1.7.1.2.1.2 Gene Function

The molecular genetics of CS1 pili biogenesis is the best studied system of any human ETEC colonisation factor. The assembly of CS1 pili relies on only four genes, **cooA**, **cooB**, **cooC** and **cooD**; which bear no homology with any of the immunoglobulin-like chaperones, or outer membrane assembly proteins characterised in other systems such as Pap pili from uropathogenic *E. coli* (Hultgren *et al.*, 1991). The **coo** genes are flanked by DNA which bears homology to a number of insertion sequence (IS) elements including IS629, IS2 and IS150 (Froehlich *et al.*, 1994; Scott *et al.*, 1992). Since the **coo** genes are flanked by remnants of transposable elements and the % G+C content of the **coo** genes averages only 38%, which is significantly lower than the average for *E. coli* of 50%, it is likely the CS1-encoding region may have been introduced into *E. coli* from another organism by transposition (Froehlich *et al.*, 1994).

The **cooA** gene encodes the 16 kDa major subunit (Marron and Smyth, 1995; Perez-Casal *et al.*, 1990), **cooB** encodes a chaperone-like protein (Voegele *et al.*, 1997), **cooC** encodes an outer membrane assembly protein and **cooD** encodes a minor pilin located at the tip of the structure (Sakellaris *et al.*, 1996). CooA has been found to occur in intermolecular complexes with the CooB assembly protein in the periplasm; however, CooA multimers can also exist but their function is currently unknown (Sakellaris *et al.*, 1996). CooA is also able to interact with
the CooD minor pilin in the periplasm. Such complexes may either be assembly intermediates, or simply play no role in biogenesis (Sakellaris et al., 1996). CooC is not required for stability of the pilin subunits, but is absolutely required for biogenesis of CS1 pili (Froehlich et al., 1994). CooC is a 94 kDa protein which was found to fractionate exclusively with the outer membrane and is predicted to be specifically involved in assembly of pilin subunits across the outer membrane (Sakellaris et al., 1996).

CooB was initially described as being required for assembly, but not transport of the CS1 pilin, since a mutation in cooB which abolished cell-surface CS1 pili expression, did not prevent CooA was transportation across the outer membrane (Scott et al., 1992). Subsequent experiments have shown that CooB is located in the periplasm and is able to specifically form intermolecular complexes with CooA, CooD and CooC, but also stabilise these proteins (Sakellaris et al., 1996; Voegele et al., 1997). In the absence of CooB, very little CooA is detected. For CooD, an increase in a 25 kDa truncated form of the protein is detected in the absence of CooB, but when CooB is restored the 38 kDa full form of CooD dominates (Voegele et al., 1997). Unlike other chaperones described, CooB also had a positive effect on CooC expression, and interacted directly with CooC. In the presence of CooB the full 94 kDa form of CooC is observed, but when CooB is absent, an accumulation of a 70 kDa truncated form occurs. If CooC was present in the system then CooB was also found to interact directly with it in the outer membrane. In the absence of CooC, CooB was localised strictly in the periplasm (Voegele et al., 1997).

CooD is required for the assembly of CS1 pili, but is not responsible for stability of the major subunit CooA in the periplasm (Froehlich et al., 1994). CS1 pili are composed almost entirely of the major subunit CooA, with CooD predicted to contribute only one subunit to the distal tip of the pilus structure (Sakellaris et al., 1996). Therefore, CooD is an integral tip-located minor pilin. CooD is necessary for the extracellular transport of CooA, which suggests that CooD is assembled first into the pilus structure (Sakellaris et al., 1999a). The level of expression of CooD was found to be correlated with the number of cell-surface-
assembled CS1 pili in *E. coli* K-12 and is therefore rate-limiting for CS1 pili biosynthesis (Sakellaris *et al.*, 1999b). Recent experiments on CS1 pili have found that a point mutation in CooD, which does not interfere with CS1 pilus formation, abolished haemagglutination, suggesting that CooD is involved in adherence (Sakellaris *et al.*, 1999a).

The pilins of CS1 (CooA and CooD) lack a number of the characteristics of the Pap and Pap-related pilins (Lindberg *et al.*, 1986; Soto *et al.*, 1998). The major pilin subunit CooA was found to lack the penultimate Tyr residue, the disulphide bond and the invariant Gly residue, while the minor pilin CooD has two Cys residues, but also lacks all other characteristics (Sakellaris *et al.*, 1996). A conserved C-terminal amino acid sequence motif is shared by the CS1 pilins (A-G-x-Y-x-G-(x)_{6}-T, where x denotes non-identical residue), which is also conserved among the characterised pilins of the CS1 family that includes CFA/I, CS2, CS4, CS14, CS17 and CS19 (Sakellaris *et al.*, 1996; Sakellaris and Scott, 1998). The conserved motif may serve as a recognition site for pilin-pilin interactions or with other components of the assembly machinery (Sakellaris *et al.*, 1996).

### 1.7.1.2.1.3 Model of CS1 Pilus Assembly

Based on the accumulation of knowledge for CS1 biogenesis, a model of assembly has been proposed (Sakellaris and Scott, 1998), which is outlined in Figure 1.5. In short, the CooA and CooD pilins are chaperoned by CooB once they enter the periplasm. This promotes the correct folding and/or prevents autoassembly and degradation of the subunits. CooB also binds and stabilises the CooC outer membrane assembly protein. It is proposed that pilus nucleation begins with a specific interaction between CooD-CooB complexes and a free CooB-CooC complex in the outer membrane, leading to the release of CooB from the pilin subunit or outer membrane where it can be recycled back to bind newly secreted subunits. Binding of CooD to CooC may lead to a conformational change in CooC such that CooC is able to bind CooA-CooB complexes and drive polymerisation of the pilus structure (Sakellaris and Scott, 1998). It has been shown in the Pap system that the outer membrane assembly
Figure 1.5 A model for CS1 pilus biogenesis.

(from Sakellaris and Scott, 1998)
protein PapC plays an active role in determining the order of pilus subunit passage across the outer membrane (Dodson et al., 1993; Jacob-Dubuisson et al., 1993). An analogous system probably occurs during CS1 pilus biogenesis (Sakellaris and Scott, 1998).

The coo system does not contain a gene that is responsible for length regulation of pilus assembly. In other systems such as Pap, type I pilis, or MR/P pilis from *Proteus mirabilis* a specific protein is required to terminate further growth of the pilus structure, thereby controlling the length of the pili (Båga et al., 1987; Li and Mobley, 1998; Russell and Orndorff, 1992). It is thought that length is controlled in the CS1 system by the relative affinity of CooA for CooD versus its affinity for CooA, and the concentrations of the pilin proteins available for assembly (Sakellaris et al., 1996).

1.7.1.2.2 CS2

CS2 pili are approximately 7 nm in diameter and contain a 15.3 kDa subunit protein with a rigid-rod like morphology (Smyth, 1982). CS2 has been shown to be a colonisation factor, since monoclonal antibodies against CS2 were able to protect against experimental infection in rabbits (Svennerholm et al., 1990). The chromosomal location of the genes required for assembly of CS2 pili is different from the majority of other colonisation factors (with the exception of CS1) which are plasmid-located (Perez-Casal et al., 1990).

The genes for CS2 biogenesis have been cloned and sequenced and are designated *cotB, cotA, cotC* and *cotD* (Froehlich et al., 1995)(Figure 1.4). The CotB protein shows significant homology with CooB and CfaA from CS1 and CFA/I pili, respectively, and is predicted to function as a chaperone. The major subunit CotA also shares homology with the corresponding pilins CooA and CfaB from CS1 and CFA/I pili, respectively. CotC and CotD also share homology with the outer membrane assembly proteins CooC and CfaC and the minor pilins CooD and CfaE from CS1 and CFA/I pili, respectively (Froehlich et al., 1995)(Figure 1.4). As with CS1, CS2 pili are also positively regulated by the plasmid located gene *rms* (Caron et al., 1989). The similarity between CS1 and CS2 pili is also reflected in the
fact that the genes for CS2 pili are interchangeable with those from CS1 pili. CotA and CotB can interact with CooC and CooD to produce cell-surface-assembled pili, and CooA and CooB can interact with CotC and CotD to form pili (Froehlich et al., 1995).

1.7.1.2.3 CS3

CS3 are wiry fibrillar structures approximately 2-4 nm in diameter (Levine et al., 1984) which are able to mediate attachment of ETEC bacteria to human enterocytes (Knutton et al., 1985). CS3 can induce a significant immune response in human volunteers who developed diarrhoea against CS3-expressing wild type ETEC in a clinical setting (Levine et al., 1984). Furthermore, animal studies have shown that monoclonal antibodies against CS3 are able to protect against experimental infection in rabbits (Svennerholm et al., 1990).

The genetic region required for biosynthesis of CS3 pili has been cloned and completely sequenced and is shown in Figure 1.6 (Boylan et al., 1988; Boylan et al., 1987; Jalajakumari et al., 1989; Manning et al., 1985). Pilus production appears to be regulated by suppression or read-through of an internal amber codon (UAG) within the cstG gene, which then allows expression of the cstH gene which encodes the major pilin subunit (Jalajakumari et al., 1989). The 15 kDa major pilin subunit, CstH was not significantly homologous at the protein level to a number of the characterised pilins, although morphologically the pili appear to be very similar to K88 and K99 (Moylan et al., 1988). Sequence analysis suggests that four genes (cstA, cstB, cstG and cstH) lie contiguously along the DNA with no overlap, while four further ORFs (cstC, cstD, cstE and cstF) are completely contained within the cstB gene (Jalajakumari et al., 1989). The periplasmically located CstA protein shows homology to the PapD chaperone of Pap pili (Lindberg et al., 1989), CstB shows homology with the outer membrane assembly protein PapC (Norgren et al., 1987) and CstG shows homology with the PapH minor pilin subunit involved in cell anchoring and length modulation (Båga et al., 1987). More recently, the features of CS3 pili have been utilised for the insertion of foreign epitopes including the mature form of the ST toxin and a conserved epitope of the B-subunit.
Figure 1.6 Genetic organisation of the cst locus encoding biosynthesis and assembly of CS3 fibrillae.

The boxes correspond to the proposed cst genes with the promoters, terminator and direction of transcription (arrowheads) also indicated (from Yakhchali and Manning, 1996).
Assembly cassette

Promoters

Major fimbrial subunit

Terminator
of LT toxin (Staendner, L. S., T. G. Duthy, P. A. Manning and M. W. Heuzenroeder, unpublished results; Yakhchali and Manning, 1996).

1.7.1.3 CFA/IV

The PCF8775 complex, now known as colonisation factor antigen IV (CFA/IV), was first described in ETEC strains of serogroups O25, O115 and O167 (Thomas et al., 1985), and is present on the surface of 10-20% of all ETEC strains. The pili produced are antigenically distinct, but consist of combinations of CS4, CS5 and CS6 (Thomas et al., 1985). CFA/IV ETEC strains express the non-pilus antigen CS6 with either CS5 or CS4 pili. There have been no reports of isolates expressing CS4 or CS5 in the absence of CS6. ST- and LT-expressing strains belonging to the serogroup O25 produce CS4 and CS6 whereas CS5 and CS6 are found on ST-expressing strains of serogroups O6, O29, O92, O114, O115 and O167, and strains of serogroups O25, O27, O92, O148, O153, O159 and O169 produce CS6 only (McConnell et al., 1985; Thomas et al., 1985; Thomas and Rowe, 1982). In O167 strains, a single plasmid encodes for CS5, CS6 and ST (Thomas et al., 1987).

Specific pilus staining and immunoelectron microscopy experiments have clearly shown that CS4 and CS5 can promote adhesion to human duodenal mucosa, while strains of serotype O27:H7, O27:H20, O148:H28 and O159:H20 expressing CS6 antigen showed good adhesion to human enterocytes (Knutton et al., 1989).

1.7.1.3.1 CS4

CS4 pili are approximately 6-7 nm in diameter and resemble CFA/I, CS1 and CS2 in morphology (Knutton et al., 1985; Levine et al., 1984). The N-terminal amino acid sequences of the four subunits are also similar (McConnell et al., 1989b). Immunoblotting techniques have revealed that CS4 is closely related to CFA/I (Gaastra and Svennerholm, 1996). In ETEC strains of serotype O25:H42, the structural and regulatory genes for CS4 are encoded on two separate plasmids, with a third plasmid encoding enterotoxin production (Willshaw et
al., 1990). The regulatory gene for CFA/I expression, cfaD/cfaR, was able to regulate the expression of CS4 pili (Willshaw et al., 1990). The CS4 regulatory gene has been cloned and gave an identical restriction pattern to cfaD/cfaR, and was able to promote expression of CS1, CS2 and CFA/I pili (Willshaw et al., 1991). At the molecular level, little is known about the genes involved in CS4 biosynthesis as the region has yet to be cloned.

1.7.1.3.2 CS5

CS5 pili have been previously shown to be approximately 7 nm in diameter, and can mediate mannose-resistant hemagglutination (MRHA) of human group A, guinea pig and bovine erythrocytes (Thomas et al., 1985). Electron microscopy has identified the CS5 pilus as a semi-rigid fibrillar structure (Heuzenroeder et al., 1989), and not rigid as first proposed (Thomas et al., 1985). Others suggest that CS5 is composed of two fine fibrils arranged in a double helical structure approximately 5-6 nm in diameter (Knutton et al., 1989).

The DNA region required for the biosynthesis of functional CS5 pili was cloned from an ETEC strain belonging to the O115:H40 serotype, originally isolated during an outbreak of diarrhoea among Aboriginal children in central Australia (Heuzenroeder et al., 1989; Manning et al., 1987). Transposon mutagenesis identified the minimal coding region of the cloned DNA required for CS5 biosynthesis to be approximately 7 kb in size (Heuzenroeder et al., 1989). Expression in minicells of the cloned genes showed that proteins of 70 kDa and 46.5 kDa co-purified with the 23 kDa major subunit and proteins of sizes 45, 31, 17 and 14 kDa were also expressed (Heuzenroeder et al., 1989).

In subsequent nucleotide sequencing of the region encoding the major CS5 pili subunit, significant homology with the corresponding major subunit of the F41 pili of porcine ETEC (Moseley et al., 1986) was shown within the signal sequence and at the carboxy terminus, as shown in Figure 1.7 (Clark et al., 1992). The CS5 major subunit was predicted to encode a mature protein of 18.6 kDa based on DNA sequence analysis; however, sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) showed the mature subunit
Figure 1.7  Protein sequence homology between the CS5 major pilin subunit and the F41 major pilin subunit.

Dots below the line correspond to similar amino acids and the asterisks correspond to identical amino acids (from Clark et al., 1992).
CS5  MKKNLLILTVLAMATVGSVLAAT-TNGQLTFNWQGVPASPVTQSSWAFVNGLDIPFTP
F41  MKKTLIALAVAAAASAVGSVMADWTEGPQGDIIGGEITSPSVKMLWKTGEGLS-SFSN
***...*...*...*...*...*...*...**

CS5  GTEQ----LNITLDSNKIDARTSVKPI-YDPFI----VP------------VSIGNVTPGA
F41  TTNEIVRKLNLISVPYDLFLAAKMSDGIKGVFVQMTLPIKNEASDGYDGSVITPSFTSNT
...*...***...*...*...*...*...*...*...*...*...*...*...*...

CS5  PVTRDTSANINSVNAFLSSVFSVNGFVGCI--------KQLTLSTAVEAAKEVAILN
F41  AMDIAVTKKSNDTNLGTKLVSPLSFNGAVATIFDFGDTTDTSAAHIIGGSATVFGFGLVN
...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...

CS5  QA---------------------------LKVGSAEPTVTVT-------------ASNKKKH
F41  PGRFTDQNIAYKWNLGSKAEMAGYVEKLMPGQSASYSGMFWDDLSHSHWNTSANKASY
...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...

CS5  ISIDMNAKAAAADVAG--------GAAIKFVAPVTFAVDI-
F41  LSYGSGYSAGSTLVZNLNKDVAGRLDWAPVTITVIYS
...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*...*
had an apparent size of 23 kDa (Clark et al., 1992). Immediately downstream of the CS5 subunit gene a potential stem-loop structure was identified with a free energy (ΔG) of -88.3 kJ/mol. This may function as a transcriptional terminator or attenuator, thereby reducing the expression of downstream genes (Clark et al., 1992). CS5 is the most hydrophobic of any ETEC pili examined, and is readily precipitated in less than 5% ammonium sulphate (Heuzenroeder et al., 1989).

Based on homology to the CFA/I regulator cfaD/cfaR, a regulatory gene has been cloned from ETEC strain O167:H5 which produces CS6 antigen and CS5 pili (Willshaw et al., 1991). This proposed plasmid-encoded regulatory gene, csvR is 87% homologous with cfaR and is able to functionally replace cfaD/cfaR to mediate CFA/I expression (de Haan et al., 1991). However, the exact role of csvR in regulating expression of CS5 pili is not known.

1.7.1.3.3 CS6

CS6 has been implicated as a colonisation factor since ETEC strains expressing only CS6 are able to colonise rabbit intestines equally as well as strains carrying CS4 and CS6 or CS5 and CS6 (Svennerholm et al., 1988). Such strains have been identified as the only pathogen in patients suffering from diarrhoea (Binsztein et al., 1991; Wolf et al., 1993). However, CS6 has not been identified as a distinct pilus structure, and is unable to mediate MRHA of human or bovine erythrocytes (Thomas et al., 1985), although CS6-expressing ETEC have been shown to bind human enterocytes in vitro (Binsztein et al., 1991; Helander et al., 1997a; Helander et al., 1997b; Wolf et al., 1993). Production of monoclonal antibodies against CS6 were able to specifically bind along the entire surface of CS6-expressing bacteria, but no structure was identified (Helander et al., 1997a). CS6 is one of the most frequently identified colonisation factors identified in clinical isolates (Gaastra and Svennerholm, 1996).

Cloning of the operon encoding for CS6 biosynthesis showed that the region consisted of four genes, cssA, cssB, cssC and cssD (Cassels and Wolf, 1995). Based on homology searches, CssC belongs to the family of molecular chaperones and CssD is the outer
membrane assembly protein. CsaA and CsbB encode two heterologous major subunit proteins (Cassels and Wolf, 1995; Wolf et al., 1997), which makes CS6 unique, since the majority of all colonisation factors encode for only a single major subunit. Antibody production studies in mice have indicated that CsbB is the most immunogenic of the two structural subunits (Helander et al., 1997b). The heterogeneity of the CS6 antigen also explains the previous reports of the CS6 antigen existing as two distinct variants of 14.5 and 16 kDa (McConnell et al., 1988; Wolf et al., 1989).

1.7.1.4 CS7 and CS8

CS7 pili were initially identified on an ST and LT expressing ETEC strain of serotype O15:H11, but are also present on serotypes O103:H49 and O114:H49, which also express ST and LT enterotoxins (Hibberd et al., 1990). CS7 pili are morphologically similar to CS5 pili with a helical structure approximately 3-6 nm in diameter. However, antisera against both pili types are only weakly cross-reactive. The major pilin subunit of CS7 is approximately 21.5 kDa. Strains expressing CS7 are able to mediate adherence of ETEC to both CaCo-2 cells and human enterocytes (Viboud et al., 1996).

CS8 or CFA/I are rod-like pili of about 7-8 nm in diameter which exhibit flexibility (Hibberd et al., 1990; Knutton et al., 1989). CS8 production is limited to a LT-producing O25:H- serotype ETEC strain, which also produces the CS6 antigen (Honda et al., 1984). A 21 kb DNA fragment containing the genes necessary for production of CS8 has been cloned (Shinagawa et al., 1993) and the sequence of the major subunit cofA determined (Taniguchi et al., 1995).

The cofA gene encodes a 26.5 kDa protein which is the precursor of CS8 pilin which, after post-translational cleavage, forms the mature 20.5 kDa pilin. The N-terminal 30 amino acid sequence of the mature CS8 pilin subunit is highly hydrophobic and shows strong homology with the type IV class B pilin subunit TcpA (toxin co-regulated pilus) from Vibrio cholerae (Taniguchi et al., 1995). Furthermore, CofA has been shown to be highly
homologous to the LngA structural protein of another ETEC type IV pili identified, longus (CS21) (Giron et al., 1994; Gomez-Duarte et al., 1999; Taniguchi et al., 1995).

1.7.1.5 CS10 (Antigen 2230), CS11 (PCFO148) and CS12 (PCFO159)

CS10 or antigen 2230 was originally identified as a non-fimbrial antigen of 16 kDa on the surface of ST expressing ETEC strains of the serotype O25:H16 (Darfeuille-Michaud et al., 1986; Forestier et al., 1987). CS10 is unable to mediate haemagglutination, while the genes encoding CS10 are located on a 66 MDa plasmid (Forestier et al., 1987).

CS11 or PCFO148 are thin curly fibrillae approximately 3 nm in diameter that are expressed by ETEC strains of serotype O148:H28 (Knutton et al., 1987). Binding studies in vitro have indicated that strains expressing CS11 are able to adhere to human enterocytes implicating a role for CS11 as a colonisation factor (Knutton et al., 1987).

CS12 or PCFO159 is present on ETEC strains of serotype O159:H4 which express both ST and LT enterotoxins (Tacket et al., 1987). CS12 is approximately 7 nm in diameter, while rabbit antiserum raised against purified CS12 pili identifying the major subunit as a 19 kDa protein (Tacket et al., 1987). The genes encoding both enterotoxins and CS12 are located on a 27 MDa plasmid. In vivo studies in rabbits showed that CS12-expressing bacteria were able to effectively colonise and elicit a significant immune response, which strongly suggests that CS12 is a colonisation factor (Svennerholm et al., 1992).

1.7.1.6 CS13

The genes encoding CS13 pili, or PCFO9 as it was formerly known, were cloned from an LT-expressing ETEC strain serotype O9:H- from central Australia (Heuzenroeder et al., 1990). Electron microscopy has revealed that CS13 is a thin flexible fibrillar structure (Heuzenroeder et al., 1990). The major subunit gene cshE encodes a 27 kDa protein which bears homology with the corresponding subunit gene in K88 fimbriae (Clark, C.A., and P. A. Manning, unpublished results). Indeed, the genetic organisation of CS13 is almost identical to
that of K88 with the genes cshA, cshF and cshG encoding minor pilins and cshB, cshC and cshD encoding the outer membrane assembly protein and periplasmic chaperones, respectively (Clark, C. A., and P. A. Manning, unpublished results). Interestingly, the CS13 outer membrane assembly protein CshB bears 97% identity with the corresponding K88 protein FaeD (Mooi et al., 1986), but the two proteins cannot functionally replace one another, which indicates that small changes in sequence can effect subunit specificity.

1.7.1.7 CS14 (PCFO166) and CS15 (Antigen 8786)

CS14 pili or PCFO166, was identified as a 7 nm pili on the surface of ETEC strain O166:H27 which produces ST (McConnell et al., 1989a). Antiserum raised against CS14 has also shown this colonisation factor is present on serogroups O20, O71 and O98 and CS14 is able to mediate MRHA of bovine and human erythrocytes (McConnell et al., 1989a). CS14 was not able to cause colonisation of the rabbit intestine (Svennerholm et al., 1992), but in vitro studies have shown that it is responsible for adhesion of ETEC to human enterocytes and CaCo-2 cells (McConnell et al., 1989a; Viboud et al., 1996).

Two major subunits have been detected by western immunoblot, which have been designated csuA1 and csuA2 and encode proteins of 15.5 and 17 kDa respectively (Gaastra and Svennérholm, 1996; McConnell et al., 1989a). CS14 is genetically related to CFA/I and CS4 pili, since the N-terminal amino acid sequence of CS14 pilin subunit is highly homologous with the corresponding sequences of CFA/I and CS4 (Sommerfelt et al., 1992). Furthermore, CS14 expression can be considerably increased when the gene which encodes the positive regulator for CFA/I expression (cfaD/cfaR) is introduced in trans (Sommerfelt et al., 1992). Likewise, gene sequences from CS14 positive strains can substitute for the cfaD/cfaR gene to induce CFA/I pilin production (Hibberd et al., 1991).

CS15 or antigen 8786 is a non-fimbrial colonisation factor originally isolated from an ST expressing ETEC strain serotype O117:H4 in Africa (Aubel et al., 1991). The antigen was identified as a 16.3 kDa protein. This antigen was shown to mediate adherence of the bacteria
to both brush border membranes of human enterocytes and to CaCo-2 cells, indicating the potential role of CS15 as a colonisation factor. The gene encoding the CS15 major subunit is located on a 70 MDa plasmid (Aubel et al., 1991).

1.7.1.8 CS17 and CS18 (PCFO20)

CS17 are rigid pili with a diameter of approximately 7 nm, and were identified on LT-producing ETEC strains belonging to the serogroups O8, O15, O48, O114, O146 (Deneko et al., 1981; McConnell et al., 1990). The CS17 major subunit pilin, called CsbA is a 17.5 kDa protein. In vitro binding studies suggest that bacteria producing CS17 did not adhere to both human enterocytes or CaCo-2 cells (Viboud et al., 1996b). Conversely, CS17 has been shown to promote colonisation of rabbit intestines in vivo (Svennerholm et al., 1992). Experiments have shown that expression of CFA/I can be induced when a plasmid containing only the operon responsible for pilus assembly is introduced into a CS17 expressing strain (Hibberd et al., 1991). This shows that a homologue of the regulator of CFA/I expression (cfaD/cfaR) is present in CS17 strains. Furthermore, CS17 pili shares antigenic homology with CFA/I, CS1, CS2, and CS4 (McConnell et al., 1989b).

CS18 or PCFO20 was initially identified from an O20:H- ETEC isolate, producing both ST and LT enterotoxins, from a child suffering from diarrhoea in Argentina (Viboud et al., 1993b). CS18 are 7 nm in diameter and are able to mediate binding to CaCo-2 cells. N-terminal sequence analysis of the 25 kDa major pilin subunit showed homology with porcine ETEC 987P pili (Viboud et al., 1993b).

Subsequent cloning and sequencing of the major pilin subunit, fotA, indicated that this protein shares 82% similarity with the porcine 987P major subunit (Viboud et al., 1996a). Furthermore, FotA was assembled and transported by the 987P machinery in porcine ETEC strain 987; however, the subunits have affinity for different host cell receptors, since CS18-producing bacteria do not bind to neonatal piglet enterocytes (Viboud et al., 1996a). A second gene, fotB, was identified immediately downstream of fotA and was predicted to encode a
periplasmic chaperone, based on strong homology with the PapD chaperone from the Pap system (Viboud et al., 1996a).

1.7.1.9 CS19 and CS20

CS19 was originally identified from the ETEC strain F595C O8 serogroup, which produces both ST and LT, on the basis of weak hybridisation to a gene probe derived from the CFA/I operon (Grewal et al., 1997). CS19 has since been shown to be expressed on ST- and LT-producing ETEC strains of the O114:H- serotype (Khalil et al., 1999). CS19 is a 7 nm diameter pili which is composed of a 16 kDa major subunit known as CsdA (Grewal et al., 1997). CS19 did not mediate MRHA on a panel of erythrocytes. However, strain F595C was able to bind CaCo-2 cells, and this binding could be inhibited by the addition of antiserum against CsdA (Grewal et al., 1997). N-terminal sequence data of CsdA shows a high degree of identity with the corresponding subunits of CFA/I, CS1, CS2, CS4 and CS14. The N-terminal sequence was identical to the CS17 subunit CsbA. Furthermore, the CFA/I regulatory gene cfaD/cfaR was able to mediate a significant increase in expression of CS19 pili which implies both CFA/I and CS19 share common regulatory mechanisms for expression (Grewal et al., 1997).

CS20 pili were first identified from a ST- and LT- expressing ETEC strain of serotype O17:H45 from a patient with diarrhoea in New Delhi, India (Valvatne et al., 1996). CS20 are approximately 7 nm in diameter and are able to specifically mediate binding of the bacteria to CaCo-2 cells in vitro. N-terminal amino acid sequence analysis of the 20.8 kDa pilin subunit shows homology to the PCFO20 (CS18) and 987P major subunits of human and porcine ETEC respectively (Valvatne et al., 1996). As with other colonisation factors, the introduction of the cfaD/cfaR regulatory gene significantly increased expression of CS20 pili (Valvatne et al., 1996).
1.7.1.10 Longus (CS21)

Longus or CS21 is a long pilus structure which was initially discovered on a human clinical ETEC isolate of serotype O8:H9, which produces LT, ST and CS3 only, by growing the strain aerobically on anaerobe blood agar (Giron et al., 1994). Longus filaments are approximately 7 nm in diameter, with a semi-flexible architecture and over 20 μm long. Unlike CSs or CFAs, longus filaments do not emanate peritrichously from the cell surface, but have a polar location (Giron et al., 1994). The longus structural gene lngA encodes a 22 kDa protein with striking N-terminal homology with the toxin-coregulated pilus (TCP) of Vibrio cholerae, the bundle forming pilus of enteropathogenic E. coli (Giron et al., 1994), and CS8 (CFA/III) pili (Gomez-Duarte et al., 1999). Longus is therefore classified into the class B type IV pilus family (Giron et al., 1995). DNA hybridisation experiments using lngA as a probe showed that longus is a widely distributed antigenic determinant in ETEC that is associated with the known plasmid-encoded CFAs, and enterotoxins (Giron et al., 1995). In particular, a high association of LngA was identified with CFA/II, and was more often associated with serogroups producing ST alone (Giron et al., 1995).

1.7.1.11 CS22

Very recently, a novel adhesin termed CS22 was identified on an ETEC O20:H- strain from Argentina (Pichel et al., 2000). The major subunit was identified as a 15.7 kDa protein, and antiserum raised against this protein inhibited the wild type ETEC strain from adhering to CaCo-2 cells (Pichel et al., 2000). Immunogold electron microscopy showed that the antiserum raised against the 15.7 kDa protein reacted against thin fibrillar-like structures on the cell surface, that were absent from a derivative strain which lacked production of the 15.7 kDa protein. N-terminal sequence analysis revealed that CS22 is 95% homologous to CS15 (antigen 8786) and 65% homologous to SEF14 fimbriae of Salmonella enterica servoar Enteritidis (Pichel et al., 2000). Both CS22 and CS15 are immunologically related, and DNA sequence analysis of the major subunit of CS22 showed it differed by only 30 residues from
the DNA sequence of the CS15 major subunit (*nfaA*). CS22 was found on 60% of the O20:H-serotype ETEC strains from Argentina, and was restricted to this serotype (Pichel *et al*., 2000).

### 1.7.2 Antigenicity of Human ETEC Colonisation Factors

ETEC pili have been shown to be antigenically distinct by using antisera or monoclonal antibodies against whole bacteria or purified pili in enzyme-linked immunosorbent assays (ELISA) or immunodiffusion techniques (McConnell *et al*., 1985). However, some colonisation factors are immunologically cross reactive. Antibodies raised against CS7 have been shown to be weakly cross-reactive with CS5 pili, which is morphologically very similar to CS7 (Hibberd *et al*., 1990). In ELISA and immunodiffusion studies CFA/I, CS1, CS2, CS4 and CS14 were antigenically distinct (McConnell *et al*., 1989; McConnell *et al*., 1990). However, monoclonal antibodies raised against isolated CFA/I subunits have been shown to be immunologically cross-reactive with CS1, CS2, CS4, CS14 and CS17 pili (Rudin *et al*., 1994). Furthermore, these antibodies were able to inhibit the binding of ETEC strains expressing these pili to both erythrocytes and CaCo-2 cells (Rudin *et al*., 1994). More detailed analyses showed that the monoclonal antibodies bound specifically to CS4 pili expressed on bacteria and were able to prevent CS4-mediated adherence to human jejunal enterocytes (Rudin *et al*., 1996). One of the monoclonal antibodies recognised an epitope at the N-terminus of the protein, which indicates that antibodies generated against N-terminal epitopes may be anti-adhesive and protective (Rudin *et al*., 1994; Rudin *et al*., 1996).

The level of cross-reactivity between CFA/I, CS1, CS2, CS4, CS14 and CS17, is not surprising given that the N-terminal regions are very similar (Hibberd *et al*., 1991; Karjalainen *et al*., 1989; Klemm *et al*., 1985; McConnell *et al*., 1989).

The different colonisation factors have now been divided into four groups, based on genetic and antigenic relationships (Gaastra and Svennerholm, 1996). The CFA/I-like group consists of CFA/I, CS1, CS2, CS4, CS14, CS17 and CS19. The ‘CS5’ group consists of CS5,
CS7, CS13, CS18 and CS20. The type IV-like group consists of CS8, longus (CS21), CS15 and CS22. Finally, the distinct group consists of CS3, CS6, CS10, CS11 and CS12. These groups are shown in Table 1.1.

1.7.3 Specific Growth Conditions and Pilus Expression

Growth conditions and the presence of certain compounds can drastically effect expression of ETEC pili. The expression of all pilus types identified (with the exception of type I pili) are temperature sensitive, with pili expression abolished at room temperatures. Thermoregulation studies have shown that a reduction in temperature effects the level of transcription of pilus operons by controlling promoter activity (van der Woude et al., 1992; White-Ziegler et al., 1998; White-Ziegler et al., 1990). The thermoregulatory response can be rapid (White-Ziegler et al., 1998). Catabolite repression has been shown to play a role in the synthesis of CFA/I and CFA/II pili (Evans et al., 1991; Karjalainen et al., 1991; White-Ziegler et al., 1998). CFA/I expression is inhibited when exogenous iron is added to the growth medium, with the global metalloregulatory protein, Fur, implicated in the specific repression of the CFA/I subunit gene (Karjalainen et al., 1991). The presence of sodium chloride in the growth medium also abolishes expression of CS3 and CS5 pili in wild type ETEC strains (Staendner, L. H. and T. G. Duthy, unpublished observations). Pili expression can be enhanced by growth on CFA agar, which may be improved further by incorporation of bile salts into the CFA agar (Evans et al., 1975; McConnell et al., 1989b).

1.7.4 Receptor Identification

Very little is known about the host receptors for the various human ETEC colonisation factors, apart from studies involving CFA/I and the CS1, CS2 and CS3 sub-components of CFA/II. CFA/I has been shown to bind free sialic acid and a 26 kDa glycoprotein containing important sialic acid moieties present on human erythrocytes (Evans et al., 1979; Pieroni et al., 1988). Reverse haemagglutination assays have also been used to study CFA/I mediated
adherence via sialoglycoconjugates (Bartus et al., 1985). Pre-treatment of human erythrocytes with neuraminidase, trypsin, chymotrypsin, papain and protease all abolished agglutination activity, along with urea and guanidine. However, compounds affecting disulphide or sulphhydral moieties did not alter reactivity. Furthermore, it was shown that sialic acid containing glycoconjugates were specifically able to inhibit CFA/I-mediated haemagglutination, as was an isolated sialoglycoprotein from the erythrocyte membrane (Bartus et al., 1985). One such sialoglycoconjugate, GM_2, has been shown to bind CFA/I (Faris et al., 1980).

Studies on CFA/I and CFA/II binding to electrophoretically-separated membrane components of rabbit intestinal brush borders or human intestinal and non-intestinal cell lines suggest that both pilus types bound to 30-35 kDa structures (Wennerås et al., 1990). Furthermore, CS2 and CS3 from CFA/II were able to bind to cell membrane components of the range 35-75 kDa, while CS3 was also able to bind strongly to structures of 120-140 kDa present in the rabbit intestinal cells, but not tissue-cultured cells (Wennerås et al., 1990). Chloroform-methanol extraction of lipids from the cell membranes did not alter CFA/I- or CFA/II-binding specificity which suggests that binding occurs via glycoproteins rather than glycolipids (Wennerås et al., 1990). CS1, CS2 and CS3 of CFA/II and CS4 of CFA/IV also have been shown to bind asialo-GM_1 (Oro et al., 1990). Further binding studies on CS3 pili have shown that the critical binding epitope of CS3 consists of the carbohydrate sequence GalNAcβ1-4Gal and that GM_1, asialo-GM_1, GM_2 and to some extent GM_3 all inhibit CS3 binding (Wennerås et al., 1995).

1.7.5 Distribution of Human ETEC Colonisation Factors

The most common colonisation factors encountered world wide are CFA/I, CFA/II (CS1, CS2 and CS3) and CFA/IV (CS4, CS5 and CS6), and the relative frequency with which each of these antigens occurs is dependent on the geographical location. The prevalence of toxin types and colonisation factors during a two year field study in Bangladesh showed that
56% of the samples were positive for one or more of the twelve colonisation factors examined 
(Qadri et al., 2000a). CFA/IV strains were the most prevalent (31%) followed by CFA/I 
(23.5%) and CFA/II (21%). CS7, CS14, CS12, CS17 and CS8 were all detected in only a 
small % of total isolates (Qadri et al., 2000). A study in Burma revealed that CFA/I was the 
most commonly detected colonisation factor detected (38%) followed by CFA/IV (11%), 
CFA/II (4%), CS7 (10%), CS17 (4%) and CS12 and CS14 (2%)(McConnell et al., 1991). 
Studies in Thailand have shown that CFA/I was the most common colonisation factor 
observed (Sunthadvanich et al., 1990).

Two studies have been performed in Africa. In central Africa (Zaire and Rwanda) 
CFA/IV was the most common colonisation factor encountered (33%) followed by CS17 
(24%) and 12% for CFA/I (McConnell et al., 1991). ETEC was also examined from a 
community-based study of paediatric diarrhoea in Egypt (Peruski et al., 1999). CFA/IV was 
again the most common colonisation factor identified (43%) followed by CFA/II (22%), 
CFA/I (13%), CS14 (13%) and CS7 (9%). No strains isolated expressed CS8, CS17 or CS12 
(Peruski et al., 1999).

In a cohort study in children from Argentina, CFA/I was the most commonly 
encountered colonisation factor (23%) followed by CFA/IV (17%), CFA/II (12%), CS14 
(9.5%), CS17 (6.7%), CS7 (5.7%) and 3.8% for CS12 (Binsztein et al., 1991; Viboud et al., 
1993a; Viboud et al., 1999). These colonisation factors were associated with ETEC strains 
that expressed either ST alone or ST and LT enterotoxins.

1.8 Pap Pili from Uropathogenic E. coli

The genetics of Pap (pili associated with pyelonephritis) pili or P-fimbriae from 
uropathogenic E. coli has been extensively studied, and is the best understood system for 
examining pilus biogenesis. Pap are expressed on the surface of uropathogenic strains of E. 
coli associated with acute pyelonephritis (Kallenius et al., 1981). Pap overcome any initial 
barriers to infection by expressing an adhesin molecule on the tip of the pilus structure, which
specifically binds to the α-D-galactopyranosyl-(1-4)-β-D galactopyranoside [Galα(1-4)Gal] moiety present in the globoseries of glycolipids on the uroepithelium cells lining the urinary tract (Eden and Leffler, 1980). This moiety is also expressed on erythrocytes and many other tissues, including the uroepithelial cells of the kidney. The host specificity of uropathogenic *E. coli* depends on the differences in binding specificity to Galα(1-4)Gal-containing isoreceptors (Stromberg *et al.*, 1990).

Pap is a composite structure composed of a flexible tip fibrillar structure joined end-to-end to pilus rods (Kuehn *et al.*, 1992). The pilus rod is composed of repeating major subunits 5-7 nm in diameter packed into a right-handed helical assembly, with X-ray diffraction experiments revealing an external diameter of 68 Å, an axial hole of 15 Å and 3.28 subunits per turn of the helical cylinder (Bullitt and Makowski, 1995; Gong and Makowski, 1992). The flexible tip fibrillum is approximately 2 nm in diameter, and this flexibility may allow the binding adhesin located at the tip of the fibrillum to interact with the digalactoside moiety on the uroepithelium (Kuehn *et al.*, 1992).

### 1.8.1 Gene Function

The genes required for biosynthesis of Pap are located in clusters at different sites of the *E. coli* chromosome. These genes have been cloned from the urinary tract *E. coli* isolate J96 (Hull *et al.*, 1981) and characterised (Båga *et al.*, 1987; Båga *et al.*, 1984; Lindberg *et al.*, 1986; Lindberg *et al.*, 1984; Lund *et al.*, 1987; Norgren *et al.*, 1987; Norgren *et al.*, 1984; Tennent *et al.*, 1990). The DNA sequence of the entire *pap* gene cluster encodes eleven genes, known as *papA-K* (Figure 1.8). The genes for *papI* and *papB* encode two regulatory proteins responsible for expression of Pap (Båga *et al.*, 1985; Goransson *et al.*, 1989a), PapA is the major subunit (Båga *et al.*, 1984), PapC is the outer membrane usher protein (Norgren *et al.*, 1987), PapD is the periplasmic located chaperone protein (Lindberg *et al.*, 1989), PapH is a minor subunit involved in cell anchoring and length regulation (Båga *et al.*, 1987), the role of PapJ is still somewhat controversial, although it may possibly act as a second chaperone
Figure 1.8 Genetic organisation of the *pap* gene locus required for biosynthesis of Pap pili from uropathogenic *E. coli*.

(from Thanassi et al., 1998a)
P pilus (pap) gene cluster

- Regulation
- Major pilus subunit
- Pilus anchor
- Outer membrane usher
- Periplasmic chaperone
- Adaptor/initiator
- Major tip component
- Galβ(1-4)Gal-binding adhesin

Tip fibrillum components
(Tennent et al., 1990), and PapG is the tip located adhesin responsible for binding digalactoside moiety on the urogenital epithelium (Kuehn et al., 1992; Lund et al., 1987). The genes of papE, papF and papK all encode minor pilins which are responsible for forming the tip fibrillae (Kuehn et al., 1992; Lindberg et al., 1986; Marklund et al., 1992).

1.8.2 Regulation of Expression

As previously mentioned, the two regulatory proteins encoded by the pap operon are PapI and PapB, which share extensive homology with the regulatory proteins FaeA and FaeB from K88 pili, DaaF and DaaA from human uropathogenic E. coli F1845 pili, and SfaC and SfaB from human uropathogenic E. coli S pili, respectively (Mol and Oudega, 1996). Expression of the pap operon is also regulated by other bacterial global regulatory proteins including H-NS (White-Ziegler et al., 1998), Lrp (Braaten et al., 1992; van der Woude et al., 1992), and Crp (Goransson et al., 1989b), but also the deoxyadenosine methylase (Dam) protein (Bllyn et al., 1990). Pap pili are subject to phase variation; however, unlike the phase variation observed in type I pili, no DNA rearrangements occur and the phase ON and phase OFF state is mediated by the methylation status of DNA within the regulatory region resulting in transcription or non-transcription of the Pap pilin (papA) gene (Bllyn et al., 1990)(Figure 1.8). It has been suggested that phase variation may allow the bacteria to escape immune responses (Hultgren et al., 1991).

Pap have also been shown to be under thermo-regulation. At 37°C Pap expression is turned on, while at temperatures below 26°C it is shut off (Blyn et al., 1989; White-Ziegler et al., 1990). H-NS has been shown to be a temperature-dependent DNA methylation blocking factor which is responsible for turning off Pap expression at temperatures below 26°C (White-Ziegler et al., 1998). This response was rapid and occurred within only one generation of growth and required H-NS and DNA sequences within papB, but the regulatory proteins PapB and PapI were not required (White-Ziegler et al., 1998). H-NS acts by binding to both GATC-I and GATC-II methylation sites to prevent methylation, thereby serving as a
methylation blocking factor to repress *pap* gene expression, through formation of a specific nucleoprotein complex, which occurs at less than 26°C (White-Ziegler *et al.*, 1998).

### 1.8.3 The PapG Adhesin

PapG is a 35 kDa minor pilin subunit, which is essential for the binding properties of Pap pili (Lindberg *et al.*, 1984; Lund *et al.*, 1987). The PapG protein is the α-D-galactopyranosyl-(1-4)-β-D-galactopyranoside [Galα(1-4)Gal] binding adhesin, which is located the distal tip of the flexible tip fibrillae of Pap, which also contains PapE, PapF and PapK (Jacob-Dubuisson *et al.*, 1993; Kuehn *et al.*, 1992; Lindberg *et al.*, 1986; Lund *et al.*, 1987)(Figure 1.9). A deletion in the *papG* gene does not affect tip fibrillae formation, but does abolish receptor binding (Kuehn *et al.*, 1992). The PapG adhesin consists of two domains (Hultgren *et al.*, 1989). The N-terminal domain contains the receptor binding site, while the C-terminal domain of PapG contains the PapD chaperone recognition surface (Hultgren *et al.*, 1989). A PapD-PapG complex does not interfere with the receptor binding N-terminal domain of PapG, such that PapG still retains same receptor binding activity in this pre-formed state, when compared to the fully assembled PapG at the tip of the pilus (Hultgren *et al.*, 1989). The specific interaction between PapG and the β-galabiose portion of the globoseries of glycolipids has been studied (Kihlberg *et al.*, 1989). A specific hydrogen bond interaction occurs between the PapG adhesin and five O₂ atoms situated on the edge of the disaccharide, and hydrophobic interactions with a non-polar cavity of the adhesin itself (Kihlberg *et al.*, 1989).

There are three characterised alleles or variants of the PapG adhesin, G-I, G-II and G-III (Stromberg *et al.*, 1990). The three adhesins recognise three different Gala(1-4)Gal-containing isoreceptors, known as globotriaosylceramide, globotetraosylceramide (globoside) and globopent-aosylceramide (Forsman antigen), respectively (Stromberg *et al.*, 1990). The dominant allele is the G-II expressing uropathogenic *E. coli* strains.
Figure 1.9  Simplified illustration of the structure and assembly of the Pap pilus from uropathogenic *E. coli*.

Nacently folded pilin subunits PapG (G), PapF (F), PapE (E), PapK (K), PapA (A) and PapH (H) are protected from periplasmic degradation by PapD (D) which delivers the subunits to the outer membrane usher protein PapC for translocation across the outer membrane, to form the mature pilus (from Soto and Hultgren, 1999).
1.8.4 The Pap Pilins- PapA, PapE, PapK, PapF and PapH

PapA is a 19.5 kDa protein identified as the major Pap pilin subunit (Båga et al., 1984). The bulk of the pilus fibre is composed of approximately 1000 copies of PapA repeating monomers, which forms a rigid rod-like structure approximately 5-7 nm in diameter (Båga et al., 1984)(Figure 1.9). A specific mutation in the papA gene results in the abolition of visible cell surface Pap pili; however, the bacteria still retains haemagglutination activity, suggesting receptor binding was not abolished (Uhlin et al., 1985).

Three-dimensional reconstructions indicate that the Pap shaft comprising of PapA is formed by the tight winding of a much thinner structure (Bullitt and Makowski, 1995). This can be demonstrated when Pap pili are subjected to mechanical or chemical stress that results in the formation of 2 nm narrow fibres which span the gap between two pilus fragments. These thread-like fibers have an open helical structure similar to the tip fibrillae and exhibit far greater flexibility than unstressed Pap pili shafts, and can extend the length of the pilus from 1 μm to up to 5 μm in length. This plasticity occurs when the bonds break between PapA on adjacent turns of the one-start helix, and a rotation occurs between adjacent domains along the one-start helix (Bullitt and Makowski, 1995). The functional significance of this plasticity is that it may facilitate the attachment of the bacteria under conditions that may otherwise remove them from the target host tissue (Bullitt and Makowski, 1995). Type I pili also exhibit a similar type of plasticity, which can be induced with glycerol as well as mechanical stress (Abraham et al., 1992; Ponniah et al., 1991).

The genes for papE, papF and papK encode proteins of sizes 16.5, 16 and 20 kDa respectively (Kuehn et al., 1992; Lindberg et al., 1986). PapE, PapF and PapK are all constituents of the heteropolymeric flexible tip fibrillae structure, in which PapG is distally located as a single subunit, with the entire structure present at the tips of Pap (Jacob-Dubuisson et al., 1993; Kuehn et al., 1992; Lund et al., 1988)(Figure 1.9).
The majority of the tip fibrillum is composed of repeating subunits of PapE, which is at least five-fold more abundant than the other tip fibrillae proteins (Jacob-Dubuisson et al., 1993; Kuehn et al., 1992). A papE mutant results in Pap which lack the tip fibrillae, although in some cases some fibrillar stubs were detected and were approximately 10 nm long, far less than the 42 nm length of the wild type tip fibrillae (Kuehn et al., 1992). Over-expression of PapE in trans in a papE mutant resulted in tip fibrillar structure approximately three times longer than the wild type.

The role of PapF and PapK was initially examined by creating mutant papF and papK genes and observing the affects of these mutations on tip fibrillae formation by high resolution electron microscopy (Kuehn et al., 1992). The papF and papK mutants did not effect formation of the tip fibrillae; however, the tip fibrillae were longer than that observed in the wild type, suggesting both proteins may play a role in length control (Kuehn et al., 1992). PapF effects pilus rod assembly since a papF mutant results in only 20% of the piliation seen in the wild type, therefore PapF may play a role in initiation of pilus shaft biosynthesis (Jacob-Dubuisson et al., 1993; Lindberg et al., 1987). More specifically though, PapF was shown to be an adaptor protein that is required to link the PapG adhesin to the distal end of the tip fibrillum (Jacob-Dubuisson et al., 1993).

The role of PapK is to initiate PapA polymerisation into the pilus rod and to regulate the length of the tip fibrillae (Jacob-Dubuisson et al., 1993). When a mutant was constructed which lacks the tip fibrillar (papE, papF, papK minus), only in trans complementation with PapK was able to induce a 20 fold increase in the accumulation of PapA, which suggests PapK can imperfectly substitute for whole tip fibrillae as a nucleator of PapA polymerisation (Jacob-Dubuisson et al., 1993). Furthermore, when PapK was overproduced in trans from a wild type pap system, the tip fibrillae became significantly shorter than the wild type, suggesting PapK terminates growth of the tip fibrillum (Jacob-Dubuisson et al., 1993). In a papF papK double mutant, no tip fibrillae were present on the very few pili observed,
suggesting initiation of the tip fibrillum requires PapF or PapK (Jacob-Dubuisson et al., 1993).

The \textit{papH} gene encodes a 21.8 kDa protein which is absolutely required for cell receptor binding and formation of cell-associated Pap pili (Båga et al., 1987). Mutants in \textit{papH} were found to produce pili which were significantly longer than wild type Pap, and 50-70\% of total pilus protein was found free of the cells. A decrease in the ratio of PapA:PapH resulted in shortened pili and when this ratio was reversed, the pili were lengthened. This indicates that PapH is a minor subunit involved in cell anchoring and length modulation (Båga et al., 1987). Therefore, PapH is the final pilin subunit assembled to terminate growth of the pilus fibre (Figure 1.9).

All the pilins of the Pap system (PapA, PapE, PapF, PapG, PapH and PapK) share both N- and C-terminal homology and significant identity when the protein sequences are aligned (Kuehn et al., 1993; Soto et al., 1998). This homology serves as recognition sites for the PapD chaperone, which delivers the pilins to the outer membrane assembly protein PapC for subsequent assembly. The PapD chaperone recognises and binds to the highly conserved C-terminal motif present in all the pilin subunits (Kuehn et al., 1993), since deletions in the C-terminus results in the abolition of chaperone-subunit interactions and PapD is able to bind synthetic peptides whose sequence corresponds to the C-terminal region of the pilin subunits (Hultgren et al., 1989; Kuehn et al., 1993). The motif is characterised by a series of alternating hydrophobic residues flanked by a glycine located 14 residues upstream from the C-terminus and a penultimate tyrosine.

Site-directed mutagenesis of the hydrophobic residues constituting the C-terminal motif of the Pap pilins was shown to specifically reduce the stability of PapG-PapK interactions within the pilus tip fibrillae, while mutations at positions flanking the C-terminal motif limit or abolish PapA-PapA interactions in the pilus shaft, but not PapA-PapK interactions (Soto et al., 1998). The C-terminal region therefore consists of the tail assembly surface, since in these variants PapA still recognised a complementary surface on the
preceding subunit in the pilus (PapK), but the ability to interact with the next incoming PapA subunit was abolished (Soto et al., 1998). Likewise, the C-terminal mutations in the PapG protein, which would be expected to possess only a tail assembly motif given that it is distally located in the pilus tip fibrillum, abolished interactions with PapK. The tip location for PapG is further supported by the fact that the N-terminal region of PapG shows no homology to the other pilin subunits, and therefore no head assembly is likely to occur (Soto et al., 1998). Therefore, the subunit C-terminal region serves a dual function for both subunit-subunit interaction and chaperone recognition. Site-directed mutagenesis also showed that the N-terminal motif of PapA is crucial for stable PapA-PapA interactions in the pilus shaft, as these subunits are assembled in a head to tail configuration (Soto et al., 1998).

An examination of periplasmic pilin intermediates complexed with PapD has been examined (Striker et al., 1994). PapA was able to form two periplasmic complexes with PapD in the form of a monomer and homodimer. The ability to form multimers was restricted to subunits that are homopolymers in the pilus, which includes PapA, but also PapE. Subunits that are present in a single or low copy in the pilus (PapK and PapG) did not form periplasmic inter-subunit interactions (Striker et al., 1994).

### 1.8.5 PapD

A great deal is known about the structure and function of the PapD chaperone (Hultgren et al., 1993; Hultgren et al., 1991; Jones et al., 1992; Soto and Hultgren, 1999; Thanassi et al., 1998a). The *papD* gene encodes a 28.5 kDa periplasmic transport protein, which is specifically involved in stabilising the major subunit PapA, along with PapE, PapF, PapK, PapH and to some extent PapG, and delivering these pilin subunits to the outer membrane assembly protein PapC (Lindberg et al., 1989; Striker et al., 1994)(Figure 1.9). In the absence of PapD, or if PapD is unable to bind the pilin subunits, the subunits misfold, aggregate and are targeted for degradation by the DegP protease (Jones et al., 1997).
The crystal structure of PapD has been determined (Holmgren and Branden, 1989). PapD consists of two globular domains orientated in the shape of a boomerang. Each domain is a β-barrel structure formed by two anti-parallel β-pleated sheets, that possess a topology similar to an immunoglobulin fold (Holmgren and Branden, 1989). It has been proposed that the two linked immunoglobulin-like domains of PapD are orientated to form a binding cleft for association with the Pap pilins (Hultgren et al., 1993). PapD is the prototypic member of a family of 26 highly conserved immunoglobulin-like pilus chaperones which includes among others, F17D from F17 pili, FaeE from K88, FanE from K99, SfaE from S pili and FimC from type I pili (Hultgren et al., 1993; Hung et al., 1996). The immunoglobulin-like chaperones have more recently been subdivided into two distinct sub-families, based upon conserved structural differences that occur within a highly conserved sheet in domain one and a conserved inter-domain hydrogen-bonding network (Hung et al., 1996).

The crystal structure of PapD complexed to a carboxy-terminal peptide corresponding to the pilus adhesin PapG showed that this peptide bound in an extended conformation with the terminus anchored in the inter-domain cleft of the chaperone via hydrogen bonds to invariant chaperone residues Arg-8 and Lys-112 (Kuehn et al., 1993). The complex was stabilised by main chain hydrogen bonds and contacts between hydrophobic residues in the peptide (Kuehn et al., 1993). The positioning of the peptide along the PapD G1 β-strand forms a β-sheet structure known as a ‘beta zipper’ (Kuehn et al., 1993). This binding has been implicated in determining binding specificity of the chaperone with the subunits. Site-directed mutagenesis of Arg-8 and Lys-112 abolished subunit recognition by PepD, showing the invariant cleft forms part of the subunit-binding pocket of PapD (Slonim et al., 1992). Another feature of the PapD structure, which is invariant in the immunoglobulin-like chaperone family is the presence of an internal salt bridge formed by Asp-196 and Arg-116 in association with the highly conserved Glu-83 (Holmgren and Branden, 1989; Holmgren et al., 1992). Site-directed mutagenesis of these residues showed that the internal salt bridge is required for chaperone stability in vivo, and may play a crucial role in a process of chaperone
dissociation before the assembly of the subunit into the pilus (Hung et al., 1999a). Residues along the G1 β-strand are required for efficient binding of the pilin subunits, which is consistent with the crystal structure of PapD-peptide complexes (Hung et al., 1996).

PapD chaperone function in pilus biogenesis depends on oxidant and chaperone-like activities of the periplasmic disulphide isomerase DsbA (Jacob-Dubuisson et al., 1994a). It was found that Pap were not assembled in a host strain that lacks DsbA, and this effect was due to the lack of DsbA dependent catalysing of disulphide bond formation in PapD. The formation of this bond was critical for correct folding of PapD in vivo; however, in vitro the absence of the disulphide bond did not prevent PapD from folding or forming a complex with PapG. DsbA may maintain nascently translocated PapD in a folding-competent conformation prior to disulphide bond formation, thereby acting as an oxidant and in a chaperone-like fashion (Jacob-Dubuisson et al., 1994a).

PapD functions by capping and uncapping interactive surfaces of nascently translocated pilus subunits (Kuehn et al., 1991). In vitro studies indicated that PapD binds PapG to form a stable complex and unlike cytoplasmic chaperones, the PapD chaperone was able to maintain PapG in a native-like conformation. The binding of PapD to PapG in vitro prevented aggregation of PapG since the dissociation of the PapD-PapG complex in vitro under reducing conditions uncapped interactive surfaces on PapG and caused the formation of large aggregates, that could be reversed by dilution of the denaturant and the addition of excess native PapD (Kuehn et al., 1991). PapD is also able to interact with itself, by capping its own subunit binding surface (Hung et al., 1999b). Crystal structure analysis of PapD dimers suggests that PapD interacts with itself by using the same interactive surfaces that it uses to bind subunits, although the interaction is a weak one (Hung et al., 1999b). The formation of pilus rods and tip fibrillae is thought to occur by PapD uncapping of the beta zipper arrangement that occurs between the PapD Gl β-strand and the C-terminus of the Pap pilins (Bullitt et al., 1996). Uncapping of the chaperone-protected PapA C-terminus led to the formation of pilus-rod sub-assemblies, while uncapping of PapE leads to the formation of tip
fibrillae sub-assemblies. Site directed mutagenesis on the leading edge of the beta zipper in PapA results in an altered helical symmetry, suggesting this surface is required for correct helical assembly of PapA (Bullitt et al., 1996).

More recently, the crystal structure of a PapD-PapK complex has been determined, which indicates that pilin subunits have the same immunoglobulin-like topology as the subunit binding N-terminal domain of the chaperone (Sauer et al., 1999). PapD functions by donating the G1 β-strand to complete the immunoglobulin-like fold of the PapK subunit in a process known as donor strand complementation. Furthermore, the structure also indicated that every subunit completes the immunoglobulin-like fold of its immediate neighbouring subunit via a mechanism termed donor strand exchange (Sauer et al., 1999).

PapD is known to function in the initial stages of biogenesis by facilitating the importation of subunits into the periplasm (Jones et al., 1997)(Figure 1.9). PapD undergoes a beta zippering interaction with the hydrophobic C-terminus of the pilin subunits, which allows their folding and release from the cytoplasmic membrane and into the periplasm (Jones et al., 1997). When PapD is absent, the subunits remain in the cytoplasmic membrane and are driven off-pathway via non-productive interactions, which affects cell growth (Jones et al., 1997). Subunit misfolding and aggregation the periplasm in the absence of PapD was sensed by two signal transduction systems: the Cpx two-component regulatory system and the σE modulatory pathway. The net result is the activation of DegP transcription, along with other chaperone-like proteins (Jones et al., 1997).

1.8.6 PapC

PapC is an 88 kDa outer membrane protein, which is essential in Pap pili expression on the cell surface since a papC mutant was devoid of Pap pili (Norgren et al., 1987). PapC is predicted to have a largely β-sheet secondary structure, typical of bacterial outer membrane proteins, and is likely to present large regions to the periplasm for specific interaction with chaperone-subunit complexes (Schifferli and Alrutz, 1994; Valent et al., 1995). Sequence
analysis showed that a particular region between amino acids 408 and 515 was very hydrophilic, with a predicted coil structure which indicates that this region of the protein may be directed towards an aqueous space, in the form of a channel. A mutation in the papC gene does not effect stability of PapA in the periplasm (Norgren et al., 1987). The term “molecular usher” has been used to describe the action of PapC, since PapC acts like a “human usher” at a cinema by specifically directing entry and assembly of pilin subunits at the proper time and place (Dodson et al., 1993)(Figure 1.9). It was found that chaperone-subunit complexes bound differentially to PapC by using an in vitro ELISA assay with purified PapC as the coating agent (Dodson et al., 1993). PapD-PapG complexes had the strongest binding to PapC, while both PapD-PapK and PapD-PapA did not bind (Dodson et al., 1993). Periplasmic PapD-PapF complexes bound PapC at an intermediate level between that of PapD-PapK and PapD-PapG while PapD-PapE complexes bound better than PapD-PapK and PapD-PapA. This result indicates that those components which are crucial for pilus tips (PapG, PapE, PapF and PapK) are made before the pilus rods, ensuring every pilus rod is joined end-to-end to an adhesive tip fibrillum. In other words, PapD-PapA complexes are not targeted to empty PapC sites, only those containing pre-formed tip fibrillae (Dodson et al., 1993).

A study of association and disassociation events of chaperone-subunit complexes with the PapC usher protein in real time using a surface plasmon resonance technique showed that PapD-PapG complexes had a 16 fold greater affinity and a 29 fold faster disassociation with the usher if compared to PapD-PapA complexes (Saulino et al., 1998). This indicates that kinetic partitioning of chaperone-adhesin complexes to the usher is a defining factor in tip localisation of the adhesin protein within the pilus structure (Saulino et al., 1998).

The structure of the PapC usher protein has been determined and was found to form a pore when reconstituted in liposomes, with a diameter of approximately 2 nm (Thanassi et al., 1998b). Purified PapC was found to oligomerise into a ring shaped hexameric complex, with a central 2 nm pore and an outer diameter of 15 nm, which confirms the pore forming ability
of PapC from reconstitution in liposomes (Thanassi et al., 1998b). With a 2 nm pore size, the 2 nm tip fibrillum would be able to pass through the PapC channel, but the 6.8 nm wide helical pilus rod would not be able to fit through. Experiments have shown that Pap pili can be unravelled into linear fibres of approximately 2 nm in the presence of glycerol or when subjected to mechanical stress thereby consisting of PapA subunits held together by head to tail interactions (Bullitt and Makowski, 1995; Thanassi et al., 1998b). The unravelled pili are then able to fit through the PapC channel and package into helical cylinders once outside the cell (Thanassi et al., 1998b).

Evidence also suggests that secretion of pili across the outer membrane may be independent of cellular energy and thermodynamically driven (Jacob-Dubuisson et al., 1994b). It has been proposed that the winding of the PapA fibre into a helix on the external surface may provide the driving force for the translocation of pili across the outer membrane (Thanassi et al., 1998a). This may only account for some of the processes occurring, since pilus tips can be assembled on the cell surface without the presence of pilus rods, so secretion energy may also come from favourable protein-protein interactions involved in assembly of pilin subunits into fibres (Saulino et al., 1999).

1.9 Vaccines Against Human ETEC Infection

ETEC is a major cause of acute diarrhoea in children in the developing world and to travellers from industrialised countries in these areas (Merson et al., 1980; Sack, 1975). Although the disease is usually self-limiting in adults and oral rehydration is an effective treatment, prevention of the illness is preferred. Existing medical treatments for protecting international travellers and children in developing countries from ETEC infection are bismuth subsalicylate or the prophylactic administration of antibiotics including trimethoprim-sulfamethoxazole or ciprofloxacin (Ericsson and DuPont, 1993). Both of these treatments have considerable disadvantages including (1) the emergence of antibiotic resistant strains, which was revealed in the Gulf War in which a large percentage of isolated ETEC strains
were resistant to trimethoprim-sulfamethoxazole, tetracycline or ampicillin, and (2) the toxic effects of long term intake of bismuth subsalicylate (Ericsson and DuPont, 1993; Hyams et al., 1991).

Vaccine prevention of ETEC disease is a more suitable alternative to antibiotic or bismuth subsalicylate therapies. Numerous ETEC vaccines have been constructed based on the two recognised virulence determinants of ETEC, namely the ST and LT enterotoxins and the 21 distinct colonisation factors. The difficulty in vaccine preparation has been to achieve broad range protection against ETEC infection in diverse geographical areas. It has been proposed that a multivalent vaccine consisting of CFA/I and CS1-6 along with an LT toxoid may protect against 80-90% of strains worldwide, and if less frequent pili antigens are also included (CS12, CS14 or CS17 for example) then greater than 90% protection could be achieved (Levine et al., 1994). No ETEC vaccine is currently in use for humans.

1.9.1 Inactivated Whole Cell and Toxoid Vaccines

One vaccine approach for protection against ETEC is to use killed bacteria alone or in combination with the B-subunit of the LT enterotoxin (LT-B) or the closely related cholera toxin B-subunit (CT-B). A monovalent vaccine consisting of colicin-killed ETEC strain H10407 (O78:H11 CFA/I*, LT+, ST+) was well-tolerated and able to elicit significant rises in intestinal sIgA antibody to both CFA/I and LT, with 75% protective efficacy upon challenge with viable ETEC of the homologous strain H10407 (O78:H11 CFA/I*, LT+, ST+) or a heterologous serotype (O63:H- LT+, ST+) expressing CFA/I (Evans et al., 1988a; Evans et al., 1988b).

A prototype ETEC vaccine containing a mixture of CT-B and formalin-killed ETEC bacteria of three different strains (O78:H12 LT*, ST*, CFA/I*; O139:H28 LT*, ST*, CS1* CS3* and O6:H16 LT, ST*, CS2*, CS3*) was both safe and induced a strong mucosal immune response against both CFA/I and CFA/II (CS1, CS2, CS3) as well as to CT-B in almost all of the vaccinees (Åhrén et al., 1993). CFA/IV expressing ETEC producing CS4 and/or CS5 pili
has also been added to the vaccine resulting in significant IgA responses (9-36 fold) in 91% of volunteers to CFA/IV, but also 84% against CFA/I, and 87% against CFA/II sub-components CS1, CS2 and CS3 (Jertborn et al., 1998).

A study of an oral inactivated ETEC plus CT-B in Bangladeshi adults and children showed that the antibody secreting cell (ASC) of the IgA isotype response against CFA/I, CFA/II and CFA/IV increased between 29- to 46-fold in adults and 13- to 24-fold in children (Qadri et al., 2000b). Furthermore, the ASC responses against CT-B increased 426 fold in adults and 46 fold in children. This showed that a single dose of ETEC vaccine was able to elicit significant mucosal immune responses in both adults and children (Qadri et al., 2000b).

1.9.2 Purified Pili Vaccines

Oral administration of purified CFA/I and CFA/II consisting of CS1 and CS3 pili in adult volunteers has been shown to be ineffective in not only manifesting rises in both serum and intestinal sIgA antibodies but also in protection against challenge with wild type ETEC strains expressing these colonisation factors (Evans et al., 1984; Levine et al., 1986). The administration of the vaccine with 2.0 g of NaHCO₃ to neutralise gastric acid, along with pre-treatment of volunteers with cimetidine to diminish gastric acid secretion resulted in only two of ten volunteers with detectable serum and intestinal sIgA antibodies (Levine et al., 1986). It was shown that gastric juice, which has been neutralised to pH 7.0, adversely affects the antigenicity of the vaccine (Schmidt et al., 1985)

To overcome the adverse effects of gastric juices on the antigenicity of purified pili vaccines, biodegradable polymers and microspheres have been used to deliver pili antigens to the small intestine (Edelman et al., 1993; Reid et al., 1993; Tacket et al., 1994). When CFA/I is incorporated into biodegradable polymer microspheres and immunised orally in rabbits, high titres of serum CFA/I antibodies are exhibited, which were not detected in rabbits immunised orally with purified CFA/I alone (Edelman et al., 1993). The microspheres have also been used to deliver an ETEC vaccine consisting of purified CFA/II composed of CS1
and CS3 pili (Reid et al., 1993). In human volunteer studies, it was shown that five out of ten vaccinees developed sIgA anti-CFA/II antibodies in jejunal fluid; however, upon challenge with an ETEC strain expressing LT, ST, CS1 and CS3, seven out of ten volunteers developed diarrhoea (Tacket et al., 1994).

1.9.3 Live Oral Vaccines

The potential use of attenuated ETEC strains for oral immunisation to protect against subsequent challenge from heterologous organisms producing homologous colonisation factors has been examined (Levine et al., 1986). The live ETEC strain used lacked the genes encoding for both LT and ST toxins, but retained the pili antigens CS1 and CS3. The protective efficacy of the strain in human volunteer studies was 75%. However, 10-15% of the volunteers receiving the oral live strain at higher dose levels developed diarrhoea, indicating adverse side effects in using this strain as a live oral vaccine (Levine et al., 1986).

The intimate association and prolonged residence in the gut-associated lymphoid tissue have led to the use of attenuated Salmonella vectors to deliver ETEC colonisation factors and toxins to the human immune system (Cardenas and Clements, 1993b; Clements and Cardenas, 1990; Giron et al., 1995; Haq et al., 1995; Maskell et al., 1987; Mason et al., 1998; Pascual et al., 1999; Tacket et al., 1997; Wu et al., 1995). Several Salmonella strains have been used as potential vaccine candidates which express ETEC colonisation factors on the cell surface including a Salmonella typhimurium aroA, asd mutant that expresses CFA/I pili (Wu et al., 1995) and a Salmonella typhi aroC, aroD mutant (CVD 908) expressing CFA/I and CS3 pili (Giron et al., 1995; Hone et al., 1991).

More recent oral vaccines against the LT toxin of ETEC have been directed towards creating “edible vaccines” in which a modified LT-B gene has been introduced into potato tubers producing protein at a level of 2 μg per gram of tuber tissue (Haq et al., 1995). Antibodies raised in response to the potato LT-B could inhibit LT activity on mammalian cells in vitro (Haq et al., 1995). Challenge experiments in mice fed on a regime of raw potato
tubers expressing LT-B constitutively, followed by challenge with 25 μg of LT indicated the mice were partially protected from fluid accumulation in the gut resulting from the effect of LT (Mason et al., 1998). The results of human trials utilising the potato tuber harbouring LT-B have yet to be published.

1.9.4 Passive Vaccines

Passive vaccines consisting of either serum or colostral bovine immunoglobulins from immunised adult cows have been used to prevent ETEC infections in adult volunteers (Freedman et al., 1998; Tacket et al., 1988). In one study, cows were immunised with a vaccine cocktail consisting of Freund’s adjuvant, LT toxin and CT toxin and suspensions of 2x 10⁹ heat or gluteraldehyde-inactivated ETEC representing several different serogroups. The colostral antibodies produced were able to specifically protect volunteers against challenge with ETEC strain H10407 (O78:H11) which produces CFA/I and ST and LT enterotoxins (Tacket et al., 1988). The product was extremely safe and well tolerated, with no side effects. In a more refined study, the protective efficacy of immunising cows with purified CFA/I pili, instead of whole cell samples was shown to provide 90% protection against clinical diarrhoea from ETEC strain H10407 (O78:H11, CFA/I⁺, LT⁺, ST⁺)(Freedman et al., 1998).

It is also clear that the titre of immunoglobulins against various CFAs is important in prophylaxis. A lack of prophylactic efficacy of an enteric-coated bovine hyperimmune milk product against ETEC challenge administered during a standard meal, has been observed using bovine immunoglobulins with activity against CFA/I, CS3, and CS6 upon challenge with ETEC strain E24377A expressing CS1 and CS3 (Tacket et al., 1999).
1.10 Aims of this Research

At the molecular level, very little is known about the biogenesis of CS5 pili from ETEC. Only very initial studies have been carried out, including the identification of the pilus structure by electron microscopy, cloning of the DNA region required for biosynthesis from wild type ETEC strain O115:H40, and sequence analysis of the gene encoding the major pilin subunit (Clark et al., 1992; Heuzenroeder et al., 1989). There is also some doubt as to the true morphology of the pilus itself (Heuzenroeder et al., 1989; Knutton et al., 1989; Thomas et al., 1985).

The overall aim of this research project is to characterise, at the molecular level, the CS5 pilus encoding region from Enterotoxigenic Escherichia coli O115:H40, culminating in a proposed model for biosynthesis of this pili. The specific aims include:

1. To sequence the entire region of DNA encompassing the CS5 pili-encoding region, along with flanking regions of DNA which may provide evidence for the origin of the region.
2. To compare the deduced sequence against other pili systems for possible homology, and to determine the size and nature of the protein products produced from the CS5 genes.
3. To characterise the function and role of individual genes towards biogenesis by creating deletion mutants in each of the genes and assessing the alternations (if any) in the specific phenotype observed.
4. To provide a model for CS5 pilus biogenesis based on the results obtained and to compare this model with existing pilus biogenesis models from other bacterial systems.

These data could then provide a useful basis for constructing suitable candidate vaccines directed against CS5-expressing ETEC.
Chapter 2

Materials and Methods

2.1 Chemicals and Reagents

Chemicals used were Analar grade. Sodium dodecyl sulphate (SDS), chloroform, ethanol, methanol, butan-2-ol, propan-2-ol (isopropanol), perchloric acid, glycerol, glucose, sucrose, ammonium sulphate, potassium acetate, glacial acetic acid, Coomassie Brilliant Blue G250 and Coomassie Brilliant Blue R250 were from BDH Chemicals. Tris (Trisma base), magnesium chloride, Tween-20, ethidium bromide, bovine serum albumin (fraction V), L(+)arabinose, imidazole and dithiothreitol (DTT) were obtained from Sigma. Ethylene-diaminetetra-acetic-acid, disodium salt (EDTA), caesium chloride, calcium chloride, sodium chloride, Triton X-100 and sodium hydroxide were from Ajax chemicals, NSW, Australia. Deoxyribonucleoside triphosphates (dATP, dCTP, dGTP and dTTP), 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal), isopropyl-β-D-thiogalacto-pyranoside (IPTG) and glycine were purchased from Boehringer Mannheim. Mineral oil was Primol 352 (Esso). Other chemicals were from Ajax, BDH or Sigma. Milli-Q (MQ) water was water purified using a Milli-Q water purification system (Millipore Corp.) with a measured resistance to conductivity of 18 MΩ/cm. L-[³⁵S]-methionine (1270 Ci/mmol) was purchased from Amersham.

Antibiotics (ampicillin, kanamycin sulphate, rifampicin) were purchased from Sigma. All other anti-microbial agents (dyes and detergents) were purchased from Glaxo, Calbiochem, Sigma Chemical Co., or BDH Chemicals Ltd.

The following electrophoresis grade reagents were obtained from the sources indicated: acrylamide and ammonium persulphate (Bio-Rad), ultra pure N,N'-methylene bis-
acrylamide and urea (BRL); high gelling temperature (HGT) agarose (Seakem); and N,N,N',N'-Tetramethyl-ethylenediamine (TEMED) was purchased from Sigma.

2.2 Enzymes

All restriction endonucleases were purchased from either Boehringer Mannheim, New England Biolabs or Progen and used according to the suppliers instructions. Lysozyme was from Sigma. Other DNA modifying enzymes were purchased from the following suppliers: New England Biolabs (T4 DNA ligase) and Boehringer Mannheim (DNA polymerase I, Klenow fragment of DNA polymerase I, molecular biology grade shrimp alkaline phosphatase and RNase A). Taq polymerase (Ampli Taq) was purchased from Perkin Elmer Cetus Corp.

Horseradish peroxidase-conjugated goat anti-rabbit IgG was obtained from Kirkegaard and Perry Laboratories Inc.

Sequencing kits, using dye-labelled terminators, and dye-labelled primers were purchased from Perkin Elmer Applied Biosystems (Foster City, California). The Erase-a-base kit used to produce nested deletions of plasmid DNA was purchased from Promega.

2.3 Bacterial Strains and Plasmids

The wild type ETEC strains and E. coli K-12 strains used for transformation of plasmids and expression of proteins are shown in Table 2.1. Plasmid vectors and plasmids which were constructed in this study are described in Table 2.2.

2.4 Growth Media

The routine bacterial cultivation of E. coli strains used the following media. General growth media (used unless otherwise stated) was nutrient broth (NB) which consisted of Lab Lemco (Oxoid) (10 g/l), Bactopeptone (Oxoid) (10 g/l) and NaCl (5 g/l). Luria-Bertani broth (LB) was used for rich growth media and consisted of Bacto Tryptone (Difco) (.0 g/l), Bacto
<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. coli K-12</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH5</td>
<td>$F^+$, *recA1, endA1, hsdR17 [r$^-$ m$^+$], supE44, $\lambda^-$, thi1, gyrA, relA1</td>
<td>Bethesda Research Laboratories</td>
</tr>
<tr>
<td></td>
<td>$F^+$, *endA1, hsdR17 [r$^-$ m$^+$], supE44, thi-1,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DH1</td>
<td>$F^+$, gyrA96, *recA1, relA1, endA1, thi-1, hsdR17, supE44, $\lambda^-$</td>
<td>B. Bachmann</td>
</tr>
<tr>
<td>TOPP 10 F'</td>
<td>$F^+$[lac$^\text{R}$ Tet$^\text{R}$], *mcrA, $\Delta$(mrr-hsd RMS-mcrBC), φ80 lacZ $\Delta$M15 $\Delta$lacX74 deoR, *recA1, araD139 $\Delta$(ara-leu), 7697 galU, galK, rpsL, endA1, nupG</td>
<td>Stratagene</td>
</tr>
<tr>
<td>BL21 (DE3)</td>
<td>$F^+$ *ompT gal [DCM] [lon] hsdS$^B$ (r$^B$ m$^B$)</td>
<td>Novagen</td>
</tr>
<tr>
<td></td>
<td>DE3 ($\lambda$ prophage carrying T7 RNA polymerase gene)</td>
<td></td>
</tr>
<tr>
<td>E2096</td>
<td>DH5 + pGP1-2</td>
<td>Laboratory Collection</td>
</tr>
<tr>
<td><strong>Wild Type ETEC</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE423</td>
<td>(isolate 576/5) O115:H40 CS5$^+$ CS6$^+$</td>
<td>Manning <em>et al.</em>, 1987</td>
</tr>
<tr>
<td>PE213</td>
<td>E6237B O78:H11 CFA/I$^+$ ST$^+$ LT$^+$</td>
<td>Willshaw <em>et al.</em>, 1982</td>
</tr>
<tr>
<td>C9216-2</td>
<td>O6:H16 rns$^-$ coo$^+$ CS1$^+$</td>
<td>J.R. Scott</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td>Source/Reference</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>pGEM-T</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PCR cloning vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pGP1-2</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; T7 RNA polymerase</td>
<td>Tabor and Richardson, 1985</td>
</tr>
<tr>
<td>pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; cloning vector</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pWKS130</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; low copy number cloning vector</td>
<td>Wang and Kushner, 1991</td>
</tr>
<tr>
<td>pGEM-7Zf&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; cloning vector</td>
<td>Promega</td>
</tr>
<tr>
<td>pREP4</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; lacI&lt;sup&gt;+&lt;/sup&gt; constitutive</td>
<td>Qiagen</td>
</tr>
<tr>
<td>pQE-31</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; His-tag vector</td>
<td>Qiagen</td>
</tr>
<tr>
<td>pBAD18-Kan</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; expression vector</td>
<td>Guzman et al., 1995</td>
</tr>
<tr>
<td>pPM1312</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; MluI-ClaI cutdown of pPM1306</td>
<td>Heuzenroeder et al., 1989</td>
</tr>
<tr>
<td>pPM5302</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PstI fragment from pPM1306 in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5303</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PstI fragment from pPM1306 in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM1306</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; cosmid clone CS5 pili region</td>
<td>Heuzenroeder et al., 1989</td>
</tr>
<tr>
<td>pPM5610</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5393</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5612</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5391</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5384</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5388</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5383</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5380</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5379</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5619</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5613</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5372</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5303</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5364</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5302</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5624</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; nested deletion product from pPM5302</td>
<td>This study (section 3.2.1.1)</td>
</tr>
<tr>
<td>pPM5631</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; SphI-Smal fragment of pPM1312 in pGEM-7Zf&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.5.1)</td>
</tr>
<tr>
<td>pPM5605</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; EcoRI-XbaI fragment of csfB in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.6.1)</td>
</tr>
<tr>
<td>pPM5636</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfC in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.6.1)</td>
</tr>
<tr>
<td>pPM5640</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfD in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.6.1)</td>
</tr>
<tr>
<td>pPM5633</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfE in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This study (section 3.2.6.1)</td>
</tr>
<tr>
<td>Plasmid</td>
<td>Feature</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>pPM5390</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; EcoRI-XbaI fragment of csfF</td>
<td>in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5647</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfD from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5644</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfD</td>
<td>in pWSK130</td>
</tr>
<tr>
<td>pPM5646</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfA from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5648</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ApaI-SacI fragment from pPM5646</td>
<td>in pBS-SK&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfE from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5657</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfE</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5684</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; KpnI-XbaI fragment of csfE</td>
<td>in pBAD18-Kan</td>
</tr>
<tr>
<td>pPM5651</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfC from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5638</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of csfC</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5661</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; PstI-XbaI fragment of cfaC</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5681</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; PstI-BamHI fragment of cooC</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5655</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfF from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5659</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; EcoRI-XbaI fragment of csfF</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5672</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; EcoRI-XbaI fragment of csfF</td>
<td>from pPM5659 in pWKS130</td>
</tr>
<tr>
<td>pPM5662</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfB from pPM5631</td>
<td></td>
</tr>
<tr>
<td>pPM5665</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt; EcoRI-XbaI fragment of csfB</td>
<td>in pWKS130</td>
</tr>
<tr>
<td>pPM5679</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; BamHI-KpnI fragment of csfB</td>
<td>in pQE-31</td>
</tr>
<tr>
<td>pPM5690</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt; ΔcsfB ΔcsfF from pPM5655</td>
<td></td>
</tr>
</tbody>
</table>
Yeast Extract (Difco) (5 g/l) and NaCl (5 g/l) as described (Miller, 1972). M9 minimal medium contained M9 salts supplemented with 0.2 mg/ml MgSO₄ and 0.5% (w/v) glucose. Liquid cultures were grown in 20 ml McCartney bottles at 37°C with moderate aeration unless otherwise stated.

Solidified media used included: nutrient agar (NA) consisting of NB with the addition of agar (BBL Technical Grade) (15 g/l); Luria-Bertani agar (LA), which was LB with the addition of Bacto Agar (Difco) (15 g/l). All specific pili expression experiments for both recombinant and wild type E. coli strains used CFA agar (Evans et al., 1979), which consisted of Difco casamino acids (10 g/l), Bacto yeast extract (1.5 g/l), MgSO₄.3H₂O (0.05 g/l) and MnCl₂.4H₂O (0.005 g/l). Ampicillin (Amp) was used at 100 µg/ml in media, kanamycin (Kan) was used at 50 µg/ml and rifampicin (Rif) at 200 µg/ml.

Colour indicator plates used were LA or CFA with the addition of 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) dissolved (20 mg/ml) in dimethyl formamide at a concentration of 40 µg/ml and isopropyl-β-D-thiogalactopyranoside (IPTG) dissolved in Milli-Q water (24 mg/ml) and filter sterilised to give a final plate concentration of 48 µg/ml. All incubations were at 37°C unless otherwise specified.

2.5 Maintenance of Bacterial Strains

For long term storage, all strains were maintained as lyophilised cultures, stored in vacuo in sealed glass ampoules. When required, an ampoule was opened and its contents suspended in several drops of the appropriate sterile broth. Half the contents were then transferred to a 10 ml bottle of nutrient broth (NB) and incubated with aeration overnight at the appropriate temperature. The other half was streaked onto two NA plates and incubated overnight at the appropriate growth temperature. Antibiotics were added to the media when appropriate. If the colony form was uniform, single colonies were selected and picked off plates for subsequent storage or use.
Bacterial strains were prepared for long-term storage by suspension of several loopfuls in a small volume of sterile skimmed milk. Approximately 0.2 ml aliquots of this thick bacterial suspension were dispensed into sterile 0.25 x 4 inch freeze drying ampoules and the end of each ampoule was plugged with cotton wool. The samples were then lyophilised in a freeze drier. After the vacuum was released, the cotton wool plugs were pushed well down the ampoule and a constriction was made just above the level of the plug. The ampoules were evacuated to a partial pressure of 30 microns and then sealed at the constriction without releasing the vacuum. Finally the ampoules were labelled and stored at 4°C.

Short-term storage of strains in routine use was as a suspension of freshly grown bacteria in glycerol (32% (v/v)) and peptone (0.6% (w/v)) at -70°C. Fresh cultures from glycerols were prepared by streaking a loopful of the glycerol suspension onto a NA or LA plate (with or without antibiotic as appropriate) followed by overnight incubation prior to use.

2.6 Synthesis of Oligonucleotides

Oligonucleotides were synthesised on an Applied Biosystems 381A DNA synthesiser in the trityl-off mode and butan-2-ol extracted prior to use. Reagents were purchased from Applied Biosystems or Ajax Chemicals. Alternatively, oligonucleotides were purchased from Geneworks (Adelaide, Australia) in a lyophilised form and resuspended to an optimal optical density (OD) in sterile MQ water. The oligonucleotides used in this study are listed in Table 2.3.

2.7 Transformation of E. coli K-12

2.7.1 Preparation of Super Competent Cells

A 10 ml overnight culture in LB was diluted 1:20 into LB with the appropriate antibiotics and incubated with aeration until an OD₆₀₀nm of 0.5 was reached. The cells were chilled on ice for 5 min, then pelleted at 4°C in a bench centrifuge (Minor MSE bench
<table>
<thead>
<tr>
<th>Oligonucleotide #</th>
<th>Sequence 5'-3'</th>
<th>Location/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M13 reverse</td>
<td>GGAAACAGCTATGACCAGT</td>
<td>pBS (nt 824-809)</td>
</tr>
<tr>
<td>-20 M13 universal</td>
<td>GTAAACGACGCCAGT</td>
<td>pBS (nt 603-617)</td>
</tr>
<tr>
<td>2510</td>
<td>CAATTGGATTTTACTATC</td>
<td>csfB (nt 2708-2687)</td>
</tr>
<tr>
<td>2524</td>
<td>GATATGAAACTATCAATTTG</td>
<td>csfC (nt 2687-2708)</td>
</tr>
<tr>
<td>2526</td>
<td>GCGATGAGACAATACATTTCC</td>
<td>csfC (nt 3578-3768)</td>
</tr>
<tr>
<td>2527</td>
<td>CGTATCTGTTGTCATCAT</td>
<td>IS66 (nt 7990-7977)</td>
</tr>
<tr>
<td>2528</td>
<td>CCTTCGCCATGCGACGC</td>
<td>IS30 (nt 9360-9377)</td>
</tr>
<tr>
<td>2529</td>
<td>GGGAACATTACACAGCAG</td>
<td>csfE (nt 5360-5380)</td>
</tr>
<tr>
<td>2532</td>
<td>CAGTCTCTGATCTTCTTACAA</td>
<td>csfB promoter (nt 1418-1399)</td>
</tr>
<tr>
<td>2540</td>
<td>CAAATTCTGAAGTAATCTTG</td>
<td>csfD (nt 6604-6593)</td>
</tr>
<tr>
<td>2541</td>
<td>CGGGCAGAAATCTCCTTTA</td>
<td>csfF (nt 6272-6289)</td>
</tr>
<tr>
<td>2542</td>
<td>CTCCTACATGCTTCCTACAG</td>
<td>csfE (nt 5803-5785)</td>
</tr>
<tr>
<td>2551</td>
<td>CGTCCGACGCGTCAG</td>
<td>csfB promoter (nt 861-876)</td>
</tr>
<tr>
<td>2552</td>
<td>GTAACGCTTCTGTAATCTAAT</td>
<td>csfB (nt 2231-2250)</td>
</tr>
<tr>
<td>2553</td>
<td>TGGATGTTTGGCGCAGATTATT</td>
<td>csfB (nt 2592-2612)</td>
</tr>
<tr>
<td>2554</td>
<td>GGTATTCGAGAAGAATAAAT</td>
<td>csfC (nt 3061-3077)</td>
</tr>
<tr>
<td>2555</td>
<td>GGCGCAGTGGCGCGATTATA</td>
<td>csfC (nt 3372-3391)</td>
</tr>
<tr>
<td>2556</td>
<td>AATATCTGTTGTCAGCC</td>
<td>csfC (nt 3658-3677)</td>
</tr>
<tr>
<td>2557</td>
<td>CAGTACTGTTTATATGTTCC</td>
<td>csfC (nt 3984-4003)</td>
</tr>
<tr>
<td>2558</td>
<td>GGGTATGAGGTCAGAGTT</td>
<td>csfC (nt 4272-4290)</td>
</tr>
<tr>
<td>2559</td>
<td>TGGTCACTACTGATTCAG</td>
<td>csfC (nt 4534-4562)</td>
</tr>
<tr>
<td>2561</td>
<td>GGAGGATCAGCAGCTGC</td>
<td>csfC (nt 5116-5132)</td>
</tr>
<tr>
<td>2563</td>
<td>GGACTCCGTGAGAATTCCT</td>
<td>csfF (nt 6032-6049)</td>
</tr>
<tr>
<td>2564</td>
<td>GGCCACACTTTTACGTACA</td>
<td>csfF (nt 6811-6830)</td>
</tr>
<tr>
<td>2565</td>
<td>CTTGTTGTAACCCGCACTTAT</td>
<td>csfF (nt 7181-7202)</td>
</tr>
<tr>
<td>2566</td>
<td>GTGGTGTTCATCCAACCC</td>
<td>csfF (nt 7408-7425)</td>
</tr>
<tr>
<td>2567</td>
<td>GAGCGCGGTATTAACTAGGTT</td>
<td>IS66 (nt 7700-7718)</td>
</tr>
<tr>
<td>2569</td>
<td>GCCCTATACAACATGTAT</td>
<td>IS66 (nt 8206-8225)</td>
</tr>
<tr>
<td>2570</td>
<td>GCGGAGGTTTTTGAATGAGA</td>
<td>IS30 (nt 8760-8777)</td>
</tr>
<tr>
<td>2571</td>
<td>GGCGCAGATGTATTC</td>
<td>IS30 (nt 9075-9091)</td>
</tr>
<tr>
<td>2578</td>
<td>CCAGGTCGAAGGCGTACA</td>
<td>ISJ (nt 546-519)</td>
</tr>
<tr>
<td>2806</td>
<td>GGAGTTGCGACAGATCCT</td>
<td>ISJ30 (nt 9666-9687)</td>
</tr>
<tr>
<td>2807</td>
<td>GGAGGACGACGCTGATAG</td>
<td>ISJ (nt 85-87)</td>
</tr>
<tr>
<td>2873</td>
<td>GGCCATTTGTAACAGCAG</td>
<td>csfA (nt 1508-1491)</td>
</tr>
<tr>
<td>2944</td>
<td>GGAAATCCCAAGGCGACTGCTGC</td>
<td>csfB T7 (nt 1968-1980)</td>
</tr>
<tr>
<td>2945</td>
<td>GGTCTAGAAGCCAGCTCACTTTTATCAGC</td>
<td>csfB T7 (nt 2856-2827)</td>
</tr>
<tr>
<td>2969</td>
<td>GGCGGCTATCTGATACGACTTCGCCGAAGA</td>
<td>ISJ (nt 437-454)</td>
</tr>
<tr>
<td>2971</td>
<td>GAATTTACATGTTAGGGGATGATG</td>
<td>csfF T7 (nt 5789-5810)</td>
</tr>
<tr>
<td>2972</td>
<td>GGTCTAGAATGACCATAATGAAAGGAAAAAG</td>
<td>csfF T7 (nt 6576-6555)</td>
</tr>
<tr>
<td>2973</td>
<td>GCTCTGGAGAAATATGGAACATATCC</td>
<td>csfET7 (nt 5177-5196)</td>
</tr>
<tr>
<td>2974</td>
<td>GGTCTAGATGGAATATTATAGCGCTTTAATAA</td>
<td>csfET T7 (nt 5879-5869)</td>
</tr>
<tr>
<td>2977</td>
<td>GGTCTAGACGGCAGCGCAGAAACCAACT</td>
<td>csfCT T7 (nt 2731-2749)</td>
</tr>
<tr>
<td>2978</td>
<td>GGTCTAGACCACTATAGAACGACAGTTTGG</td>
<td>csfCT T7 (nt 5263-5244)</td>
</tr>
<tr>
<td>2979</td>
<td>GGTCTAGACGAGATGTCACGAGCCTT</td>
<td>csfDT7 (nt 6455-6472)</td>
</tr>
<tr>
<td>2980</td>
<td>GGTCTAGAAGCACTACGAGGAAATTTG</td>
<td>csfDT7 (nt 7676-7658)</td>
</tr>
<tr>
<td>3062</td>
<td>AAGTGCGGAGCGGTTGTGATGCTG</td>
<td>ccoC (nt 1341-1357)</td>
</tr>
<tr>
<td>3063</td>
<td>GGGAATCCGATAGAAGAAAATATAATATC</td>
<td>ccoC (nt 4028-4007)</td>
</tr>
<tr>
<td>3064</td>
<td>AATCGATCGCTCCTTTTATCAGGAA</td>
<td>ccoC (nt 2141-2160)</td>
</tr>
<tr>
<td>3065</td>
<td>GGTCCTAGAAGCCGATACGCGAATGTA</td>
<td>ccoC (nt 4750-4733)</td>
</tr>
<tr>
<td>3110</td>
<td>GGCAGATCCGTCATGTTGATTCAGTGTAATG</td>
<td>csfB-His (nt 2144-2164)</td>
</tr>
<tr>
<td>3111</td>
<td>GGGTCGAGGGTTCTTATGAGTCATAT</td>
<td>csfB-His (nt 2784-2765)</td>
</tr>
<tr>
<td>3196</td>
<td>CATACTCTATCGAATGCG</td>
<td>IS911 (nt 8693-8675)</td>
</tr>
<tr>
<td>3197</td>
<td>CGCCCTAACCCAATCTGCTAT</td>
<td>IS30 (nt 9240-9221)</td>
</tr>
</tbody>
</table>

nt: denotes nucleotide position corresponding to the designated sequence derived for the CS5 pili encoding region (Accession. no. AJ224079). The ccoC sequence is from Accession. no. 76908 (Froehlich et al., 1994) and cfaC is from Accession. no. M55661 (Jordi et al., 1992b). pBS denotes pBS-SK+ cloning vector (Stratagene). Restriction enzyme sites are underlined.
centrifuge, 10 min, 4500 rpm) and resuspended in 10 ml of ice cold solution α (30 mM KAc, 100 mM KCl, 10 mM CaCl₂, 50 mM MnCl₂, 15% glycerol). The solution was then pelleted for 10 min at 4°C in a bench centrifuge (Minor MSE bench centrifuge, 4500 rpm) and resuspended in 1 ml of ice-cold solution β (10 mM MOPS, 75 mM CaCl₂, 10 mM KCl, 15% glycerol) and left at 4°C for 2 hr. The final solution was aliquoted (200 μl) into 1.5 ml reaction tubes and either stored at -70°C or used immediately.

2.7.2 Transformation Procedure

Transformation was performed essentially according to the method described (Brown et al., 1979). Competent cells were left on ice for at least 2 hr before transformation (frozen cells were thawed on ice). Competent cells (200 μl) were then mixed with DNA and left on ice for at least 1 hr. The cell/DNA mixture was heated at 42°C for 1.5 min, cooled on ice for another 10 min, then 1 ml of LB was added and the mixture incubated with aeration for 1 hr at 37°C. Aliquots of the cultures were then plated directly onto selection plates. Addition of sterile buffer to competent cells was included as a negative control. Transformation controls consisted of competent cells plus pure cloning vector.

2.8 Plasmid Preparation Procedures

2.8.1 Small Scale Preparation

Small scale plasmid purification was performed by a modification of the three step alkali lysis method (Garger et al., 1983). A total of 1.5 ml of an overnight bacterial culture was transferred to a 1.5 ml reaction tube and harvested by centrifugation (3 min, 8000 rpm, Heraeus Biofuge 15). The cells were resuspended in 100 μl of Solution I (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA) and left at room temperature for 5 min. Cell lysis occurred after the addition of 200 μl of Solution II (0.2 M NaOH, 1% (w/v) SDS) and a 5 min incubation on ice. To precipitate protein, chromosomal DNA and high molecular weight
RNA, 150 μl of Solution III (3M KAc, pH 4.8, 11.5% (v/v) glacial acetic acid) was added, and after a 5 min incubation on ice, the mixture was centrifuged (Eppendorf 5417R centrifuge, 14000 rpm, 5 min, 4°C) and the supernatant transferred to a fresh 1.5 ml reaction tube. After addition of 250 μl of 7.5 M NH₄Ac, the tubes were mixed by inversion 4 times then incubated on ice for 10 min. Further protein and other cell debris was pelleted by centrifugation (14000 rpm, 15 min, 4°C) and the supernatant transferred to a fresh tube. Isopropanol was then added (700 μl) and the tube placed on ice for 15 min. Plasmid DNA was collected by centrifugation (15000 rpm, 15 min, 4°C), washed with 300 μl of 70% (v/v) ethanol and dried in vacuo or at 65°C for 10 min. Pellets were resuspended in 40 μl of sterile MQ water and stored at -20°C until further use.

2.8.2 Medium Scale Preparation for Sequence Analysis

10 ml of an overnight bacterial culture was harvested by centrifugation (10 min, 4500 rpm, Centra4X IEC bench centrifuge), and the cells were resuspended in 300 μl of Solution I (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1x 10⁻⁴ g/ml RNase A), and incubated at room temperature for 5 min. Cell lysis was induced by the addition of 300 μl of Solution II (0.2 M NaOH, 1% (w/v) SDS), and a 5 min incubation on ice. To precipitate protein, chromosomal DNA and high molecular weight RNA, 300 μl of Solution III (5 M KAc, pH 4.8) was added, and after a 10 min incubation on ice, the mixture was centrifuged (Eppendorf 5417R centrifuge, 14000 rpm, 10 min, 4°C) and the supernatant transferred to a fresh 1.5 ml reaction tube. Chloroform (400 μl) was added, to extract further protein and cell debris, and the tube was mixed gently by inversion for 30 sec and centrifuged (15000 rpm, 3 min, Heraeus Biofuge 15). The aqueous layer was transferred to a fresh 1.5 ml reaction tube, and the chloroform extraction was repeated. Cold isopropanol (-20°C) (700 μl) was then added and the sample incubated on ice for 10 min, and the tube centrifuged (Eppendorf 5417R centrifuge, 14000 rpm, 5 min, 4°C) to collect plasmid DNA. The pellet was washed with 300
µl of 70% (v/v) ethanol and dried in vacuo or at 65°C for 10 min. Pellets were resuspended in 50 µl of sterile MQ water and stored at -20°C until further use.

2.9 Analysis and Manipulation of DNA

2.9.1 DNA Quantitation

DNA concentration in solutions was determined by measurement of absorption at 260 nm and assuming an OD_{260} of 1.0 is equal to 50 µg DNA/ml dsDNA or 33 µg/ml of ssDNA (Miller, 1972).

2.9.2 Restriction Endonuclease Digestion of DNA

All cleavage reactions using restriction enzymes were performed in reaction buffers as recommended by the manufacturer. 0.1-0.5 µg of plasmid DNA was incubated with 2 units of each restriction enzyme in a final volume of 20 µl, at 37°C, for 2 hr. For restriction of > 1 µg of plasmid DNA, 4-8 units of each restriction enzyme was used in a final volume of 30-50 µl, and incubated overnight at 37°C. The reactions were terminated by heating at 65°C for 20 min, or 85°C for 20 min for enzymes with higher inactivation temperatures. Prior to loading on a gel, a 1/10 volume of tracking dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1 mg/ml RNase A) was added to each sample.

2.9.3 Analytical and Preparative Separation of Restriction Fragments

Electrophoresis of digested DNA was carried out at room temperature on horizontal, 0.8%, 1% or 2% (w/v) agarose gels (Seakem HGT). Gels were electrophoresed at 120 V for 1-3 hr in either 1x TBE buffer (67 mM Tris base, 22 mM boric acid and 2 mM EDTA, final pH 8.8) or 1x TAE (40 mM Tris acetate, 2 mM EDTA). After electrophoresis, the gels were stained in distilled water containing 2 µg/ml ethidium bromide. DNA bands were visualised
by an ultraviolet transilluminator (UVT), and photographed using a Tracktel gel documentation video imager and thermal printer (Mitsubishi).

2.9.4 Calculation of Restriction Fragment Size

The size of restriction enzyme fragments were calculated by comparing their relative mobility with that of EcoRI digested *Bacillus subtilis* bacteriophage SPP1 DNA (Ratcliff *et al.*, 1979). The sizes (kilobases, kb) used were: 8.5; 7.35; 6.1; 4.84; 3.59; 2.81; 1.95; 1.86; 1.51; 1.39; 1.16; 0.98; 0.72; 0.48; 0.36 and 0.09 for SPP1 DNA.

2.9.5 Purification of DNA Fragments

DNA fragments (~2-5 µg) for cloning procedures (from digests) were extracted from 1% TAE-agarose gels using a QIAquick gel extraction kit (QIAGEN) as described by the manufacturer.

2.9.6 Dephosphorylation of DNA

Dephosphorylation of restriction endonuclease digested DNA was carried out using shrimp alkaline phosphotase (SAP). 1-2 µg of digested DNA was added to a mixture containing 2 units of SAP, and 1x dephosphorylation buffer (50 mM Tris-HCl, 0.1 mM EDTA, pH 8.5) in 20 µl. The mixture was incubated for 15 min at 37°C for cohesive DNA ends, or 1 hr at 37°C for blunt DNA ends. The SAP was inactivated at 65°C for 15 min. The dephosphorylated DNA was then used directly in a ligation reaction, or stored at -20°C until further use.

2.9.7 Ligation of DNA Fragments

Ligation reactions with 1-2 units of T4 DNA ligase were performed in 1x ligation buffer A (20 mM Tris-HCl, 10 mM MgCl₂, 0.6 mM ATP, 10 mM DTT and BSA (100 mg/ml)) for cohesive ends, or ligation buffer B (1 mM Tris, pH 7.5, 1 M MgCl₂, 50% PEG, 0.1 M ATP
and 1 M DTT) for blunt end ligations in a volume of 20 μl, and incubated at 4°C for 16 hr. A ratio of approximately 1:5 vector DNA to insert DNA was routinely used. Restriction enzymes were heat inactivated at 65°C or 85°C prior to ligation.

2.9.8 DNA Cloning Procedures

2.9.8.1 In vitro Cloning

DNA to be sub-cloned (200 ng) was cleaved in either single or double restriction enzyme digests. This was combined with 20 ng of similarly cleaved vector DNA, then ligated with 1-2 units of T4 DNA ligase in a volume of 20 μl in ligation buffer A. The ligated DNA was then used directly for transformation of E. coli strains. Transformants were screened for insertional inactivation of LacZα expression wherever possible, prior to plasmid DNA isolation. After isolation plasmids were screened for the correct insert by restriction enzyme digestion or prior to plasmid isolation by PCR screening using small scale lysates.

2.9.8.2 PCR Screening

Lysates were prepared by resuspending a single colony in 50 μl of sterile MQ water followed by heating at 100°C for 5 min, and a brief centrifugation (1 min, 15000 rpm, Heraeus Biofuge 15). 5 μl of the supernatant was used in standard PCR reactions using primers flanking the multiple cloning sites (M13-20 universal primer and M13 reverse primer with pBluescript) or internal gene primers. 10 μl of the PCR reaction was electrophoresed in order to assess the size of the cloned DNA.

2.9.8.3 Generation of Step-Wise Deletions

Uni-directional or bi-directional deletions were generated in plasmid DNA using the Erase-a-base kit from Promega (Madison, Wisconsin). 10 μg of plasmid DNA digested with the appropriate restriction enzymes, and a 3 μl sample was electrophoresed on a 1% agarose gel to ensure the plasmid DNA was completely digested. The digested DNA was extracted
with one volume of TE-saturated phenol:chloroform:isoamylalcohol (25:24:1) and precipitated in 2 volumes of 100% ethanol and 0.1 volume of 3 M sodium acetate, pH 5.2 (Eppendorf 5417R centrifuge, 14000 rpm, 15 min, 4°C). The DNA pellet was resuspended in 10 µl of sterile MQ water. Digestion with exonuclease III was carried out at 37°C (450 bp/min) using 300-500 units of exonuclease III in the presence of exonuclease III buffer (66 mM Tris-HCl, pH 8.0, 0.66 mM MgCl₂). Samples were taken every 15-30 sec and placed in a reaction tube containing 7.5 µl of S1 nuclease mix (0.94 M potassium acetate, pH 4.6, 0.34 M NaCl, 1.35 mM ZnSO₄, 6.75 % (v/v) glycerol containing 2.25 units of S1 nuclease) on ice. All samples were incubated at room temperature for 30 min then 1µl of stop buffer was added (0.3 M Tris base, 0.05M EDTA) and the mixture heated at 70°C for 10 min. The deletion rate was confirmed by electrophoretic analysis of 1 µl of each sample. The DNA was precipitated with 0.3 volume of 7.5 M ammonium acetate and 2 volumes of 100% ethanol and pelleted (Eppendorf 5417R centrifuge, 14000 rpm, 15 min, 4°C). DNA staggered ends were end filled with 3-5 units of Klenow enzyme at 37°C for 3 min in the presence of 20 mM Tris-HCl, pH 8.0 and 100 mM MgCl₂ and then a further 5 min upon the addition of 0.125 mM of dNTPs. The reaction was inactivated at 65°C for 10 min. Samples were then ligated using T4 DNA ligase, transformed into E. coli K-12 strain DH5α and plated onto NA containing Amp (100 µg/ml). Plasmid DNA was prepared from the transformants and analysed electrophoretically to determine the sizes of each deletion, or alternatively deletion sizes were screened by PCR on cell lysates using specific oligonucleotides.

2.10 Polymerase Chain Reaction (PCR) Protocol

2.10.1 Standard PCR Reaction

The protocol used for PCR is that described for the generation of PCR products with cohesive ends (Delidow, 1993). The PCR reaction was performed in reaction tubes (0.5 ml, Perkin Elmer Cetus) in a 50 µl volume containing Taq buffer (50 mM KCl, 10 mM Tris-HCl,
pH 8.3, 1.1 mM MgCl₂, 0.01% (w/v) gelatin), 2 mM final concentration of each deoxynucleoside triphosphate (dNTP), 100 pmol each primer, 200 ng of plasmid template or genomic DNA and 2 U of Taq polymerase (Perkin Elmer Cetus). The reaction was overlaid with a drop of Nujol light mineral oil (Perkin Elmer) and following an initial denaturation period of 5 min at 95°C, was subjected to 25 cycles of amplification (95°C for 30 sec, 55°C for 30 sec, 72°C for 1 min per 1 kb of DNA template to be amplified, followed by a final extension at 72°C for 5 min) using a DNA thermal cycler (Perkin Elmer Cetus). Following PCR, the reaction was carefully removed from under the oil and 10 μl of the reaction product was analysed on a 1% agarose gel, and the remainder was purified by using a QIAquick PCR purification kit (Qiagen) according to manufacturer’s instructions and stored at -20°C until further use.

2.10.2 Cloning of PCR Products

For efficient ligation, a molar ratio of 1:3 of pGEM-T vector (Promega) to purified PCR product was routinely used. A mixture comprised of pGEM-T vector (1 μl), T4 DNA ligase 10x buffer (1 μl), T4 DNA ligase (1 μl), the PCR product and MQ water to 10 μl was left overnight at 4°C. The ligation mixture was then transformed into DH5α and plated onto NA/ Amp/ X-gal/ IPTG plates to select for vector with insert (white colonies).

2.11 DNA Sequencing Protocols

2.11.1 Sequencing Using Dye-Labelled Primers

Sequencing reactions were carried out on 1 μg of double stranded plasmid DNA using the protocol provided by Applied Biosystems. In dye-labelled primer sequencing the DNA was divided into four tubes containing 160 ng (A and C) and 320 ng (G and T) of DNA respectively. To each tube, Ready reaction mix and DNA template were added as follows:
<table>
<thead>
<tr>
<th>Reagent</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ready Reaction Mix</td>
<td>4 µl</td>
<td>4 µl</td>
<td>8 µl</td>
<td>8 µl</td>
</tr>
<tr>
<td>DNA Template</td>
<td>1 µl</td>
<td>1 µl</td>
<td>2 µl</td>
<td>2 µl</td>
</tr>
<tr>
<td>Total Volume</td>
<td>5 µl</td>
<td>5 µl</td>
<td>10 µl</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

Each reaction was overlayed with ~20 µl of light mineral oil and centrifuged briefly. Samples underwent 15 cycles (96°C for 10 sec; 55°C for 5 sec; 70°C for 60 sec), followed by 15 cycles (96°C for 10 sec; 70°C for 60 sec; 15 cycles total), and were then held at 4°C. Reactions were combined in 80 µl of 95% (v/v) ethanol with 3 µl of 3 M sodium acetate and precipitated on ice. The DNA was pelleted at 14000 rpm for 15 min (Eppendorf 5417R centrifuge, 4°C). Samples were dried in vacuo and stored at -20°C.

2.11.2 Sequencing with Dye-Labelled Terminators

Plasmid DNA was purified by ethanol precipitation prior to dye terminator sequencing. Sequencing was performed with kits supplied by Applied Biosystems. 0.5 ml thin walled tubes (Gene Amp, Perkin Elmer Cetus) containing 1-2 µg of template DNA and 3.2 pmol primer, made up to a final volume of 20 µl with 8 µl of pre-mix (Applied Biosystems) and sterile water, were overlaid with mineral oil (Nujol, Perkin Elmer) and subjected to 25 cycles (96°C for 30 sec; 50°C for 15 sec; 60°C for 4 min) before adding 2 µl of 3 M sodium acetate and 50 µl of ice cold 100% ethanol and precipitating for 2 hr at -20°C. Reactions were then centrifuged at 15000 rpm for 15 min at 4°C before washing with 300 µl of 70% (v/v) ethanol and drying in vacuo. Samples were stored at -20°C until further use.

2.11.3 Electrophoresis and Analysis of DNA Sequences

The dried DNA pellets were resuspended in 4.5 µl of loading buffer (83% deionised formamide, 8.3 mM EDTA, pH 8.0) and heated to 95°C for 2 min. Samples were
electrophoresed on a 6% polyacrylamide-8M urea gel in an Applied Biosystems 373A or 377 DNA sequencer. Raw sequence data generated were analysed using Applied Biosystems Seq Ed Program Version 6.0. The electrophoresis was carried out by the Molecular Pathology Sequencing Laboratory, Institute of Medical and Veterinary Science, Adelaide.

DNA sequence data generated was analysed using the LKB DNA and protein analysis programs DNASIS and PROSIS (Hitachi Software). Multiple sequence alignments was carried out using CLUSTAL W (Thompson et al., 1994). Hydropathy plots were generated by the Kyte and Doolittle program in PROSIS (Kyte and Doolittle, 1982) and aligned using PROFILEGRAPH (Hoffmann and Stoffel, 1989). Signal peptide cleavage site predictions were carried out based on the constraints of von Heijne (1985) and the program SignalP Version 1.1 (Nielsen et al., 1997). Secondary structure analysis on predicted proteins was carried out using the Predict Protein Server program PHDSec (Rost et al., 1994). DNA and amino acid sequence homologies were detected using the basic local alignment search tool (BLAST)(Altschul et al., 1990).

2.12 Protein Analysis

2.12.1 SDS-Polyacrylamide Gel Electrophoresis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on 5% stacking and 10 or 15% separating polyacrylamide gels using a modification of the procedure described by Lugtenberg et al. (1975). Samples were heated at 100°C for 5 min in 1x SDS-sample buffer consisting of 25 mM Tris-HCl, pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 5% (v/v) β-mercaptoethanol, and 15% (w/v) bromophenol blue prior to loading. Samples were electrophoresed through the gel at 180 V for 2-3 hr. Gels were stained with Coomassie Brilliant Blue G250 consisting of 5 ml 0.4% Coomassie Brilliant Blue G250 in distilled water (dH₂O), 10 ml perchloric acid and 85 ml of dH₂O. Staining was performed overnight at room temperature with gentle agitation or for 30 min with heating (2x 1.5 min in a 700W microwave oven). Gels were de-stained with several changes of 5% (v/v) acetic acid at room
temperature for 1-2 hr with intermittent heating (1 min in a 700 W microwave oven) with agitation, or overnight.

Size markers (Pharmacia) were α-lactalbumin (14.4 kDa), soybean trypsin inhibitor (20.1 kDa), carbonic anhydrase (30 kDa), ovalbumin (43 kDa), bovine serum albumin (67 kDa) and phosphorylase B (94 kDa). Pre-stained molecular mass standards (New England Biolabs) were: triosephosphate isomerase (32.5 kDa), aldolase (47.5 kDa), glutamic dehydrogenase (62 kDa), MBP-paramyosin (83 kDa) and MBP-β-galactosidase (175 kDa).

2.12.2 Western Immunoblot Analysis

Proteins were transferred to nitrocellulose (Schleicher and Schuell) at 200 mA for 2 hr in a Trans-blot cell (Bio-Rad). The transfer buffer used consisted of 25 mM Tris-HCl, pH 8.3, 192 mM glycine and 5% (v/v) methanol (Towbin et al., 1979). The blot was incubated for 1 hr in 5% skim milk powder in TTBS (0.05% (v/v) Tween-20, 20 mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl) to block non-specific binding sites before incubating in primary antiserum (diluted 1:1000 to 1:5000) overnight at room temperature with gentle agitation. Unbound antibody was removed by washing the filter (3 x 10 min) in TTBS before incubation with HRP-conjugated goat anti-rabbit IgG (diluted 1:10000 to 1:15000) for 2 hr at room temperature. Prior to detection, the filter was washed (5 x 5 min) with TBS (20 mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl). For enhanced chemiluminescence detection (ECL), filters were soaked in Chemiluminescence Blotting Substrate (POD) (Boehringer Mannheim) for 1 min in a transparent plastic bag. Filters were exposed to X-ray film (Cronex) at room temperature for at least 30 sec, and up to 10 min. Films were developed in D19 developer (Kodak) for 2-5 min before being washed and fixed for 5 min in negative fixer (Kodak).
2.12.3 Protein Expression Systems

2.12.3.1 Standard T7 Polymerase-Mediated Expression

The method employed is essentially that described by Tabor and Richardson (1985). The plasmid pGP1-2 carries the T7 RNA polymerase gene under the control of the lambda P_L promoter (Tabor and Richardson, 1985). The E. coli DH5 strain containing pGP1-2 (E2096) was transformed with plasmids encoding the specific genes of interest under control of the T7 RNA polymerase promoter. A 10 ml LB broth with Amp (100 μg/ml) and Kan (50 μg/ml) was inoculated with a single colony of E2096 containing T7 expression constructs, and grown with aeration at 30°C overnight. The culture was centrifuged (10min, 4500 rpm, Centra4X IEC bench centrifuge) and washed in fresh LB (10 ml) then subcultured 1:20 and incubated with constant aeration at 30°C. When an OD_{600nm} of 0.6 was reached, the cultures were shifted to 42°C for 20 min to induce expression of the T7 RNA polymerase. The culture was then transferred to 37°C for a further 3 hr. Cells (1.5 ml) were then pelleted (15000 rpm, 2 min, Heraeus Biofuge 15), and resuspended in 120 μl of 1x SDS-sample buffer. Samples were heated at 100°C for 5 min prior to electrophoresis on SDS-15% polyacrylamide gels.

2.12.3.2 L-[^35S]-Methionine Labelling of Over-expressed Proteins

A modification of the method of Tabor and Richardson (1985) was used to detect radio-labelled proteins. Strain E2096 containing expression constructs were grown at 30°C overnight in LB with Amp (100 μg/ml) and Kan (50 μg/ml). The culture was centrifuged (10min, 4500 rpm, Centra4X IEC bench centrifuge) and washed in fresh LB (10 ml) then subcultured 1:20 in fresh LB with antibiotics, grown to an OD_{600nm} of 0.6, and 600 μl of the culture was pelleted (8000 rpm, 5 min Heraeus Biofuge 15). The cells were washed 3x in 1.5 ml of M9 minimal media and resuspended in 1 ml of methionine assay medium (Difco) and M9 minimal media in the ratio 1:50. Cells were grown at 30°C for a further 2 hr with vigorous aeration. Cells were then shifted to 42°C for 20 min to induce expression of the T7
RNA polymerase, and Rif was added to a final concentration of 200 µg/ml, and incubation continued for a further 20 min at 42°C. The culture was then transferred to 37°C for a further 2 hr followed by pulsing with 10 µCi of L-[^35]S-methionine (Amersham) for 5 min at 30°C. Cells were then pelleted (15000 rpm, 1 min Heraeus Biofuge 15), and resuspended in 120 µl of 1x SDS-sample buffer. Samples were heated at 100°C for 5 min prior to electrophoresis on SDS-15% polyacrylamide gels.

2.12.3.3 (Histidine)₆ Tagged Protein Expression

*E. coli* strain BL21 (DE3) harbouring pREP4, along with pQE31 (Qiagen) based expression constructs were grown at 37°C with aeration overnight in LB (10 ml) containing Kan (25 µg/ml), and Amp (100 µg/ml). Cells were centrifuged (10 min, 4500 rpm, Centra4X IEC bench centrifuge), and the pellet was resuspended in fresh LB (10 ml), and 2 ml was added to a fresh 50 ml broth of LB containing Kan (25 µg/ml) and Amp (100 µg/ml). The culture was grown with aeration at 37°C until the OD₆₀₀nm reached 0.6, when IPTG was added to a final concentration of 1 mM, and the culture grown for a further 3 hr. The culture was centrifuged (10 min, 8000 rpm, Beckman JA-14 rotor) and cells were fractionated to determine the cellular location of the over-expressed protein. A sample of the inclusion body fraction was resuspended in 8 ml of Buffer C (8 M urea, 100 mM NaH₂PO₄, 10 mM imidazole, 10 mM Tris-HCl, pH to 6.3 immediately before use) and 200 µl of washed Ni-NTA agarose (with Buffer C) was mixed at 4°C for 1 hr with gentle agitation in a 10 ml yellow capped tube (Sarstedt). The Ni-NTA agarose was pelleted at 3500 rpm for 5 min (Heraeus Labofuge 400R) and the supernatant gently aspirated. The beads were then washed twice in 4 ml of Buffer C and re-centrifuged each time with the supernatant removed. The bound His-Tagged proteins were eluted in 4x 0.5 ml of Buffer D (100mM NaH₂PO₄, 8 M urea, 10 mM Tris-HCl, pH to 5.9 immediately prior to use) or Buffer E (100mM NaH₂PO₄, 8 M urea, 10 mM Tris-HCl, pH to 4.5 immediately prior to use). A sample of eluted protein
was mixed with an equal volume of 2x SDS-sample buffer and examined by SDS-15% PAGE.

2.13 Autoradiography

Destained SDS-PAGE gels were soaked in 3% (v/v) glycerol for 3 hr prior to being overlayed with cellophane and dried onto Whatman 3MM chromatography paper at 60°C for 3 hr on a Bio-Rad gel drier. L-[35S]-methionine labelled proteins were detected by autoradiography which was performed at -70°C for 2-14 days without intensifying screens using Kodak XR-100 film.

2.14 Cell Fractionation

The cell fractionation procedure was a modification of that described by (Morona et al., 1995). Cells (50 ml) were grown in LB to mid-exponential phase at 37°C (OD600nm of 0.6). Cells were pelleted in a Beckman JA-10 rotor, (7000 rpm, 10 min, 4°C) and resuspended in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1 then transferred to SM-24 tubes and chilled on ice. Cells were converted to sphaeroplasts with 0.1 ml of 1 mg/ml lysozyme in 0.1 M EDTA, pH 7.3 for 30 min on ice. Cells were centrifuged as above and the supernatant collected (periplasmic fraction). The cell pellet was resuspended vigorously in 3 ml of 3 mM EDTA, pH 7.3 and freeze-thawed in an ethanol dry ice bath four times. Cells were lysed with a Branson Ultrasonifier (50% cycle, intermittent) and unlysed cells and inclusion bodies were removed by centrifugation in a Beckman JA-20 rotor (7000 rpm, 10 min, 4°C). The supernatant containing the membranes and the cytoplasm was centrifuged at 35000 rpm using a 50Ti or 80Ti rotor for 90 min at 20°C in a Beckman L8-80 ultracentrifuge. The supernatant (cytoplasmic fraction) was collected and the whole membrane pellet was resuspended in 1 ml H2O. 500 μl of Triton solution (4% Triton X-100, 2 mM MgCl2, 50 mM Tris, pH 7.5) was added to an equal volume of the whole membrane sample which was vortexed intermittently.
for 30 min and kept on ice. The inner (Triton X-100, soluble) membrane fraction was separated from the outer (Triton X-100, insoluble) membrane fraction by centrifugation at 35000 rpm for 90 min in a 50Ti or 80Ti rotor (Beckman L8-80) at 20°C. The pelleted outer membrane fraction was resuspended in 500 μl of MQ water. All fractions were placed at -20°C until further use. A sample of the fractions was mixed with an equal volume of 2x SDS-sample buffer and heated at 100°C for 5 min prior to SDS-15% PAGE.

2.15 N-Terminal Sequence Analysis

Whole membrane extracts were prepared from E. coli K-12 harbouring pPM5631 and electrophoresed on SDS-15% PAGE. After Coomassie Brilliant Blue G250 staining, and destaining with 5% (v/v) acetic acid, the protein of interest (39 kDa) was excised from the gel and re-electrophoresed to remove any contaminating proteins. The protein in the second gel was transferred to pre-wetted (5 min 100% methanol, 2 min transfer buffer) PVDF protein sequencing membrane (BioRad) via electro-blotting at 200 mA for 2 hr in a Trans-Blot cell (BioRad). Transfer buffer was 25 mM Tris-HCl, pH 8.3, 192 mM glycine, and 10% (v/v) methanol. The proteins were detected by staining for 10 min in 0.025% (w/v) Coomassie Brilliant Blue R250 in 40% (v/v) methanol. The N-terminal sequence was determined on an Applied Biosystems Procise-HT 494 Sequencer at the Australian Proteome Analysis Facility, Macquarie University, Sydney by Dr Brad Walsh.

2.16 Polyclonal Antisera Preparation

2.16.1 Protein Preparation and Immunisation Schedule

New Zealand outbred white rabbits (female) were obtained from the Central Animal House, University of Adelaide. SDS-15% PAGE purified protein were used to immunise the rabbits. Briefly, the protein was electrophoresed on SDS-15% PAGE and stained with Coomassie Brilliant Blue G250, destained with 5% (v/v) acetic acid and the protein of interest
was excised from the gel and re-electrophoresed to remove any contaminating proteins. The protein band (corresponding to approximately 250 µg) was excised from the gel and emulsified with 0.5 ml of Freund’s complete adjuvant and an equal volume of 1x PBS in an Ultra-Turrax Blender (Janke and Kunkel). At day 0 the rabbit was immunised subcutaneously with 10 injections of the protein homogenate at different sites (100 µl per injection). The rabbit was boostered 3 times at day 14, day 41 and day 60 with 10 subcutaneous injections at different sites on the rabbit, with a protein homogenate in Freund’s incomplete adjuvant. The rabbit was exsanguinated by cardiac puncture under anaesthesia 14 days after the last immunisation and the serum obtained was incubated at 37°C for 2 hr before storage at -20°C.

2.16.2 Absorption

Rabbit antiserum was absorbed with live E. coli K-12 strain DH5α harbouring plasmid vectors used in the specific cloning. An overnight culture of DH5α was adjusted to 5x 10^10 bacteria/ml and centrifuged (10 min, 4500 rpm, Centra4X IEC bench centrifuge) then resuspended in 10 ml of serum in the presence of 0.02% sodium azide. Absorptions were carried out by incubating the serum at 37°C for 4 hr or alternatively overnight at 4°C. After each absorption, the bacteria were removed by centrifugation (10 min, 4500 rpm, Centra4X IEC bench centrifuge). The process was repeated 4-8 times, and after the final absorption the serum was passed through a 0.2 µm Millipore filter.

2.16.3 Affinity Purification of Antisera

Antisera was purified by a modification to immunoaffinity chromatography by the removal of non-specific antibodies present in the serum (Salamitou et al., 1994). Cell extracts from E. coli K-12 DH5α containing the plasmid vector used for protein expression were electrophoresed on SDS-15% PAGE and transferred onto a nitrocellulose membrane. The membrane was incubated for 1 hr at room temperature in TTBS (0.05% (v/v) Tween-20, 20
mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl) and 5% (w/v) skim milk powder to block non-specific sites. The membrane was then incubated overnight at room temperature with 10 ml of absorbed polyclonal antiserum. Excess antisera was removed, and the membrane washed three times in TTBS followed by one wash in TBS (20 mM Tris-HCl, pH 7.4, 0.9% (w/v) NaCl). The membrane was then stripped of non-specific antibodies by treatment with 10 ml of 0.2 M glycine-HCl buffer (pH 2.2) at room temperature for 15 min. The filter was washed 3 times in TBS and blocked with 5% skim milk in TTBS and the process repeated 3-6 times. The absorbed serum was filtered through a 0.2 μm Millipore filter, and was tested by western blot analysis. A working stock was stored at 4°C with the remainder stored in 1.5 ml aliquots at -20°C until required.

2.17 Pili Preparation

2.17.1 Crude Pili Preparation Method 1

Bacterial strains were grown to confluence overnight at 37°C on CFA agar, containing the appropriate antibiotics where necessary. Cells were harvested in 1 ml of 1x PBS and incubated with shaking at 56°C for 20 min (Eppendorf Thermomixer model 5436). Bacterial cells were removed by centrifugation at 10000 rpm for 10 min in an Heraeus Biofuge 15 bench centrifuge. The supernatant containing the heat extracts was collected and an aliquot mixed with an equal volume of 2x SDS-sample buffer for SDS-15% PAGE. The remainder was stored at -20°C until further use.

2.17.2 Crude Pili Preparation Method 2

Bacterial strains were grown to confluence overnight at 37°C on CFA agar, containing the appropriate antibiotics where necessary. Cells were harvested in 1 ml of 1x PBS and incubated with shaking at 56°C for 20 min and pelleted by centrifugation at 10000 rpm for 10 min (Heraeus Biofuge 15). Trichloroacetic acid (50% v/v) was added to the supernatant to give a final concentration of 5% and incubated on ice for 1 hr. The pili was precipitated by
centrifugation at 14000 rpm for 15 min at 4°C (Eppendorf 5417R centrifuge). The pellet was washed with 100% ethanol and air dried for several hours. The pellets were resuspended in 1x SDS-sample buffer and stored at -20°C until further use.

2.17.3 Caesium Chloride (CsCl) Purification of CS5 Pili

A recombinant *E. coli* K-12 strain which expresses significant amounts of CS5 pili was grown to confluence on large CFA agar plates (30 x 30 cm) with the appropriate antibiotics by plating 1ml of an overnight culture incubated at 37°C. The cells were harvested in 10 ml of 1x PBS and placed into a shaking waterbath at 56°C for 20 min (orbital shaking bath, Puton Industries). The bacterial cells were then removed by centrifugation at 10000 rpm for 10 min at room temperature (Heraeus Biofuge 15). The supernatant was collected and ammonium sulphate was added to 5% saturation with constant stirring for several hours at 4°C. The precipitated protein was pelleted by centrifugation at 15000 rpm for 15 minutes at 4°C (Eppendorf 5417R centrifuge). The protein was then resuspended in 2.5 ml of 1x PBS and mixed with CsCl in 1x PBS to give a final volume of 11 ml and a final concentration of 40% (w/v) CsCl. Samples were loaded into open-top, thin wall ultratubes (Nunc) and spun in an SW-41 rotor at 35000 rpm for 18 hr at 4°C using a L8-80 ultracentrifuge (Beckman). The CS5 pili band was removed from the tube via insertion of a 19 gauge needle attached to a 2 ml syringe and dialysed twice against 1x PBS overnight at 4°C. The excised band containing CS5 pili was checked for purity by SDS-15% PAGE and Coomassie Brilliant Blue (G250) staining and the amount quantified by using a BCA protein assay estimation kit (Pierce). The purified pili was stored at -20°C until further use. A sample of purified CS5 pili (1 mg) was sent to Professor John Heuser at the Washington University School of Medicine, St. Louis, Missouri, USA for Quick-freeze Deep-etch electron microscopy. Samples of purified pili were routinely used in SDS-15% PAGE.
2.18 Slide Agglutination

One half loopful of bacterial cells grown overnight on CFA agar at 37°C were gently resuspended in 5 μl of 1x PBS on a glass slide. A sample of 5 μl of polyclonal antiserum was then mixed with the suspension and the immediate agglutination of the bacteria at room temperature was considered a positive result.

2.19 Mannose Resistant Haemagglutination (MRHA)

Slide haemagglutinations were performed by gentle re-suspension of bacterial cells in 1x PBS after overnight growth at 37°C on CFA agar to an OD

2.20 Immunogold Electron Microscopy

Colloidin-coated copper-palladium grids (200 mesh, Centre for Electron Microscopy and Microstructure Analysis, Adelaide) were placed coat side down onto 40 μl of poly-L-lysine (100 μg/ml) for 15 min and washed twice in MQ water (40 μl per wash), removing excess liquid between each successive wash with pieces of Whatman 3MM chromatography paper. The grid was then placed onto a 40 μl suspension of bacterial cells grown on CFA agar, which had been carefully resuspended in 1x PBS for 5 min. The grid was then blocked twice in 40 μl of 1% (w/v) BSA-PBS before incubation with 40 μl of antiserum (diluted 1:10 in 1x PBS) for 15 min. The grids were washed a further two times in 1% BSA-PBS and incubated for 15 min with 20 μl of a 1:40 dilution of protein A-gold particles (10-20 nm) purchased from Amersham. The grids were washed 2x in 1x PBS (40 μl) then a further wash in MQ water (40 μl) before incubation on 1% uranyl acetate (w/v) for 1 min. The grids were
then inverted and left to dry. Samples were examined using a Phillips CM-100 transmission electron microscope at an acceleration voltage of 80 kV.
Chapter 3

Sequence Analysis and the Initial Characterisation of the CS5 Pilus Encoding Region from Enterotoxigenic *E. coli* O115:H40

3.1 Introduction

Enterotoxigenic *E. coli* (ETEC) is a major cause of acute diarrhoea in children in the developing world and international travellers to these areas (Merson *et al*., 1980; Sack, 1975). ETEC bacteria mediate the disease process by adhering to the epithelial brush border membranes of the small intestine and releasing heat-stable (ST) or heat-labile (LT) enterotoxins, or both. Specific colonisation of the epithelium is mediated through the action of proteinaceous surface structures known as pili or fimbriae. In ETEC strains isolated from humans, such pili are referred to as colonisation factor antigens (CFAs). The CFAs have been implicated in the establishment of infection by interacting with epithelial glycolipids and glycoproteins in a lectin-like fashion, thus promoting bacterial colonisation of the intestine (Levine, 1987).

A number of the gene clusters encoding CFA biosynthesis have been characterised and include CFA/I (Jordi *et al*., 1992b) and CFA/II strains expressing combinations of the surface (CS) antigens CS1 (Froehlich *et al*., 1994; Sakellaris *et al*., 1996; Scott *et al*., 1992), CS2 (Froehlich *et al*., 1995) and CS3 (Jalajakumari *et al*., 1989). CFA/IV is present on the surface of 10-20% of all ETEC strains. CFA/IV ETEC strains express the non-pilus antigen CS6 with either CS5 or CS4 pili, and all are antigenically distinct (Thomas *et al*., 1985). CS5 pili and CS6 are found on strains of serotypes O6, O29, O92, O114, O115 and O167 (McConnell *et al*., 1985; Thomas *et al*., 1985).

The region required for the biosynthesis of functional CS5 pili from ETEC belonging to the O115:H40 serotype, isolated originally during an outbreak of diarrhoea among
Aboriginal children in central Australia, has been cloned (Heuzenroeder et al., 1989; Manning et al., 1987). Transposon mutagenesis identified the minimal coding region of the cloned DNA required for CS5 biosynthesis to be approximately 7 kb in size (Heuzenroeder et al., 1989). The nucleotide sequence of the 23 kDa major CS5 pilin subunit, has been determined (Clark et al., 1992). Immediately downstream of the CS5 major subunit gene a potential stem-loop structure was also identified, which may function as a transcriptional attenuator, thereby reducing the expression of downstream genes (Clark et al., 1992).

This chapter describes the sequencing and initial characterisation of the remaining region of DNA required for biosynthesis of CS5 pili from Enterotoxigenic E. coli O115:H40.

3.2 Results

3.2.1 Sequencing of the CS5 Region from Enterotoxigenic E. coli O115:H40

3.2.1.1 Sub-cloning of the CS5 Region

To facilitate the sequencing of the CS5 region, it was first necessary to generate smaller sub-clones of pPM1312 (Heuzenroeder et al., 1989). This plasmid was digested with PstI to generate two fragments of 4.5 kb and 7.5 kb which were ligated directly into PstI cut pBS-SK+ which had been treated with shrimp alkaline phosphotase (section 2.9.6.), to generate plasmids pPM5302 and pPM5303 respectively (Figure 3.1). Both plasmids contain not only the CS5 region, but also some of the cloning vector from pPM1312.

3.2.1.2 Generation of Nested Deletions from pPM5303 and pPM5302 and DNA Sequencing

Nested deletions were carried out according to section 2.9.8.3. In short, 10 µg of pPM5302 and pPM5303 were digested independently with SacI and XbaI, which both lie within the polylinker region of the vector pBS-SK+, then treated with exonuclease III to delete from the susceptible XbaI 5' overhang of DNA for a given period of time (~450 bp per minute at 37°C). The linear fragment of DNA was then treated with Klenow enzyme and S1
Figure 3.1 Sub-cloning of the CS5 region for sequencing analysis.

pPM1312 (Heazenroeder et al., 1989) was digested with Psfl and the two generated fragments (approximately 4.5 kb and 7.5 kb) which both contain part of the CS5 region, were cloned into Psfl cut pBS-SK⁺. The two plasmids constructed were named pPM5302 and pPM5303 respectively.
Cleave with PstI

Ligation

Cleave with PstI
nuclease, ligated with T4 DNA ligase and transformed into *E. coli* K-12 strain DH5α. The selected nested deletion products were then directly used for sequencing (section 2.11) using either the -20 M13 universal primer or the M13 reverse primer (Table 2.3). Gaps within the sequence were covered by designing synthetic oligonucleotides to sequence across non-overlapping nested deletions. Figure 3.2 outlines the nested deletion products obtained and the synthetic oligonucleotides used.

To ensure unambiguous sequence data, the sequence of the second strand of DNA was determined by designing the following synthetic oligonucleotides #2510, #2541, #2542, #2551, #2552, #2553, #2554, #2555, #2556, #2557, #2558, #2559, #2561, #2563, #2564, #2565, #2566, #2567, #2569 and #2571 (Table 2.3) based on the sequence of the first strand of DNA (not shown). The complete nucleotide (nt) sequence of 9936 bp from the CS5 region is given in Appendix I and has been submitted to the databases at Genbank/EMBL under Accession no. AJ224079. Open reading frames (ORFs) which are thought to be involved in CS5 pili biogenesis have been assigned the genetic nomenclature of *csf* for *coli-surface-factor-five*. The previously described and sequenced major subunit pilin (Accession no. X63411)(Clark *et al.*, 1992) has been designated *csfA* in accordance with the nomenclature proposed for pili biogenesis systems (Gaastra and Svennerholm, 1996). A summary diagram of the entire region is given in Figure 3.3.

### 3.2.2 Analysis of the DNA Sequence Flanking the csf Cluster

The sequence of DNA flanking the *csf* cluster was examined by Genbank/EMBL database searches, with homology identified to a number of insertion sequence (IS) elements or IS-like elements, which are shown in Appendix I. Nucleotides 1-768 represents an intact IS1 insertion sequence, complete with inverted repeats (IRs). This region shares 99% sequence identity and 98% protein identity with the published IS1 sequence (Accession no. V00609)(Ohtsubo and Ohtsubo, 1978). There are two conservative amino acid substitutions (Ser→Lys due to a change of nt C to T at position 263 and His→Gln due to a change of C to A
Figure 3.2 Sequencing strategy for the CS5 region.

Sub-clones generated from the nested deletions of pPM5302 and pPM5303 were sequenced using either the -21 M13 universal primer or the M13 reverse primer. Where required, gaps in the sequence were filled using the synthetic oligonucleotides indicated. Further sequence analysis was performed either upstream or downstream of the CS5 region from pPM1306 (Heuzenroeder et al., 1989) using the synthetic oligonucleotides indicated. Second strand DNA sequencing was carried out using synthetic oligonucleotides designed on the basis of the first strand sequence (not shown). The nucleotide sequence of the major subunit gene has been previously determined (Clark et al., 1992). Relevant restriction sites are indicated.
Figure 3.3 Summary of the genetic organisation of the CS5 region from Enterotoxigenic *Escherichia coli* O115:H40.

Genes belonging to the CS5 cluster are denoted *csfA-F*. The direction of transcription is indicated. The region is flanked by a variety of insertion sequence elements. + region homologous to part of *orfB* from Tn21, * region homologous to part of IS66, # region homologous to part of IS911. IS1 and a defective IS30 are indicated. The upper line refers to the position of the genes relative to the sequence length in kb, with relevant restriction sites indicated. The bold *SphI* and *SmaI* restriction sites were utilised in the construction of pPM5631 (section 3.2.5.1).
at position 468), within one of the genes (insB) required for transposition. It is unknown what effect (if any) this may have on the activity of the transposase. From nt 849-1188 is a sequence which shares 77% protein identity with part of orfB of transposon Tn21 from Shigella flexneri (Accession no. U42226; R.M. Hall, unpublished). A very similar sequence precedes the cooB gene from the CS1 pili operon (Accession no. X62495; Scott et al., unpublished). Immediately downstream of the cpy cluster from nt 8014-8305 is a region which is 34% identical and 54% similar at the protein level to part of an ORF from IS66 (Accession no. M10204)(Machida et al., 1984).

Immediately following IS66 from nt 8377-8710 is a region that shares 96% sequence identity and 96% protein identity to the IS911 left IR (IS911 IR_L) from Shigella dysenteriae and part of the first gene from IS911, orfA (Accession no. X17613)(Prère et al., 1990). Translation of orfA is interrupted by an apparent insertion of a second IS element, IS30. However, this does not result in the termination of translation of orfA, instead a potential hybrid protein could be formed which is unrelated to the remaining amino acids from orfA, and terminates at nt 8816, within the coding sequence for the IS30 transposase (Appendix I). This hybrid protein is identical to a number a hypothetical IS911 sequences from E. coli submitted to the database (Accession no. U70214, AE000499), which suggests a possible form of hot spotting of IS30 elements into IS911 elements. It is unknown whether such a hybrid protein retains transposase activity.

A defective IS30 element is located from nt 8711-9936 which is 98% identical in DNA sequence to the complete IS30 (Accession no. X00792)(Dalrymple et al., 1984), and includes a set of IRs (IS30 IR_L and IR_R). An apparent insertion of 5 nt (5' GCTCA 3') between nt 8947-8948 would be predicted to result in the premature termination of the IS30 transposase at nt 8954 (Appendix I).

To ensure that the IS elements described in this section are not just cloning artefacts derived from E. coli K-12, but also exist in the wild type ETEC strain PE423, from which the CS5 region was derived (Heuzenroeder et al., 1989), a series of PCR reactions was carried out...
Specific oligonucleotides were designed which link each of the IS elements or IS related-sequences with the native csf cluster in PE423 (Table 2.1) and the cloned csf cluster in pPM1312. IS1 (#2969, #2873), IS66 (#2566, #2527), Tn21 (#2551, #2873), IS911 (#2566, #3196) and IS30 (#2566 and #3197) were all detected flanking the csf cluster in PE423, along with the positive control for the PCR, pPM1312. This indicates that the IS elements flanking the csf region are also present in the wild type strain PE423 (data not shown). Oligonucleotide sequences are shown in Table 2.3.

3.2.3 Analysis of the DNA Sequence Constituting the csf Gene Cluster

A summary of each of the newly characterised csf ORFs including the previously characterised csfA (Clark et al., 1992) is shown in Table 3.1. Immediately upstream of the csfA gene are two potential sigma (σ)70 promoter sequences from nt 1262-1291 and nt 1375-1402. The former has −35 (TTGACg) and −10 (TATgT) sites with an 18 nt spacer region and the latter has −35 (TTtCA) and −10 (TATgT) sites separated by a 16 nt spacer region (Appendix I). There is a high degree of identity for both the −35 and −10 consensus sequences (TTGACA and TATAAT), although the spacing between them is not optimal (17 nt), but acceptable (Hawley and McClure, 1983). The identification of only one set potential promoters immediately upstream of csfA, along with overlapping reading frames between csfB-csfC and csfF-csfD ORFs (Appendix I) indicates that this region is likely to constitute an operon.

It has been previously shown that immediately downstream of the major subunit (csfA) was a proposed transcriptional attenuator sequence and the beginning of a second ORF with a good ribosome binding site (5′ GGAA 3′) located 8 nt upstream of the AUG initiation codon, with a predicted signal sequence cleavage site between Phe25-Ser26 (Clark et al., 1992).

This next ORF of the region is csfB, extends from nt 2096-2770 and is predicted to encode a 24 kDa periplasmic protein resulting from a signal sequence cleavage event between Phe25-Ser26 (von Heijne, 1985), with a pI value of 6.42 and a mean hydrophobicity index of −0.41.
Table 3.1 Summary of each of the newly sequenced *csf* ORFs.

<table>
<thead>
<tr>
<th>ORF</th>
<th>No. of amino acids in ORF</th>
<th>Molecular mass (kDa)$^a$</th>
<th>Molecular mass (kDa) of mature protein$^b$</th>
<th>Hydrophobicity mean index$^c$</th>
<th>Predicted pI$^d$</th>
<th>Position in sequence$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>csfA</em></td>
<td>203</td>
<td>20.9</td>
<td>18.6$^e$</td>
<td>0.32</td>
<td>5.74</td>
<td>1427-2038</td>
</tr>
<tr>
<td><em>csfB</em></td>
<td>224</td>
<td>25.8</td>
<td>24.0</td>
<td>-0.41</td>
<td>6.42</td>
<td>2096-2770</td>
</tr>
<tr>
<td><em>csfC</em></td>
<td>814</td>
<td>90.3</td>
<td>88.0</td>
<td>-0.44</td>
<td>8.61</td>
<td>2767-5211</td>
</tr>
<tr>
<td><em>csfE</em></td>
<td>203</td>
<td>22.8</td>
<td>20.3</td>
<td>-0.01</td>
<td>7.72</td>
<td>5227-5838</td>
</tr>
<tr>
<td><em>csfF</em></td>
<td>224</td>
<td>25.5</td>
<td>23.3</td>
<td>-0.21</td>
<td>8.68</td>
<td>5852-6526</td>
</tr>
<tr>
<td><em>csfD</em></td>
<td>373</td>
<td>41.0</td>
<td>38.9</td>
<td>-0.30</td>
<td>8.3</td>
<td>6523-7644</td>
</tr>
</tbody>
</table>

$^a$ Predicted molecular mass of putative protein encoded by the ORF.


$^c$ Generated by the Kyte and Doolittle program (PROSIS) (Kyte and Doolittle, 1982).

$^d$ Numbers are in nucleotides.

$^e$ Confirmed size from SDS-15% PAGE by Clark *et al.*, (1992) indicated a molecular mass of 23 kDa.
The following ORF, termed *csfC*, which overlaps *csfB* by 4 nt, extends from nt 2767-5211. A potential ribosome binding site (5' AAGA 3') is located 8 nt upstream of the AUG start codon (Shine and Dalgarno, 1974). The *csfC* ORF is predicted to encode a 90.3 kDa protein with a pI value of 8.61 and a mean hydrophobicity index of -0.44. A potential signal sequence cleavage site between Ala21-Asp22 would result in a mature protein of 88 kDa in the periplasm. The overlap of the reading frames for *csfB* and *csfC* is suggestive of translational coupling (McCarthy and Gualerzi, 1990).

Immediately downstream of *csfC* initiating, within 17 nt, is a fourth ORF denoted *csfE*, which extends from nt 5227-5838. Analysis of this ORF suggests it may encode a protein of 22.8 kDa with a pI value of 7.72 and a mean hydrophobicity index of -0.01. A potential signal sequence cleavage site exists between Leu20-Gln21 giving rise to a mature protein of 20.3 kDa in the periplasm. The *csfE* reading frame is preceded by a potential ribosome binding site (5' GAAGAG 3') located at an optimal distance of 7 nt from the AUG start codon.

The fifth ORF in the cluster termed *csfF*, lies 13 nt downstream of *csfE* and extends from nt 5852-6526, giving rise to a 25.5 kDa protein with a pI of 8.68 and a mean hydrophobicity index of -0.21. A potential signal sequence cleavage site located between Ala19-Phe20 would give rise to a 23.3 kDa protein in the periplasm. A potential ribosome binding site (5' GAGA 3') is located 8 bases upstream from the AUG start codon.

The final ORF in the cluster is *csfD*, which overlaps *csfF* by 4 nt, is again indicative of translational coupling. This ORF extends from nt 6523-7644 and is predicted to encode a protein of 41 kDa with a pI value of 8.3 and a mean hydrophobicity index of -0.3. A 38.9 kDa mature protein in the periplasm would be produced from the potential signal sequence cleavage site located between Ala19-Ala20. The potential ribosome binding site is located 10 nt upstream of the AUG start codon (5' AAGG 3'), which is not optimal for efficient translation (Gold *et al.*, 1981; Kozak, 1983), but may be offset by translational coupling with *csfF*.

The average G+C content of the six ORFs comprising the *csf* cluster is 37.1% compared with the average G+C content of 50.8% in *E. coli* (Blattner *et al.*, 1997). The pattern of codon
usage within the csf cluster also significantly differs from the codon usage of E. coli. The rare or modulator codons of E. coli include CUA (leucine), AUA (isoleucine), AGA/AGG/CGA/CGG (arginine) and GGA/GGG (glycine), since their isoaccepting tRNAs are a very minor species (Grosjean and Fiers, 1990). These codons are used much more frequently in the csf cluster when compared to E. coli highly and minimally expressed genes (Anderson and Kurland, 1990) as shown in Table 3.2. This may indicate that the csf cluster was introduced recently into E. coli from another organism by transposition. This may be further supported by the presence of several IS elements flanking the csf cluster.

3.2.3.1 Transcriptional Terminators

An analysis of the DNA sequence downstream of the stop codon of the last ORF in the csf cluster, csfD, did not identify any Rho-independent terminator sequences, characteristically composed of secondary stem-loop structures (Rosenberg and Court, 1979). It is therefore possible that termination may occur via a Rho-dependent mechanism, in which termination occurs from a C over G rich region upstream of the 3' endpoint of the mRNA (Alifano et al., 1991). However, in this system the 3' end of the mRNA is unknown.

3.2.4 Homologies

3.2.4.1 CsfC is Homologous to the CooC Membrane Usher Protein of CS1 Pili

A search of the databases at Genbank/EMBL through NCBI with the derived sequence of the CsfC protein showed greatest homology with the membrane usher protein CooC from CS1 pili (Accession no. 76908)(Froehlich et al., 1994) with 19% protein identity and 38% similarity (Figure 3.4). Weak homology with other proposed membrane usher proteins including CfaC from the CFA/I gene cluster (Accession no. M55661)(Jordi et al., 1992b) and CotC from the CS2 gene cluster (Accession no. Z47800)(Froehlich et al., 1995) was also identified.
Table 3.2  Codon usage of each of the csf genes compared to E. coli high and low expressed genes.

The percentage represents the usage of a particular codon encoding an amino acid against other codon groups which encode the same amino acid. The table compares the codon usage of the csf genes relative to those described for E. coli high and low expressed genes (Anderson and Kurland, 1990) for those codons that constitute rare or modulator codons (Grosjean and Fiers, 1990), which are CUA (leucine), AUA (isoleucine), AGA/AGG/CGA/CGG (arginine) and GGA/GGG (glycine).

<table>
<thead>
<tr>
<th>Codon</th>
<th>Amino Acid</th>
<th>% in csfA</th>
<th>% in csfB</th>
<th>% in csfC</th>
<th>% in csfD</th>
<th>% in csfE</th>
<th>% in csfF</th>
<th>% in E. coli (high)</th>
<th>% in E. coli (low)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUA</td>
<td>Leu</td>
<td>7.7</td>
<td>15.4</td>
<td>12.1</td>
<td>33.3</td>
<td>21.1</td>
<td>29.2</td>
<td>0.6</td>
<td>3.5</td>
</tr>
<tr>
<td>AUA</td>
<td>Ile</td>
<td>27.2</td>
<td>32.0</td>
<td>28.0</td>
<td>33.3</td>
<td>35.5</td>
<td>28.6</td>
<td>0.2</td>
<td>10.9</td>
</tr>
<tr>
<td>CGA</td>
<td>Arg</td>
<td>0.0</td>
<td>30.0</td>
<td>21.9</td>
<td>0.0</td>
<td>11.1</td>
<td>21.4</td>
<td>0.0</td>
<td>9.6</td>
</tr>
<tr>
<td>CGG</td>
<td>Arg</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>11.1</td>
<td>7.4</td>
<td>0.2</td>
<td>10.8</td>
</tr>
<tr>
<td>AGA</td>
<td>Arg</td>
<td>0.0</td>
<td>40.0</td>
<td>18.8</td>
<td>22.2</td>
<td>33.3</td>
<td>42.9</td>
<td>0.0</td>
<td>4.6</td>
</tr>
<tr>
<td>AGG</td>
<td>Arg</td>
<td>0.0</td>
<td>10.0</td>
<td>18.8</td>
<td>44.4</td>
<td>11.1</td>
<td>7.4</td>
<td>0.0</td>
<td>2.9</td>
</tr>
<tr>
<td>GGA</td>
<td>Gly</td>
<td>30.8</td>
<td>45.4</td>
<td>37.5</td>
<td>66.7</td>
<td>61.5</td>
<td>32.2</td>
<td>0.5</td>
<td>14.6</td>
</tr>
<tr>
<td>GGG</td>
<td>Gly</td>
<td>23.1</td>
<td>0.0</td>
<td>28.1</td>
<td>11.1</td>
<td>7.7</td>
<td>19.4</td>
<td>1.1</td>
<td>16.7</td>
</tr>
</tbody>
</table>
Figure 3.4 Comparison of amino acid sequences of CsfC and CooC.

The homologous proteins were initially found with BLAST (Altschul et al., 1990) searches of databases at NCBI. Identical amino acids are indicated by an asterix (*), conserved residues with a dot (.), gaps (-) have been introduced to optimise alignment. The analysis was performed with CLUSTAL W (Thompson et al., 1994).
Secondary structure analysis using PHDSec (Rost et al., 1994) predicted a high β-sheet (44.5%) and low α-helical (9.8%) content for this protein, along with a high percentage of charged amino acids (18.4%) without long regions of hydrophobic amino acids (data not shown). Similar properties were defined for the CooC protein (Froehlich et al., 1994), and such features are commonly found in other known outer membrane proteins (Nikaido and Vaara, 1985).

Although both CsfC and CooC share a common percentage of both β-sheets, α-helices and charged residues, when the two proteins were aligned on the basis of the respective hydropathy plots, the CsfC protein did not appear to share any structural similarities with either CooC or CfaC (data not shown). The CooC and CfaC proteins are structurally related, which is not unexpected given that CooC is 64% identical and 78% similar to CfaC (Froehlich et al., 1994).

3.2.4.2 CsfB is Homologous to CsfF

BLAST searches of the NCBI databases (Altschul et al., 1990) revealed that the amino acid sequence of CsfB shared 30% protein identity and 51% protein similarity to CsfF (Figure 3.5). Sequence alignment between the two proteins shows several stretches of identical residues and/or conserved residues, which may possibly serve as functional domains for the proteins. Hydropathy plot alignment between CsfB and CsfF shows that the two proteins are also structurally related (Figure 3.6).

Database searches using advanced BLAST at EMBL for CsfB showed no near homologues other than CsfF (P> 0.98). Interestingly, when the same search was performed with CsfF, a region of this protein was weakly homologous to part of the PapD chaperone (P=0.89)(Accession no. X61239)(Lindberg et al., 1989). PapD was then aligned with CsfF and CsfB (Figure 3.7). Very little protein identity or conservation existed when all three proteins were aligned, with CsfF and CsfB were found to share only three out the ten and four out of ten respectively of the invariant residues associated with the 26 known members of the superfamily
Figure 3.5 Comparison of the amino acid sequences of CsfB and CsfF.

The homologous proteins were initially found with BLAST [Altschul, 1990 #191] searches of databases at NCBI. Identical amino acids are indicated by an asterix (*), conserved residues with a dot (.), gaps (-) have been introduced to optimise alignment. Analysis was performed with CLUSTALW software [Thompson, 1994 #118].
| CsfB | MKILYSPLLPLPPS---CASVDSMIKFSGEDDFPLVIENGKREYIYVTLSELISEKNN 57 |
| CsfF | MFFRAILIVFPPGGQDQIDTLTKVIDKDTQYIEITGEYERIEYIYTQLTVQTLTVK 60 |
| CsfB | RDEIFYNAQHPVTASAEADIIISSGEEQVKIKIKNMYTPVGDRIFGINFSPDTLDN 117 |
| CsfF | GLREIPFPNPCISSWPIUEPGEIVLDDKAKIIRVQKIRNPQODRVSGLAFIEPKVR- 119 |
| CsfB | NDRQ---YNIPFGYKAILVPGTESESTVDVSKEKDNSNINNVMDVADYCG 174 |
| CsfF | RKIQDGLQISVGYKVLIPGKSPLKQIKASKKS---GNITENGNKLRIVPPDCS 177 |
| CsfB | SYNNKCRVQLTRPYSEKKIEISNENPIEFTFSIYIGERKLIKRIKIL 224 |
| CsfF | GKNKSECAGAVILLPYTSK--QID---DSEHVQTLSTYLINDLHKIRKVTITL 224 |
Figure 3.6 Hydropathy plot alignment of CsfB and CsfF.

The hydropathy plots were generated by the method of Kyte and Doolittle (1982) and aligned using PROFILEGRAPH [Hoffmann, 1989 #924]. Positive numbers on the Y-axis indicate hydrophobic regions. The position of every 10th amino acid is marked on each X-axis. The length of each protein in amino acids (aa) is indicated.
Figure 3.7 Comparison of the amino acid sequences of CsfF and CsfB with PapD.

The homologous proteins were initially found and with advanced BLAST searches of databases at EMBL (Altschul et al., 1990). Identical amino acids are indicated by an asterix (*), conserved residues with a dot (.), gaps (-) have been introduced to optimise alignment. Analysis was performed with CLUSTALW software (Thompson et al., 1994). The ten invariant residues associated with the superfamily of immunoglobulin-like chaperones [Lindberg, 1989 #223] are indicated within PapD as boxed amino acids, and where identity exists, the boxed residues extend into CsfB or CsfF or both.
of the immunoglobulin-like chaperones, of which PapD is the prototypic member (Hung et al., 1996).

### 3.2.4.3 CsfA Shares Weak Homology with CsfD

Searches of the databases at NCBI indicated no significant protein or DNA homologues to the CsfD protein (P>0.65). However, BLAST protein sequence alignment between CsfA and CsfD showed the proteins were weakly homologous with 14% protein identity and 24% similarity (Figure 3.8). Although the protein identity between CsfA and CsfD is low, this identity is similar in percentage to the minor tip-associated pilin subunits CooD, PapG and FimH with their cognate major pilins in CS1, Pap and type I pili respectively (Girardeau and Bertin, 1995; Sakellaris and Scott, 1998). This may give an indication that CsfD could possibly function as a minor pilin subunit, given that CsfA has been previously shown to encode the major pilin subunit (Clark et al., 1992).

CsfE showed no significant homology to any relevant protein sequences lodged into the NCBI databases using BLAST or when using advanced BLAST at EMBL (P>0.21).

### 3.2.5 Construction of a Suitable CS5 Pili Expression System

#### 3.2.5.1 Re-Cloning of the csf Region without Flanking IS Elements

To establish a suitable expression system for the csf genes only, it was necessary to remove the flanking IS elements. Therefore, pPM1312 was digested with Sphi and SmaI, which isolates the csf genes (including the potential promoters) from the IS elements (Figure 3.3). The Sphi-SmaI fragment was cloned into Sphi-SmaI cut pGEM-7Zf* in the opposite orientation to the lac promoter (ie behind the T7 promoter) to generate pPM5631 (Figure 3.9).

#### 3.2.5.2 Expression of pPM5631 in E. coli K-12

To examine whether pPM5631 contains all the information necessary for biogenesis of CS5 pili, slide agglutinations, haemagglutinations and immunogold electron microscopy were
Figure 3.8 Comparison of the amino acid sequences of CsfA and CsfD.

Identical amino acids are indicated by an asterix (*), conserved residues with a dot (.), gaps (-) have been introduced to optimise alignment. The analysis was performed with CLUSTAL W software (Thompson et al., 1994).
<table>
<thead>
<tr>
<th></th>
<th>CsfA</th>
<th>CsfD</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>MKKN------------------</td>
<td>SVLAMATVSGS-----------</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>VLAAVTNGQLTF</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>NWQG------VVPSAPVTQS</td>
<td>LDIPFTPG-TEQLNIT-----</td>
</tr>
<tr>
<td></td>
<td>SWAFFNG---------</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>TVRSNNGCCSVSSAYLGN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AYTLEYEIRSSNDVTRNC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GSQTQYTHMPNNSQVNV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGYR</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>-----LDS---NKDITARSVK</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PYDFF-----------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</td>
<td>ANINSVNAFLSSVPVSNGF</td>
<td>AVEAAKGEVAILNGQALK---</td>
</tr>
<tr>
<td></td>
<td>VSNFKQTLST---------</td>
<td>VGSAS</td>
</tr>
<tr>
<td>300</td>
<td>VTLPTVDNIFDRTIGRQIQ</td>
<td>KLQSENVGSASSA</td>
</tr>
<tr>
<td></td>
<td>QNVLIPFVINGHFTPYN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TISLQVISING?KLQSEN</td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>PTVVTVASNK-KESHISID</td>
<td>AAAADV-AEGAAINFVAPVF</td>
</tr>
<tr>
<td></td>
<td>MNAK-------AAAADV-</td>
<td>AVDI----------</td>
</tr>
<tr>
<td></td>
<td>AEGAAINFVAPVF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AVDI----------</td>
<td></td>
</tr>
<tr>
<td>360</td>
<td>TIPYLSAMNTGISBERRY</td>
<td>IDKNTAVTGDY</td>
</tr>
<tr>
<td></td>
<td>SLATNGNGLNVTTINNL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ESQGYS1QGRFNAFD.L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IDKNTAVTGDY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>373</td>
<td>ADTLTAIFQISLL</td>
<td></td>
</tr>
</tbody>
</table>

**Alignment Scores:**

<table>
<thead>
<tr>
<th></th>
<th><strong>CsfA</strong></th>
<th><strong>CsfD</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>180</td>
<td>180</td>
<td>180</td>
</tr>
<tr>
<td>104</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>240</td>
<td>240</td>
<td>240</td>
</tr>
<tr>
<td>157</td>
<td>157</td>
<td>157</td>
</tr>
<tr>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>203</td>
<td>203</td>
<td>203</td>
</tr>
<tr>
<td>360</td>
<td>360</td>
<td>360</td>
</tr>
<tr>
<td>373</td>
<td>373</td>
<td>373</td>
</tr>
</tbody>
</table>
Figure 3.9 Cloning of the csf region including potential promoters without flanking IS elements.

A SphI-SmaI fragment containing all of the csf genes without flanking IS elements from pPM1312 (Heuzenroeder et al., 1989) was ligated into SphI-SmaI cut pGEM-7Zf+ to give pPM5631 with the csf genes orientated from the T7 promoter.
Cleave with **SmaI**, **SphI**

**Ligation**

Cleave with **SmaI**, **SphI**

**pPM5631**

**CS5 major subunit**
performed. This plasmid was transformed into *E. coli* K-12 strain DH5α for the assessment of CS5 pili biogenesis. The formation of cell surface pili was initially examined by slide agglutinations (section 2.18) by resuspending half a loop full of bacteria grown overnight on CFA agar in 5 µl of 1x PBS on a flat glass microscope slide and adding an equal volume of rabbit anti-native CS5 pili antiserum (Heuzenroeder *et al.*, 1989). Immediate agglutination was considered a positive result. *E. coli* K-12 harbouring pPM5631 strongly agglutinated, the wild type CS5 expressing strain PE423 also agglutinated, while *E. coli* K-12 harbouring pGEM-7Zf+ did not agglutinate, even after several minutes of observation. This suggests that the *csf* cluster may encode all the information necessary for cell surface expression of CS5 pili.

To assess whether the CS5 pili produced showed adhesive properties, slide haemagglutinations were performed (section 2.19). Bacterial cells were resuspended in 1x PBS at an OD$_{600nm}$ = 4.0 and a 10 µl sample was mixed with an equal volume of washed 5% human group A+ erythrocytes in the presence of 0.1M D(+)-mannose. CS5 pili has been previously shown to exhibit mannose resistant haemagglutination (MRHA) of human group A+ erythrocytes (Manning *et al.*, 1987). Visible haemagglutination of the erythrocytes after one minute of gentle rocking was considered a positive result. *E. coli* K-12 with pPM5631 was positive, while the same strain with pGEM-7Zf+ was clearly unable to mediate haemagglutination even after several minutes of observation (Figure 3.10). Therefore pPM5631 contains all the information necessary to form a functional CS5 pilus.

The same strains were examined for CS5 pili expression by immunogold electron microscopy using the rabbit anti-native CS5 pili serum (section 2.20). *E. coli* K-12 containing pPM5631 exhibited peritrichous pili (Figure 3.11, panel B) some of which were found detached in the surrounding milieu (Figure 3.11, panel D) and were in far greater amounts than the corresponding wild type strain PE423 from which the DNA region was cloned (Figure 3.11, panel A). *E. coli* K-12 containing pGEM-7Zf+ was devoid of CS5 pili (Figure 3.11, panel C). These data clearly indicate that the *csf* cluster alone encodes all information required for expression of functional CS5 pili in *E. coli* K-12.
Figure 3.10 Haemagglutination activity mediated by pPM5631 in *E.coli* K-12.

Slide haemagglutinations were performed using a 10 μl sample of resuspended bacteria in PBS (OD₆₀₀ₙ㎜= 4.0) mixed with an equal volume of 5% human group A⁺ erythrocytes in the presence of 0.1M D(+) mannose. Agglutinated erythrocytes are visible as granules on the slide.
Figure 3.11 Immunogold electron microscopy of PE423 and E. coli K-12 derivatives.

Polyclonal antiserum raised against native purified CS5 pili was used at a concentration of 1:10, with protein A-20 nm gold conjugate used at a dilution of 1:40. A. PE423. B. E. coli K-12 containing pPM5631. C. E. coli K-12 containing pGEM-7Zf+. D. CS5 pili in the cell milieu. Bar represents 500 nm.
Interestingly, expression of CS5 pili in *E. coli* K-12 results in the pili forming bundled structures, which in some cases were over 20 nm in width. This is not unexpected, given the high level of expression achieved in *E. coli* K-12, but also since CS5 pili is readily precipitated at < 5% ammonium sulphate making it one of the most hydrophobic pili characterised to date (Heuzenroeder *et al.*, 1989). Similar bundled structures were observed in the wild type strain PE423, particularly so when this strain was grown on CFA agar in the presence of 0.15% bile salts.

### 3.2.6 Protein Expression Studies

#### 3.2.6.1 Cloning of Individual *csf* Genes Behind the T7 Promoter

The *csfB, csfC, csfD, csfE* and *csfF* genes are predicted to encode proteins of sizes 25.8 kDa, 90.3 kDa, 41 kDa, 22.8 kDa and 25.5 kDa respectively (Table 3.1). In order to confirm this, each gene was PCR amplified using specific oligonucleotides and cloned into pBS-SK+ orientated behind the T7 promoter. The plasmids constructed could then be compared directly with pPM5631, which contains all of the *csf* cluster orientated behind the T7 promoter.

*csfB* was PCR-amplified from pPM5631 with oligonucleotides #2944 and #2945 (Table 2.3), which included *EcoRI* and *XbaI* sites, respectively. This fragment was digested with the appropriate restriction enzymes and cloned into *EcoRI* and *XbaI* cut pBS-SK+ to generate pPM5605 (Figure 3.12).

pPM5636 was constructed by using oligonucleotides #2977 and #2978 (Table 2.3) which include *PstI* and *XbaI* sites respectively to PCR-amplify *csfC* from pPM5631. The PCR fragment was then digested with the appropriate enzymes and cloned into *PstI* and *XbaI* cut pBS-SK+ (Figure 3.13).

*csfD* was PCR-amplified from pPM5631 with oligonucleotides #2979 and #2980 (Table 2.3), which include *PstI* and *XbaI* sites, respectively. This fragment was digested with the appropriate restriction enzymes and cloned into *PstI* and *XbaI* cut pBS-SK+ to generate pPM5640 (Figure 3.14).
Figure 3.12 Cloning strategy of the csfB gene behind the T7 promoter.

The gene was initially PCR amplified from pPM5631 using the synthetic oligonucleotides indicated. The PCR fragment was then digested with EcoRI and XbaI and ligated into EcoRI and XbaI cut pBS-SK+ orientated behind the T7 promoter to give pPM5605. The arrow above the gene indicates the orientation.
PCR with oligonucleotides #2944, #2945

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

pPM5631

pBS-SK+

pPM5605
Figure 3.13 Cloning strategy of the csfC gene behind the T7 promoter.

The gene was initially PCR amplified from pPM5631 using the synthetic oligonucleotides indicated. The PCR fragment was then digested with PstI and XbaI and ligated into PstI and XbaI cut pBS-SK\(^+\) orientated behind the T7 promoter to give pPM5636. The arrow above the gene indicates the orientation.
PCR with oligonucleotides #2977, #2978

Cleave with PstI and XbaI

Ligation

Cleave with PstI and XbaI

Ligation

pPM5631

pBS-SK+

pPM5636
Figure 3.14 Cloning strategy of the \textit{csfD} gene behind the T7 promoter.

The gene was initially PCR amplified from pPM5631 using the synthetic oligonucleotides indicated. The PCR fragment was then digested with \textit{Pst}I and \textit{Xba}I and ligated into \textit{Pst}I and \textit{Xba}I cut pBS-SK\textsuperscript{*} orientated behind the T7 promoter to give pPM5640. The arrow above the gene indicates the orientation.
pBS-SK$^+$
lacZ$^a$
pPM5631

(+) origin
-RAmp

T3

T7

Psfl
cslD
PsI

PCR with oligonucleotides #2979, #2980

XbaI

ColE1 origin

Cleave with PstI and XbaI

Ligation

Cleave with PstI and XbaI

pPM5640

CSF region

Amp$^R$
csfD

Amp$^R$

XbaI

PstI
csfE was PCR-amplified from pPM5631 with oligonucleotides #2973 and #2974 (Table 2.3), which include PstI and XbaI sites, respectively. This fragment was digested with the appropriate restriction enzymes and cloned into PstI and XbaI cut pBS-SK+ to generate pPM5633 (Figure 3.15).

Finally, the construction of pPM5390 required oligonucleotides #2971 and #2972 (Table 2.3) which PCR-amplified csfF from pPM5631 and included EcoRI and XbaI sites respectively. The PCR fragment was digested with EcoRI and XbaI and ligated into EcoRI and XbaI cut pBS-SK+ (Figure 3.16).

3.2.6.2 Specific Expression of CsfB, CsfC, CsfD, CsfE and CsfF Proteins

Expression of the csf genes was carried out by transforming each of the plasmids into the T7 RNA polymerase strain E2096 (Table 2.1) containing pGP1-2. Over-expressions and L-[^35S]-methionine labelling of proteins were carried out as described in section 2.12.3.2.

Several protein bands were detected from pPM5631, which correlated well with the sizes of the individual over-expressed gene products (Figure 3.17). CsfE and CsfD proteins migrated with an apparent molecular mass in approximate agreement based on their respective DNA sequences (Figure 3.17). A band corresponding with the expected size of CsfE in pPM5631 may either be CsfE or CsfA, which also migrates at approximately 23 kDa (Clark et al., 1992). CsfF migrated with an apparent molecular weight of 27 kDa, slightly higher than that predicted based on DNA sequence alone. Similarly, the apparent molecular mass of CsfB was 26 kDa which was slightly higher than its predicted molecular mass from DNA sequence studies (Figure 3.17). The CsfC protein was predicted to have a molecular mass of 90.3 kDa and the mature form 88 kDa, however T7 analysis indicates an apparent molecular mass of only 71 kDa (Figure 3.17).
The gene was initially PCR amplified from pPM5631 using the synthetic oligonucleotides indicated. The PCR fragment was then digested with PstI and XbaI and ligated into PstI and XbaI cut pBS-SK$^+$ orientated behind the T7 promoter to give pPM5633. The arrow above the gene indicates the orientation.
PCR with oligonucleotides #2973, #2974

Cleave with PstI and XbaI

Cleave with PstI and XbaI

Ligation
Figure 3.16 Cloning strategy of the csfF gene behind the T7 promoter.

The gene was initially PCR amplified from pPM5631 using the synthetic oligonucleotides indicated. The PCR fragment was then digested with EcoRI and XbaI and ligated into EcoRI and XbaI cut pBS-SK+ orientated behind the T7 promoter to give pPM5390. The arrow above the gene indicates the orientation.
PCR with oligonucleotides #2971, #2972

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

PCR with oligonucleotides #2971, #2972

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI

LIGATION
Figure 3.17 T7 RNA polymerase / promoter expression of *csf*B, *csf*C, *csf*D, *csf*E and *csf*F gene clones in pBS-SK+.

*E. coli* strains carrying either pGEM-7Zf+, pPM5631, pBS-SK+, pPM5390, pPM5636, pPM5605, pPM5633 or pPM5640 with pGP1-2 were over-expressed using the T7 polymerase/promoter system of Tabor and Richardson (1985). Proteins were labelled with L-[35S]methionine (Amersham) and whole cell fractions were subjected to SDS-15% PAGE followed by autoradiography of the dried gel. Molecular mass standards are indicated. Arrows indicate the positions of the over-expressed proteins.
pGEM-7Zf+
pPM5631
pBS-SK+
pPM5390 (CsfF)
pPM5636 (CsfC)
pPM5605 (CsfB)
pPM5633 (CsfE)
pPM5640 (CsfD)
3.3 Summary and Conclusions

This chapter described the sequencing and initial analysis of the CS5 region from enterotoxigenic *E. coli* O115:H40. The CS5 region consists of five further ORFs (*csfB, csfC, csfE, csfF* and *csfD*) downstream of the previously characterised major subunit *csfA* (Clark *et al.*, 1992). Analysis of the DNA sequence indicated that all the ORFs were capable of encoding proteins, with characteristic signal sequence cleavage sites (von Heijne, 1985). Based on sequence analysis alone, the *csfB, csfC, csfE, csfF* and *csfD* genes were predicted to encode 24 kDa, 88 kDa, 20.3 kDa, 23.3 kDa and 38.9 kDa mature proteins, respectively. Furthermore, the *csfB* and *csfC* ORFs and *csfF* and *csfD* genes were shown to overlap, which is suggestive of translational coupling. This, along with the identification of one set of putative promoter sequences upstream of *csfA*, indicate the CS5 region is likely to constitute an operon.

T7 polymerase-mediated over-expression of individually cloned genes was shown to produce protein products in approximate agreement with the predicted sizes, with the exception of CsfC, which migrated faster than predicted. The 71 kDa CsfC product may represent a degradation product which is readily observed in the *E. coli* K-12 background, while the full length form is not detected. Proteolytic degradation products are a common occurrence among other known outer membrane usher proteins including CooC (Voegele *et al.*, 1991), PapC (Dodson *et al.*, 1993), FaeD (Valent *et al.*, 1995) and FasD (Schifferli and Alrutz, 1994). However, it is worth noting there is an unusual distribution of charged residues (18.4%) within the coding sequence for CsfC, which may also account for the faster migrating band observed.

As with other ETEC pili systems, the *csf* cluster is flanked by a number of distinct IS elements, which includes a complete IS1 element, a defective IS30 element and several DNA regions which show homology to other known IS elements including IS66, IS911 and *orfB* from *Tn*21. The presence of flanking IS elements has been noted in the ETEC pili systems CS1 (Froehlich *et al.*, 1994; Scott *et al.*, 1992), CS2 (Froehlich *et al.*, 1995), CS3 (Jalajakumari *et al.*, 1989) and CFA/I (Hamers *et al.*, 1989; Jordi *et al.*, 1992b).
CS5 pili may be representative of a distinct class of pili, since the protein products encoded by four of the six csf genes comprising the CS5 region show very little homology to any other pili related sequences lodged into the Genbank/EMBL databases. CsfB was shown to be homologous to CsfF in terms of protein sequence (30% identity, 51% similarity), but also the two proteins appeared to share structural homology when hydrophobicity plots were aligned. Furthermore, CsfA was shown to bear some weak protein homology (14% identity, 24% similarity) with CsfD, which is similar to that shared between major and minor pilins in other systems characterised to date. Also based on homology searches, the csf cluster does not encode a regulatory protein, although it is possible that any potential regulator may not necessarily located on this contiguous region of DNA, but elsewhere.

Expression of the cloned csf region in E. coli K-12 produced cell surface CS5 pili in far excess of that observed in the wild type strain PE423. This may be due to a gene dosage effect, a consequence of the expression of the csf cluster from a high copy number plasmid versus expression from the single or low copy virulence plasmids found in the wild type ETEC strains. Alternatively, in E. coli K-12 there may be a lack of regulation on this region compared to the wild type strain. CS5 pilus biogenesis from the high copy number plasmid in E. coli K-12 provides a useful system for investigating phenotypic changes in expression resulting from mutational and complementation experiments. Consequently, the next four chapters describe the characterisation of the unknown genes of this region, namely csfB, csfC, csfD, csfE and csfF.
Chapter 4

The Role of CsFD in CS5 Pilus Biogenesis

4.1 Introduction

In a number of the well described pilus biogenesis systems, the pili are principally composed of a single repeating polypeptide known as the major subunit, but also one or more minor components are incorporated into the pilus fibre. Both Pap pili and type I pili are composite fibres with distinct fibrillar tips (Jones et al., 1995; Kuehn et al., 1992). The tip fibrillum from Pap pili is composed of four distinct subunits, PapK, PapE, PapF and the adhesin PapG (Jacob-Dubuisson et al., 1993). PapF and PapK not only function as adaptor proteins for the ordered assembly of the fibrillar structure, but are also required to initiate the formation of the tip fibrillae and pilus rods (Jacob-Dubuisson et al., 1993).

In contrast, CS1 pili are composed almost entirely of the major subunit CooA, with a distally located tip protein CooD predicted to contribute only one subunit to the pilus structure (Sakellaris et al., 1996). CooD is necessary for the extracellular transport of CooA, while the level of expression of CooD dictates the number of CS1 pili assembled in E. coli. Therefore, CooD expression is rate-limiting for CS1 pili biosynthesis (Sakellaris et al., 1999b).

The role of the minor pilins in pilus-mediated adherence to host epithelial cells depends on the pilin type. For instance, the minor pilin FimH from type I pili (Hanson and Brinton, 1988), PilC from Neisseria type IV pili (Rudel et al., 1995) and PapG from Pap pili (Lund et al., 1987) have been identified as pilus tip adhesin proteins essential for adherence to host tissues. Recent experiments on CS1 pili have found a point mutation in CooD, which does not interfere with CS1 pilus formation, abolished haemagglutination, suggesting that
CooD is involved in adherence (Sakellaris et al., 1999a). Conversely, specific adhesion of K99 pili to host tissues is determined by the major pilin subunit (Jacobs et al., 1987b).

This chapter describes the characterisation of the last gene constituting the csf cluster, csfD, which is shown to encode an integral minor subunit of CS5 pili. Furthermore, the morphology of the CS5 pilus is elucidated by high resolution electron microscopy on rotary shadowed pili samples.

4.2 Results

4.2.1 Construction of a csfD Deletion Mutant

Due to the lack of any significant homologous proteins in the databases at Genbank or EMBL (P>0.65), the csfD gene was targeted for deletion in the working CS5 pili expressing plasmid pPM5631, to determine its role in the production of CS5 pili in E. coli K-12. pPM5647 was constructed from pPM5631, by removing the last 563 nt of the csfD gene by digesting pPM5631 with BamHI (within the vector sequence) and AccI which removes from nt 7082 onwards of the characterised CS5 region (Figure 4.1, Appendix I). The cut plasmid was then agarose gel-purified and treated with Klenow enzyme, ligated using T4 DNA ligase and transformed into E. coli K-12 strain DH5α (Table 2.1). No polar effects on downstream genes would be expected, since csfD has previously been demonstrated to be the last gene constituting the csf cluster (Figure 3.3).

4.2.2 Construction of a csfD Complementing Plasmid

pPM5644, which contains the csfD gene under control of the inducible lac promoter on the low copy number plasmid pWKS130 (Wang and Kushner, 1991), was constructed by PCR amplifying the csfD gene from pPM5631 using oligonucleotides #2979 and #2980 (Table 2.3) which contain PstI and XbaI sites respectively. The PCR fragment was digested with the appropriate restriction enzymes and cloned into PstI and XbaI cut pWKS130 to give pPM5644.
Figure 4.1 Construction of a csfD deletion mutant.

Plasmid pPM5631 was digested with AccI and BamHI which deletes 563 nt of the csfD gene from nt 7082 onwards of the characterised sequence (Appendix I). The cut plasmid was treated with Klenow enzyme and re-ligated to give pPM5647 which contains the entire csf region with the modified csfD gene. The orientation of the csfD gene is indicated by an arrow.
Cleave with *BamHI, AccI*

Religation

modified *csf region (ΔcsfD)*
(Figure 4.2). This plasmid was used in \textit{trans} to complement the introduced deletion of the \textit{csfD} gene.

4.2.3 The Role of CsfD in CS5 Pili Biogenesis

The construction of a deletion derivative of \textit{csfD} in the working plasmid pPM5631 provided a system in which the effects of such a deletion on CS5 pili biogenesis could be examined experimentally. The following sections describe the characterisation of the CsfD protein in detail.

4.2.3.1 Analysis of Pilus Assembly

To initially investigate the effect of a deletion in \textit{csfD} on CS5 pili biogenesis, heat extracts were prepared by gently harvesting confluent bacteria grown on CFA agar (with the appropriate antibiotics) with 1x PBS and adjusting the concentration of bacteria to 1x $10^{10}$ cells in a 500 µl volume. The strains were then incubated with mild shaking at 56°C for 20 minutes. The supernatant fraction resulting from centrifugation is the heat extract sample (section 2.17.1). Cell supernatants were prepared by adjusting the concentration of bacteria to 1x $10^{10}$ cells in a 500 µl volume, then pelleting the cells and collecting the supernatant. A sample of 40 µl of either heat extracts or supernatants was mixed with an equal volume of 2x SDS-sample buffer and loading directly onto SDS-15% PAGE gels for western immunoblot analysis (section 2.12.2).

Heat extracts and cell supernatants were examined for the presence of the major subunit CsfA by western immunoblot using rabbit polyclonal antiserum raised against purified native CS5 pili (Heuzenroeder \textit{et al.}, 1989). CsfA was detected in both heat extracts and supernatants in \textit{E. coli} K-12 strains containing pPM5631 and \textit{E. coli} K-12 with pPM5647 complemented with pPM5644 (Figure 4.3). No CsfA was detected in the heat extracts or supernatants derived from \textit{E. coli} K-12 containing pPM5647, which suggests that CsfD is absolutely required for the transport of the CsfA subunits across the outer membrane (Figure 4.3).
Figure 4.2 Construction of csfD in the low copy number vector pWKS130.

The csfD gene was PCR amplified using oligonucleotides #2979 and #2980 which contain PstI and XbaI restriction sites respectively, from the template pPM5631. The amplified csfD gene fragment, which is orientated according to the direction of the arrow, was then digested with the appropriate restriction enzymes and ligated into PstI and XbaI cut pWKS130 (Wang and Kushner, 1991) to generate pPM5644 which contains the csfD gene behind the lac promoter.
PCR with oligonucleotides #2979, #2980

Cleave with PstI and XbaI

Ligation

Cleave with PstI and XbaI

Kan R
Figure 4.3 Effect of CsfD on assembly and release of CS5 pili in *E. coli* K-12.

Heat extracts and cell supernatants were prepared from the various strains (section 2.17.1). The western immunoblot was performed using rabbit anti-native CS5 pili antiserum (1:5000) on either heat extracts or cell supernatants of *E. coli* K-12 harbouring either pGEM-7Zf, pPM5631, pPM5647 or pPM5647 + pPM5644. Samples represent 1x 10⁹ cells. The solid arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
The anti-CS5 pili antiserum was found to contain antibodies specific for both CsfA and CsfD. Deletion of the csfD gene results in the abolition of a protein band corresponding to the confirmed size of the CsfD protein; however, CsfD is observed in immunoblots after complementing with pPM5644 (Figure 4.3). Therefore CsfD may represent a contaminant of the original pili purification used to immunise rabbits or more likely be intimately associated with the pilus structure (discussed later).

The formation of cell surface CS5 pili in this mutant was also assessed by slide agglutination (section 2.18). *E. coli* K-12 harbouring plasmids pGEM-7Zf+ or pPM5647 did not agglutinate, whereas strong agglutination was observed with *E. coli* K-12 harbouring pPM5631, but also *E. coli* K-12 with pPM5647 complemented with pPM5644. This shows that CsfD is required for formation of CS5 pili on the cell surface.

The same strains were also examined visually for surface CS5 pili formation by immunogold electron microscopy using the anti-CS5 pili antiserum (section 2.20). In short, bacterial strains were coated onto Cu/Pd 200 mesh grids (Centre for Electron Microscopy and Microstructure Analysis, University of Adelaide) then incubated with a 1:10 dilution of the anti-CS5 pili polyclonal antiserum followed by several blocking and washing steps and incubation with a 1:40 dilution of protein A-20 nm gold conjugate, followed by negative staining with 1% uranyl acetate. *E. coli* K-12 containing pPM5631 showed peritrichous pili (Figure 4.4, panel A), while *E. coli* K-12 containing pPM5647 did not exhibit visible pili either attached to the cells, or fixed in the milieu (Figure 4.4, panel B). Formation of CS5 pili was restored when *E. coli* K-12 harbouring pPM5647 was complemented with pPM5644 (Figure 4.4, panel C). Therefore, CsfD is essential for the assembly of CS5 pili.

Both the csfD deletion strain and the complemented strain were also tested for adhesive properties by the ability to exhibit MRHA of red blood cells (section 2.19). *E. coli* K-12 harbouring pPM5647 and pPM5644 caused the haemagglutination of human group A+ erythrocytes, while the csfD mutant strain (*E. coli* K-12 with pPM5647) was unable to mediate haemagglutination (Figure 4.5).
Figure 4.4 Immunogold electron microscopy.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pPM5631, B. pPM5647, C. pPM5647 + pPM5644, or D. pGEM-7Zf+. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
Figure 4.5 Requirement of CsfD for CS5 pilus mediated haemagglutination.

Samples of *E. coli* K-12 containing pPM5647 or pPM5647 + pPM5644 were prepared by gently resuspending the bacteria to an OD_{600nm} = 4.0 in 1x PBS. 10 µl was mixed with an equal volume of human group A+ erythrocytes in the presence of 0.1 M D(+)-mannose and gently rocked (section 2.19). Clear haemagglutination of the erythrocytes, which appear as darkened granules only occurs when the *csfD* mutation is restored (pPM5647 + pPM5644).
4.2.3.2 Stability of CsfA Subunits in the Periplasm

Since a deletion in the csfD gene abolished the formation of cell surface CS5 pili, it was determined whether this effect was mediated through periplasmic degradation of CsfA subunits, therefore abrogating their delivery and assembly across the outer membrane. Periplasmic extracts were prepared from E. coli K-12 strains by resuspending 2x 10^10 cells in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1. Cells were converted to sphaeroplasts with 100 µg/ml of lysozyme in 0.1 M EDTA, pH 7.3 and the resulting supernatant collected after centrifugation was the periplasmic fraction (section 2.14). Whole cell samples were prepared by adjusting the harvested bacteria to 1x 10^10 cells in a volume of 500 µl. For both periplasmic and whole cell extracts, 40 µl of sample was mixed with an equal volume of 2x SDS-sample buffer and analysed by western immunoblots following SDS-15% PAGE. E. coli K-12 harbouring pPM5647 resulted in approximately equal amounts of CsfA protein in both whole cell and concentrated periplasmic preparations when compared to E. coli K-12 containing pPM5631 or pPM5647 complemented with pPM5644 in western immunoblot analysis (Figure 4.6). Therefore CsfD does not affect both the delivery of CsfA into the periplasm and its stability within the periplasm.

4.2.3.3 Cell Fractionation

Since removal of CsfD has such a dramatic effect on CS5 pili biogenesis it was necessary to determine its cellular location. E. coli K-12 harbouring pPM5631 was fractionated into the following components: whole cells (w/c), inclusion bodies (i/b), supernatants (s/n), cytoplasms (cyto), periplasms (peri), whole membranes (w/m), inner membranes (i/m) and outer membranes (o/m) according to section 2.14. The various fractions were analysed by western immunoblot (section 2.12.2) using the anti-CS5 pili antiserum, which has specificity for the CsfD protein. CsfD was not only present in periplasmic fractions, but also readily detected in outer membrane fractions (Figure 4.7). A doublet corresponding to the processed and
Figure 4.6 Effect of CsfD on the stability of CsfA major subunits in the periplasm.

Western immunoblot analysis of A. Concentrated periplasmic extracts and B. Whole cell samples from *E. coli* K-12 harbouring either pGEM-7Zf+, pPM5631, pPM5647 or pPM5647 + pPM5644, using anti-native CS5 pili antiserum. The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Each sample represents 2 x $10^9$ cells (periplasm) and 1 x $10^9$ cells (whole cells). Molecular mass standards are indicated.
Figure 4.7 Cell fractionations.

Western blot analysis of whole cell fractionations derived from E. coli K-12 harbouring pGEM-7Zf* or pPM5631 to determine the cellular location of CsfD using rabbit anti-native CS5 pili antiserum. Bacteria were fractionated according to section 2.14. Whole cell (w/c); inclusion body (i/b); supernatant (s/n); cytoplasm (cyto); periplasm (peri); whole membrane (w/m); outer membrane (o/m) and inner membrane (i/m) fractions are indicated. The empty arrow indicates CsfD, the solid arrow CsfA.
unprocessed forms of CsfA is evident in the whole cell and inclusion body fractions. The two major outer membrane porins of *E. coli* (OmpC/OmpF) served as controls for outer membrane and inner membrane fraction contamination in Coomassie Brilliant Blue staining and were not detected in the inner membrane fraction (data not shown).

4.2.3.4 Association of CsfD with the CS5 Pilus

The association of CsfD with CS5 pili was determined experimentally. A sample of 10 μg of whole purified CS5 pili (section 2.17.3) in 1% SDS-sample buffer without β-mercaptoethanol was directly compared to depolymerised CS5 pili in 1% SDS-sample buffer without β-mercaptoethanol boiled for 5 min. CS5 pili would not be expected to enter a SDS-15% PAGE gel in a polymerised (native) form, but only in a denatured monomeric form.

In a Coomassie Brilliant Blue stained gel, no native CS5 pili entered the gel, although a small amount of monomeric CsfA was detected (Figure 4.8A, lane 1). When a sample of the heat treated denatured pil: was run in parallel, two distinct protein monomer bands were detected (Figure 4.8A, lane 2). The major band corresponded in size to CsfA while the minor band corresponded to the size expected for CsfD. These bands were confirmed by western immunoblot analysis to be CsfA and CsfD (Figure 4.8B, lane 2). These data show that CsfD is an integral pilus protein, and is a minor component of the pilus structure. Furthermore, the relative amounts of CsfD detected compared to CsfA indicate that CsfD may constitute a significant proportion of the pilus structure. It is also clear that CsfD, in the context of the CS5 pilus, is relatively stable and is not destabilised during purification, since CsfD is only released from the pilus structure in denaturing conditions, which includes heating the pili at 100°C for 5 min.

4.2.3.5 The Ratio of CsfD to CsfA in the CS5 Pilus

The contribution of CsfD to the total amount of CS5 pili in the caesium chloride purified samples was determined semi-quantitatively by densitometry analysis against known quantities
Figure 4.8 Association of CsfD with CS5 pili.

A. Coomassie Brilliant Blue staining of SDS-15% PAGE with 10 µg of CS5 pili in the presence of 1%-SDS loaded. Lane 1, unboiled sample; lane 2, boiled sample. B. Western immunoblot of (A) using anti-native CS5 pili antiserum (1:5000). Lane 1, unboiled sample; lane 2, boiled sample. The solid arrow indicates CsfA protein, the empty arrow CsfD. Molecular mass standards are indicated.
of bovine serum albumin (BSA) (Crawford and Beckerle, 1991) using a Bio-Rad FX Scanner and Quantity One software (Bio-Rad). Samples of purified BSA ranging from 1-60 μg were run on SDS-15% PAGE and stained with Coomassie Brilliant Blue and analysed against CS5 pili samples up to 150 μg. The large quantity of purified pili samples was necessary to adequately resolve CsfD and assess the contribution of this protein to the total amount. Comparison against the BSA standard curve of densitometry values obtained for the CsfD band indicated that approximately 2.25 μg of CsfD is present in 125 μg of pili and 2.6 μg in 150 μg of pili, thereby showing that CsfD may constitute between 1.2-1.8 % of the total pilus antigen (data not shown).

It must be noted that throughout the thesis, western immunoblot analysis using enhanced chemiluminescence (ECL) probably over represents the true contribution of CsfD to the pilus structure, due to the non-quantitative nature of ECL. However, Figure 4.8A shows that the contribution of CsfD to the total pilus protein is significant by Coomassie Brilliant Blue staining.

4.2.4 Generation of Specific Antiserum to CsfD and N-terminal Sequence Analysis

To assess the location of CsfD in the pilus it was necessary to generate specific polyclonal antiserum against the CsfD protein. Attempts at immuno-affinity purifying significant quantities of CsfD antibodies from the polyclonal CS5 pili antiserum failed. Therefore, CsfD was over-expressed from plasmid pPM5640, which contains the csfD gene behind the T7 promoter (Figure 3.14) according to section 2.12.3.1. The induced bacterial cells were then fractionated into the following: whole cell (w/c); inclusion body (i/b); periplasm (peri); cytoplasm (cyto); whole membrane (w/m); outer membrane (o/m); and inner membrane (i/m) samples (section 2.14). The over-expressed CsfD protein localised with both the inclusion body and whole membrane fractions (Figure 4.9).

The band corresponding to CsfD was gel excised from SDS-15% PAGE whole membrane samples following Coomassie Brilliant Blue staining and re-run on a second SDS-15% PAGE. The protein was then transferred onto PVDF (polyvinylidene difluoride) membranes and subjected to N-terminal sequencing by Dr. Brad Walsh, Macquarie University,
Figure 4.9 Over-expression and fractionation of CsfD from pPM5640.

Coomassie Brilliant Blue stained SDS-15% PAGE of *E. coli* K-12 harbouring pBS-SK* or pPM5640. The 39 kDa CsfD protein is indicated by an asterix (*). Over-expression of CsfD from pPM5640 by temperature shift (30-42°C) to induce the λP₇ promoter is as described in section 2.12.3.1. The bacteria were fractionated as described in section 2.14. Whole cells (w/c), inclusion body (i/b), periplasm (peri), cytoplasm (cyto) and whole membrane (w/m) fractions are shown. Molecular mass standards are indicated.
New South Wales, Australia (section 2.15). N-terminal sequence analysis of CsFD confirmed the signal sequence cleavage site of the CsFD protein to be between Ala$_{19}$-Ala$_{20}$ as predicted based on sequence data alone, but also further verified the 39 kDa protein recognised by the anti-CS5 pilus antiserum as being CsFD (Figure 4.10).

For generation of the polyclonal antiserum, the gel excised CsFD band was homogenised with 500 μl of Freund's complete adjuvant and an equal volume of 1x PBS and used to immunise a rabbit subcutaneously as described in section 2.16.1. The antiserum was then absorbed extensively against whole cells (section 2.16.2) and neutralised across cell extract samples of E. coli K-12 containing pBS-SK$^+$ which were electrophoresed on SDS-15% PAGE and transferred onto nitrocellulose membranes prior to non-specific blocking with 5% (w/v) skim milk in TTBS (section 2.16.3). The absorbed antiserum obtained was able to slide agglutinate E. coli K-12 harbouring pPM5631, but was unable to agglutinate the negative control strain (E. coli K-12 with pGEM-7Zf$^+$) which suggested that the antiserum was binding cell surface CsFD in the context of CS5 pili attached to the bacterium. The antiserum was specific for CsFD and not the major subunit CsFA, since western blot analysis using denatured purified CS5 pili showed a single band corresponding to the 39 kDa CsFD protein with no cross-reactivity against the 23 kDa major subunit CsFA (data not shown).

4.2.4.1 Specificity of Anti-CsFD Antibodies in Immunogold Electron Microscopy

To examine the possible location of the CsFD protein within the CS5 pilus structure, immunogold electron microscopy was performed according to section 2.20. Firstly, the bacteria were incubated with anti-CsFD antiserum which was bound by protein A conjugated with 20 nm diameter gold particles and then the step was repeated using anti-CS5 pili antiserum which was bound by protein A conjugated to 10 nm diameter gold particles. Interestingly, very little anti-CsFD 20 nm gold conjugate was bound along the CS5 pilus structures in E. coli K-12 with pPM5631 in most cases. This may indicate that either CsFD is not located along the pilus shaft consisting of CsFA, which was labelled with 10 nm gold, or the antibodies are sterically hindered
Figure 4.10 N-terminal sequence analysis of the CsfD protein.

The N-terminal sequence of the CsfD protein was determined as described in section 2.15. The derived sequence showed that the signal sequence cleavage site is between Ala₁₉-Ala₂₀ of the pre-CsfD form of the protein, which is indicated by an arrow.
-Ser$_{14}$-Phe$_{15}$-Met$_{16}$-Val$_{17}$-Gln$_{18}$-Ala$_{19}$-Ala$_{20}$-Thr$_{21}$-Thr$_{22}$-
from accessing CsfD epitopes within the pilus itself, due to the presence of CsfA. Of the bacteria observed, only a relatively small proportion of the pili was labelled with the 20 nm gold particles bound to rabbit anti-CsfD antibodies, which were located predominantly at the distal ends of pili bundles, or on the cell surface (Figure 4.11, panel A-C). Immunogold analysis using anti-CsfD antiserum conjugated with protein A-20 nm gold alone showed gold particles away from the bacteria towards the tip regions of the shadowed CS5 pili (Figure 4.11, panel D). The cell surface location of CsfD antibodies may indicate initiation points for CS5 pili, since CsfD has been shown to be required for CS5 pili initiation.

Since CS5 pili show a tendency to bundle due to the hydrophobicity of the pilus itself, it was not possible to determine whether the antibody had truly bound across the tip regions, or whether the antibodies were binding across fracture points in the pilus thereby exposing the CsfD protein. It can be considered that there is two potential locations of CsfD. Firstly, CsfD is located at the tip of the CS5 pilus, assembled as a fibrillar structure analogous to Pap or type I pili (Jones et al., 1995; Kuehn et al., 1992), or alternatively CsfD may be located throughout the pilus, which is similar to the minor pilins of K88 or K99 pili (Oudega et al., 1989; Simons et al., 1991). The data presented is unable to distinguish between the two possibilities, given the subjectivity of immunogold analysis. Therefore, purified CS5 pili was examined by high resolution quick-freeze deep-etch electron microscopy, a technique which enables visualisation of thin fibrillar structures to a high degree of detail (Kuehn et al., 1992).

4.2.5 Ultrastructure Analysis of CS5 Pili by Quick-Freeze Deep-Etch Electron Microscopy

Given the difficulty in establishing the exact location of CsfD within CS5 pili, a sample of purified CS5 pili (section 2.17.3) was sent to Professor John Heuser (Washington University School of Medicine, St. Louis, Missouri, USA) to examine the pili by high resolution quick-freeze deep-etch electron microscopy (Heuser, 1989). The technique employs the adsorption of the pili to mica chips that were quick-frozen, then freeze-fractured and deep-etched before rotary replication with platinum (Heuser, 1989). This technique has been used to successfully resolve
Figure 4.11 Immunogold electron microscopy using anti-CsfD antiserum.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring pPM5631. The bacteria were reacted with rabbit anti-CsfD antiserum (1:10) and then with protein A coupled to 20 nm diameter gold particles (1:40)(panel D). Following washing, the bacteria were then reacted with rabbit anti-CS5 pili antiserum (1:10) followed by the addition of protein A coupled to 10 nm diameter gold particles (1:40)(panel A-C). Arrows indicate the positions of several 20 nm gold particles binding at the tip regions of CS5 pili in all panels. Bar represents 200 nm.
the fine structures of Pap, type I and *Haemophilus* type IV pili (Jones et al., 1995; Kuehn et al., 1992; St. Geme et al., 1996). This showed that these pili are composite structures composed of a rigid shaft approximately 6 nm in diameter and a thin distinct fibrillar tip approximately 2 nm in diameter at the distal end of the pilus shaft.

The CS5 pili examined were thin fibrillar structures approximately 2 nm in diameter and arranged in a helical structure (Figure 4.12). The numerous bends and turns exhibited by the pili indicates that this fibrillar structure is flexible. Therefore, in morphological terms, CS5 pili is similar to the fibrillar structures of K88 or K99 (de Graaf and Gaastra, 1994). In most cases, no distinct tip region could be identified on the pili, which suggests that CsfD does not assemble to form a distinct fibrillar structure, but is incorporated directly into the growing pilus as a minor component. However, several distinct regions were identified on individual pilus structures (Figure 4.12, panel A-B) but in the majority of cases such regions were present on shortened fractured pili. Such regions may therefore be the remnants of fracture points along the pilus and could possibly result from the heterogeneity of the pilus composed of CsfA and CsfD, leading to weak fracture points between these two proteins. Such fracture points may then expose CsfD for access by the anti-CsfD antiserum, as observed in immunogold experiments (Figure 4.11). The open helical form was the dominant species of pili observed under the electron microscope, however a tightly packed helical form was also detected (Figure 4.12, panel C).

### 4.2.6 Secretion of CsfD in the Absence of CsfA

To verify that CsfD is the first pilin subunit across the outer membrane, and therefore initiates pilus biogenesis, a deletion in the csfA gene was constructed. Plasmid pPM5631 was digested with *ApaI*, which is located across nt 2069-2074 immediately between the csfA and csfB genes within the proposed stem loop attenuator structure (Clark et al., 1992), and also within the vector sequence. The cut plasmid was then re-ligated with T4 DNA ligase to generate pPM5646 (Figure 4.13). This plasmid lacks the predicted CS5 promoter sequences, and therefore pPM5646 was digested with *ApaI* and *SacI* and cloned directly into *ApaI* and *SacI* cut
Figure 4.12 Morphology of the CS5 pilus by high resolution electron microscopy.

Purified CS5 pili samples were subjected to quick-freeze deep-etching followed, by rotary shadowing according to Heuser et al. (1989). Panels A-B highlight distinct tip regions present on several of the individual flexible, open helical pilus rods (arrows and the circled region). Panel C shows two tightly packed helical pilus samples (arrows) surrounded by the open form pilus structures which dominate. Bar represents 40 nm.
Figure 4.13 Construction of a *csfA* deletion mutant.

Plasmid pPM5631 was digested with *ApaI*, which is located between *csfA* and *csfD* (Appendix I), but also in the vector sequence, and re-ligated to remove the *csfA* gene and give pPM5646. This plasmid was then digested with *ApaI* and *SacI* to excise the remaining *csf* genes and cloned into *ApaI-SacI* cut pBS-SK+ to generate pPM5648 in which the remaining *csf* genes are orientated from the *lac* promoter.
Digest with Apal and SacI

Digest with Apal and SacI

Ligation

modified csf region (ΔcsfA)

modified csf region (ΔcsfA) from lac
PBS-SK+ to generate pPM5648 in which the entire CS5 gene sequence minus csfA is orientated from the inducible lac promoter. The removal of the stem-loop structure from pPM5631 by digestion with ApaI may serve to increase expression of the downstream csf genes.

Slide agglutinations were performed on E. coli K-12 containing pPM5648 using either rabbit anti-CsfD or anti-CS5 pili antiserum (section 2.18), and in both cases no agglutination of the bacteria resulted. Furthermore, when the same strains were tested for MRHA of human group A+ erythrocytes (section 2.19), no haemagglutination resulted (data not shown). Therefore CsfA is required to provide a structural pilus framework for either recognition by specific antibodies or for cross-linking human erythrocytes.

Immunogold electron microscopy (section 2.20) was performed on E. coli K-12 harbouring pPM5648 using both the rabbit anti-CsfD antiserum and the anti-CS5 pili antiserum. In no cases did the antisera recognise CsfD on the cell surface on the bacterial cell envelope, which suggests that CsfD cannot be assembled as discrete pilus-like structures in the absence of CsfA, or these structures, if present, are not detected by the antiserum used (Figure 4.14A). Although no CsfD was detected on the surface of the bacterial cells, CsfD is released by these strains, since supernatant samples from E. coli K-12 harbouring pPM5648 or pPM5648 complemented with csfD on the low copy number plasmid pPM5644, contained CsfD (Figure 4.14B). This data provides further evidence that CsfD is the first CS5 pilin assembled into the CS5 pilus structure, and therefore initiates pilus biogenesis.

4.3 Summary and Conclusions

In this chapter, one of the genes constituting the csf gene cluster, csfD was characterised further. CsfD was shown to be an integral minor pilin subunit required for the assembly of CS5 pili, but not for transport of the major subunit CsfA into the periplasm and its stability within the periplasm. It was also shown that CsfD is required for initiation of pilus nucleation, since deletion of csfD results in the complete absence of CsfA subunits in the supernatant or assembly into functional pili. Furthermore, when a specific deletion in the csfA gene was
Figure 4.14 Assembly of CS5 pili and cell supernatant analysis of the csfA mutant.

A. Immunogold electron microscopy was performed according to section 2.20 on *E. coli* K-12 harbouring pPM5648 grown on CFA agar in the presence of ampicillin and IPTG. The bacteria were reacted first with rabbit anti-CS5 pili and anti-CsfD antisera (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles followed by negative staining with 1% uranyl acetate. Bar represents 200 nm. B. Western immunoblot on cell supernatants (equivalent to 5x 10⁶ cells) from *E. coli* K12 harbouring either pPM5648 or pPM5648 + pPM5644 using rabbit anti-CsfD antiserum (1:2500). The CsfD protein is indicated by an arrow.
constructed, although no detectable cell surface-associated CsfD was present, CsfD was
observed in the supernatant of the csfA deletion strain, thereby further verifying that CsfD is the
first pilin subunit assembled across the outer membrane.

This observation is similar to the proposed role of the distally located CooD minor pilin
of CS1 pili, which is essential for the extracellular transport of CooA, and functions as a rate-
limiting initiator of pilus assembly (Sakellaris et al., 1999b). However, unlike CooD, the
relative amounts of CsfD detected in purified depolymerised CS5 pili suggests CsfD
contributes more than a single subunit per pilus. In support of this, semi-quantitative
densitometry analysis on purified CS5 pili samples indicated that the contribution of CsfD is
between 1.2-1.8 % to the total pilus antigen.

Several experiments were carried out to determine the location of CsfD within the pilus
structure. Immunogold electron microscopy using anti-CsfD antiserum did show some binding
along the tips of pili bundles, but these results were largely subjective, given the occurrence of
multiple pili associating directly with each other. Interestingly, the antiserum against CsfD was
able to inhibit MRHA of human group A* erythrocytes by E. coli K-12 expressing CS5 pili to a
dilution of 1:32 (data not shown), and although this may implicate CsfD as a possible adhesin,
the steric effects of the antibodies on important receptor recognition sites cannot be ruled out.

High resolution electron microscopy using quick-freeze deep-etched pili, which were
then rotary replicated with platinum, identified the CS5 pilus as a 2 nm flexible fibrillar
structure, and not the semi-rigid pilus morphology as earlier proposed (Heuzenroeder et al.,
1989). Other morphological studies on CS5 pili suggested the pili were 5-6 nm in diameter
and appeared to be composed of two fine fibrils arranged in a double helical structure (Knutton
et al., 1989). This observation may have been the result of two individual pili bundling
together. There were two forms of CS5 pili observed. The first was a tightly packed helical
conformation and the second was a slightly unravelled open helical form of the same pilus.
Both pili had approximately the same diameter of approximately 2 nm.
The high resolution electron microscopy did identify several putative tip regions on the CS5 pilus which may be composed of CsfD; however, these regions were mostly on broken pili which suggests they may be an artefact. The majority of the pili did not contain distinct tip regions. Therefore, the favoured location of CsfD in CS5 pili is that it is located distally and then incorporated along the growing pilus at intervals determined by the stoichiometric ratio of CsfD and CsfA in the periplasm. The incorporation of CsfD along the shaft may explain the flexibility of the CS5 fibrillar structure, but also may lead to fragile regions which are susceptible to breakage under stressful conditions.
Chapter 5

The Role of CsfE in CS5 Pilus Biogenesis

5.1 Introduction

There are several theories by which the length of a polymeric structure such as pili can be determined. Firstly, the length of a pilus may be the result of spontaneous fractures of the fibre, which would presumably become more widespread with increasing length. Alternatively, the length may be determined by an equilibrium between polymerisation and depolymerisation of the fibre, which occurs in actin filaments (Pollard and Cooper, 1986), but also has been shown in F pili and Pseudomonas pili (Bradley, 1972; Novotny and Fives Taylor, 1974). However, the control of pilus length usually depends on a specialised minor component of the pilus which, when incorporated into the polymerising structure, terminates further growth of the pilus.

PapH from Pap pili has been shown to be a minor subunit involved in length modulation, but also cell anchoring (Båga et al., 1987). It was suggested that the stoichiometric relationship between PapH and the major pilin subunit PapA determines the pilus length (Båga et al., 1987). PapH is located at the base of the pilus rod. In other analogous systems, MrpB from Proteus mirabilis, which is homologous to PapH, also terminates assembly of MRP-like fimbriae (Li and Mobley, 1998). In mrpB mutants, there were significantly fewer pili on the cell surface, however, the pili were up to six times longer than the corresponding wild type strain (Li and Mobley, 1998). FimG from type I pili has been demonstrated to be a minor pilus component that regulates pilus length. fimG mutants produce pili three times longer than the parent strain, but have the same number of pili per cell with normal type I morphology and retain receptor binding (Klemm and Christiansen, 1987;
Maurer and Orndorff, 1987; Russell and Orndorff, 1992). Conversely, CS1 pili lacks any proteins involved in length regulation. It has been hypothesised that the relative affinity of the major subunit CooA for the minor subunit CooD versus its affinity for CooA and the concentrations of these proteins will determine the length, since CooD is expressed at very low levels compared to CooA (Sakellaris et al., 1996).

This chapter describes the characterisation of the fourth gene constituting the CS5 pilus encoding region, csfE, and specifically shows the requirement of CsfE for the termination of assembly of CS5 pili.

5.2 Results

5.2.1 Construction of an In-Frame Deletion in the csfE Gene

A single NruI site located within the csfE gene across nt 5597-5602 of the characterised sequence (Appendix I) was utilised for exonuclease III mediated deletions, since NruI digestion produces blunt end fragments which are susceptible to exonuclease III. Nested deletions were carried out according to section 2.9.8.3. In short, 10 µg of pPM5631 was digested with NruI, and the DNA fragment treated with exonuclease III that deletes from both ends of the cut plasmid, for approximately 60 sec at 37°C, with samples taken every 15 sec. The DNA was then treated with Klenow enzyme and S1 nuclease, ligated with T4 DNA ligase and transformed into E. coli K-12 strain DH5α. Transformants were selected which retained at least part of the csfE gene by utilising the specific csfE oligonucleotides #2973 and #2974 (Table 2.3) in a PCR reaction on boiled lysates from the transformants (section 2.9.8.2). Chosen csfE deletions were sequenced to ensure the deletion retained the correct reading frame, using oligonucleotides #2561, #2542 and #2974 (Table 2.3). One such deletion, which juxtaposed nt 5428 and 5806 together, thereby deleting 378 nt of the csfE gene and retaining the correct reading frame, was selected for further use and termed pPM5654 (Figure 5.1).
Figure 5.1 Construction of an in-frame csfE deletion mutant.

Plasmid pPM5631 was digested with NruI, which is located across nucleotides 5597-5602 within the csfE gene (orientated according to the arrow) and subjected to bi-directional deletions using exonuclease III. The DNA was then treated with Klenow enzyme, S1 nuclease and T4 DNA ligase to give pPM5654, which contains an in-frame deletion of the csfE gene, by the juxtapositioning of nucleotides 5428 and 5806.
Cleave with NruI, exonuclease III deletion

Ligation

modified csf region (ΔcsfE)
5.2.2 Construction of a csfE Complementing Plasmid

The wild type csfE gene was PCR amplified from pPM5631 using oligonucleotides #2973 and #2974, which contain PsI and XbaI sites respectively (Table 2.3). The fragment was then digested with the appropriate restriction enzymes and cloned into PsI and XbaI cut pWKS130 (Wang and Kushner, 1991) to give pPM5657, with the csfE gene orientated behind the inducible lac promoter (Figure 5.2). The plasmid was used in trans to complement the introduced in-frame deletion of csfE.

5.2.3 Effect of a Deletion in csfE on CS5 Pilus Biosynthesis

The deletion in the csfE gene was assessed for any phenotypic effects on CS5 pilus biogenesis by examining heat extracts, periplasms, whole cells and supernatants for any alterations from wild type pili expression by western immunoblots, along with immunogold electron microscopy, and haemagglutination experiments. The following sections describe the role of CsfE in CS5 pilus biogenesis.

5.2.3.1 Examination of Pilus Morphology and Function

The constructed csfE mutant was initially assessed for any phenotypic effects on CS5 pilus assembly by slide agglutinations (section 2.18). E. coli K-12 harbouring the csfE deletion plasmid (pPM5654) did not agglutinate with the rabbit anti-CS5 pili antiserum, whereas strong agglutination was observed in E. coli K-12 harbouring pPM5631 or pPM5654 complemented with pPM5657. This indicates that the introduced csfE mutant may alter assembly of CS5 pili.

Immunogold electron microscopy was then performed on the strains using rabbit anti-CS5 pili antiserum, as described in section 2.20. E. coli K-12 harbouring pPM5631 or pPM5654 complemented with pPM5657 produced peritrichous CS5 pili, while the negative control strain E. coli K-12 harbouring pGEM-7Zf+ was devoid of pili (Figure 5.3, panel A-C). E. coli K-12 harbouring pPM5654 did produce CS5 pili on the cell surface, but in far less
Figure 5.2 Construction of csfE in the low copy number vector pWKS130.

The csfE gene was PCR amplified using oligonucleotides #2973 and #2974 which contain PstI and XbaI restriction sites respectively, from the template pPM5631. The amplified csfE fragment, which is orientated according to the direction of the arrow, was then digested with the appropriate restriction enzymes and ligated into PstI and XbaI cut pWKS130 (Wang and Kushner, 1991) to generate pPM5657, which contains the csfE gene behind the lac promoter.
PCR with oligonucleotides 
#2973, #2974

Cleave with 
PstI and XbaI

Cleave with 
PstI and XbaI

Ligation

pPM5657
Figure 5.3 Immunogold electron microscopy analysis.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pGEM-7Zf*, B. pFM5654 + pFM5657, C. pFM5631, or D-E pFM5654. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
numbers, while the morphology of the bundled pili suggested they were much longer than the corresponding CS5 pili in *E. coli* K-12 with pPM5631 or the complemented *csfE* mutant (Figure 5.3, panel D-E). This data shows that CsfE may function in the control of the length of CS5 pili, thereby acting as a terminator of CS5 pili assembly. The apparent drastic reduction in the number of cell surface associated of CS5 pili may explain why *E. coli* K-12 with pPM5654 did not readily agglutinate with the rabbit anti-CS5 pili antiserum.

The *csfE* mutant strain was also tested for the ability to cause MRHA of human group A⁺ erythrocytes (section 2.19). Despite the ability of *E. coli* K-12 with pPM5654 to assemble CS5 pili on the cell surface, albeit in fewer quantities and much longer than the wild type, no haemagglutination ability was conferred (Figure 5.4). The ability to haemagglutinate was restored when the *csfE* mutant was complemented with a wild type copy of the *csfE* gene on pPM5657 (Figure 5.4). The longer CS5 pili produced by the *csfE* mutant strain are likely to be more susceptible to mechanical shearing caused by resuspending the bacteria in 1x PBS, which may explain why this strain was unable to mediate haemagglutination.

Western immunoblot analysis following SDS-15% PAGE using rabbit anti-CS5 pili antiserum was also performed on both heat extracts (section 2.17.1) and cell supernatants prepared from *E. coli* K-12 harbouring either pGEM-7Zf⁺, pPM5631, pPM5654 or pPM5654 complemented with pPM5657. A 40 μL of sample was mixed with an equal volume of 2x SDS-sample buffer prior to SDS-15% PAGE. *E. coli* K-12 harbouring pPM5654 showed a reduction in the accumulation of CsfA and CsfD pilin subunits in both heat extracts and cell supernatants, when compared to *E. coli* K-12 harbouring pPM5631 (Figure 5.5). This reduction is examined further by preparing whole cell and periplasmic extracts to ascertain whether the accumulation of CsfA and CsfD pilin subunits is reduced in a *csfE* deletion strain.
Figure 5.4 Requirement of CsfE for CS5 pilus mediated haemagglutination.

Samples of E. coli K-12 containing pPM5654 or pPM5654 + pPM5657 were prepared by gently resuspending the bacteria to an OD_{600nm} = 4.0 in 1x PBS. 10 μl was mixed with an equal volume of human group A+ erythrocytes in the presence of 0.1M D(+)-mannose and gently rocked (section 2.19). Clear haemagglutination of the erythrocytes, which appear as darkened granules, only occurs when the csfE mutation is restored (pPM5654 + pPM5657).
Figure 5.5 Effect of a csfE deletion on assembly and release of CS5 pili in *E. coli* K-12.

Heat extracts and cell supernatants were prepared from the various strains according to section 2.17.1. Western immunoblot analysis was performed using rabbit anti-native CS5 pili antiserum (1:5000) on either heat extracts (A) or cell supernatants (B) of *E. coli* K-12 harbouring pGEM-7Zf+, pPM5631, pPM5654, and pPM5654 + pPM5657. Samples represent $1 \times 10^9$ cells. The dark arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
5.2.3.2 Periplasmic and Whole Cell Analysis

Periplasmic extracts were prepared from the various strains, to assess whether a mutation in csfE effects the delivery of major and minor subunits into the periplasm, but also the stability of the subunits within the periplasm.

Periplasmic extracts were prepared from E. coli K-12 strains by resuspending 2x $10^{10}$ cells in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1. Cells were converted to sphaeroplasts with 100 µg/ml of lysozyme in 0.1 M EDTA, pH 7.3 and the resulting supernatant collected after centrifugation was the periplasmic fraction (section 2.14). Whole cell samples were prepared by adjusting the harvested bacteria to 1x $10^{10}$ cells per 500 µl. For both periplasmic and whole cell extracts, 40 µl of sample was mixed with an equal volume of 2x SDS-sample buffer and analysed by western immunoblots following SDS-15% PAGE.

E. coli K-12 harbouring pPM5654 did not result in any change in the accumulation of both CsfA and CsfD in the periplasm when compared to E. coli K-12 harbouring pPM5631 or pPM5654 complemented with pPM5657 (Figure 5.6). This suggests that CsfE is not required for stability of CsfA or CsfD in the periplasm of E. coli K-12. There was an increased amount of CsfA and CsfD detected in whole cells of E. coli K-12 with pPM5654 compared to E. coli K-12 harbouring pPM5631 or pPM5655 complemented with pPM5657. This is likely a reflection of the overall hydrophobicity of CS5 pili, where detached longer CS5 pili reassociate with cell surface pili and are centrifuged with the whole cells. Examination of electron micrographs of the csfE mutant strain shows a mesh-like network of associated pili on these strains (Figure 5.3, panel D-E), which provides some evidence for long cell-free pili re-associating with cell surface pili attached to the bacteria.

5.2.4 Specific Modulation of CsfE with P\textsubscript{BAD} Promoters

This next section examines in more detail the outcome of modulating CsfE levels on CS5 pili numbers and lengths, by western immunoblot analysis and making specific counts and measurements of the pili under the electron microscope.
Figure 5.6 **Effect of CsfE on the stability of CsfA major subunits in the periplasm.**

Western immunoblot analysis of A. Whole cell samples and B. Concentrated periplasmic extracts from *E. coli* K-12 harbouring pGEM-7Zf*, pPM5631, pPM5654 and pPM5654 + pPM5657, using anti-native CS5 pili antiserum (1:5000). The dark arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Sample represents $2 \times 10^9$ cells (periplasm) and $1 \times 10^9$ cells (whole cells). Molecular mass standards are indicated.
5.2.4.1 Construction of csfE in pBAD18-Kan

The csfE gene was excised from pPM5633 (Figure 3.15) by digestion with KpnI and XbaI and purified following agarose gel electrophoresis from the remaining plasmid. The fragment was then cloned directly into the multiple cloning site of pBAD18-Kan (Guzman et al., 1995) such that the csfE gene was under control of the P_BAD promoter from the araBAD (arabinose) operon. This plasmid was termed pPM5684 (Figure 5.7) and used to directly complement the csfE mutant (pPM5654) in E. coli K-12.

5.2.4.2 The Effect of Specific Modulation of CsfE on CS5 Pilus Biogenesis

The expression of CsfE from pPM5684 in E. coli K-12 containing pPM5654 was accomplished by growing the bacteria on CFA agar containing the appropriate antibiotics along with 0.002%, 0.01% or 0.05% arabinose. In the presence of arabinose, transcription from the P_BAD promoter is turned on, allowing expression of the CsfE protein. Western immunoblots using rabbit anti-CS5 antiserum were performed on both heat extracts and cell supernatants (section 2.17.1), prepared under the different concentrations of arabinose, along with E. coli K-12 with pPM5631 as a comparison.

For the different concentrations of arabinose used, CsfA and CsfD could be detected in heat extracts from E. coli K-12 harbouring pPM5654 complemented with pPM5684; however, the relative amounts of each was below that observed for E. coli K-12 with pPM5631 (Figure 5.8A). When cell supernatants were examined from each of the strains grown under the different concentrations of arabinose, the amount of CsfA was also reduced, while CsfD was not detected when compared to E. coli K-12 with pPM5631 (Figure 5.8B). This is likely to be the result of a reduction in plasmid copy number. pPM5631 has a CoIE1 origin of replication, while pPM5684 contains a pBR322 origin that is incompatible with pPM5631; but is retained given the difference in antibiotic resistances. This incompatibility would be expected to lead to an overall reduction in copy number, thereby reducing expression of the pili, as observed in
Figure 5.7 Construction of pPM5684 containing $csfE$ from the $P_{BAD}$ promoter.

The $csfE$ gene was excised from pPM5633 as a $KpnI-XbaI$ fragment and cloned directly into $KpnI$ and $XbaI$ cut pBAD18-Kan (Guzman et al., 1995) to create pPM5684, which contains the $csfE$ gene under the influence of the $P_{BAD}$ promoter derived from the $araBAD$ (arabinose) operon.
Cleave with *KpnI, XbaI*  

Ligation  

Cleave with *KpnI, XbaI*
Figure 5.8 Western immunoblot analysis of CS5 pili expression by modulating CsfE.

Heat extracts (A) and cell supernatants (B) were prepared from either *E. coli* K-12 harbouring pPM5631 or *E. coli* K-12 harbouring pPM5654 + pPM5657, which was grown in the presence of 0.002%, 0.01% or 0.05% arabinose. Western immunoblot analysis was performed using rabbit anti-native CS5 pili antiserum (1:5000). Samples represent 1x $10^9$ cells. The dark arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
western immunoblots of both heat extract and cell supernatant samples. However, since a significant amount of surface pili is still produced in strains harbouring both plasmids, by immunogold electron microscopy analysis (see Figure 5.9, panel C), the effect of modulating CsfE levels on CS5 pilus morphology can be readily determined.

5.2.4.3 Immunogold Electron Microscopy and Quantitative Analysis of Pili Numbers and Lengths

As a more quantitative measure of the effect of modulating CsfE levels, the bacteria were examined under the electron microscope following immunogold labelling and negative staining with 1% uranyl acetate, with pili numbers and lengths recorded. The mean number of surface located pili was calculated on fifty randomly selected bacteria ± SEM, while the mean length of pili was calculated by recording the length of five pili on ten randomly selected bacteria (n=50) ± SEM. Since CS5 pili bundle, the mean number and length is representative of discrete bundles of pili. Also included was E. coli K-12 with pPM5654 and pPM5657. This strain contains a wild type csfE gene on the low copy number plasmid pWKS130 (pPM5657)(Figure 5.2), that is under the influence of the lac promoter, and has been previously used to complement the introduced csfE deletion on pPM5654 (section 5.2.3.1). Specific induction of the csfE gene from pPM5657 was carried out by the addition of 40 µg/ml IPTG to the CFA agar.

Immunogold analysis was performed according to section 2.20 using rabbit anti-CS5 pili antiserum (1:10) and protein A-20 nm gold conjugate (1:40), followed by negative staining with 1% uranyl acetate. E. coli K-12 expressing pPM5631 had a mean number of 41.34 ± 1.349 pili per cell and length of 0.55 ± 0.024 µm (Figure 5.9, panel B; Table 5.1). However, in the csfE mutant strain, E. coli K-12 with pPM5654, the pili were much longer (mean = 1.859 ± 0.069 µm) and in far less abundance (mean = 5.22 ± 0.475) than the wild type (P< 0.0001)(Figure 5.9, panel A; Table 5.1). When the csfE mutant was complemented in trans with pPM5657 in the presence or absence of IPTG, the mean number of pili per cell
**Table 5.1 Effect of CsfE on CS5 pilus length.**

<table>
<thead>
<tr>
<th>E. coli K-12 containing</th>
<th># pili per cell&lt;sup&gt;a&lt;/sup&gt;</th>
<th>length of pili (µm)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPM5631</td>
<td>41.34 ± 1.349</td>
<td>0.55 ± 0.024&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5631 + pBAD18-Kan + 0.01% arabinose</td>
<td>11.82 ± 1.09&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N.D</td>
</tr>
<tr>
<td>pPM5631 + pPM5684 + 0.01% arabinose</td>
<td>18.46 ± 1.128&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.539 ± 0.018</td>
</tr>
<tr>
<td>pPM5654</td>
<td>5.22 ± 0.475&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.859 ± 0.069&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5657 - IPTG</td>
<td>19.26 ± 0.834&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.775 ± 0.042&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5657 + IPTG</td>
<td>22.72 ± 1.085&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.771 ± 0.045&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5684 + 0.002% arabinose</td>
<td>10.62 ± 0.603&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.036 ± 0.079&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5684 + 0.01% arabinose</td>
<td>9.88 ± 0.579&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.721 ± 0.035&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5684 + 0.05% arabinose</td>
<td>6.36 ± 0.447&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.735 ± 0.038&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mean number of pili on 50 randomly selected bacteria ± SEM.

<sup>b</sup> Mean length of 5 pili on 10 randomly selected bacteria (n=50) ± SEM.

<sup>c</sup> P< 0.0001 compared to wild type (pPM5631) by using the unpaired 2 tailed t test.

<sup>d</sup> P< 0.0001 compared to wild type (pPM5631) by using the unpaired 2 tailed t test.

<sup>e</sup> P= 0.002 compared to wild type (pPM5631) by using the unpaired 2 tailed t test.

_N.D_ Not Determined.
Figure 5.9 Visual CS5 pili production following CsfE expression by immunogold electron microscopy.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pPM5654, B. pPM5631, C. pPM5631 + pBAD18-Kan, D. pPM5631 + pPM5684, E. pPM5654 + pPM5657 (- IPTG), F. pPM5654 + pPM5657 (+ IPTG), G. pPM5654 + pPM5684 (0.002% arabinose), H. pPM5654 + pPM5684 (0.01% arabinose), I. pPM5654 + pPM5684 (0.05% arabinose). The different bacterial strains were first reacted with rabbit antiserum raised against native CS5 pili (1:10), then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
increased (22.72 ± 1.085 and 19.26 ± 0.834 respectively), while the mean length of the pili decreased (0.771 ± 0.045 μm and 0.775 ± 0.042 μm respectively), if compared to E. coli K-12 harbouring pPM5654 (Figure 5.9, panel E-F; Table 5.1). However, both the mean number and length of the pili were not restored to that observed in the wild type strain (P< 0.0001), which may be a consequence of csfE expressed from a low copy number plasmid versus the high copy number of pPM5631. The addition of IPTG to the growth media of E. coli K-12 with pPM5654 and pPM5657 did not result in significant differences between the mean number and length of CS5 pili (Table 5.1).

Expression of CsF from the PBad promoter on pPM5684 was carried out by the addition of either 0.002%, 0.01% or 0.05% arabinose, which specifically induces CsF expression. E. coli K-12 harbouring pPM5654 and pPM5684 was grown on the different concentrations of arabinose, along with E. coli K-12 with pPM5631 and pPM5684. E. coli K-12 with pPM5631 and pBAD18-Kan served as a control to show that co-expression of pBAD18-Kan with the ColE1 based plasmids pPM5631 or pPM5654 reduces expression of CS5 pili due to the incompatibility problem of the plasmids, as observed in western immunoblots in Figure 5.8. Clearly, co-expression of pBAD18-Kan and pPM5631 in E. coli K-12 leads to a decrease in cell surface CS5 pili numbers when compared to E. coli K-12 with pPM5631 alone (Figure 5.9, panel C; Table 5.1).

E. coli K-12 harbouring pPM5654 and pPM5684 in the presence of 0.002% arabinose, the mean length of the pili was 1.036 μm which decreased further when the concentration of arabinose was increased to 0.01% (0.721 ± 0.035 μm) or 0.05% (0.735 ± 0.038 μm)(Figure 5.9, panel G-I; Table 5.1). The length of the pili produced were still significantly longer than that observed in the wild type (P< 0.0001). When pPM5684 was expressed with pPM5631 in the presence of 0.01% arabinose, the mean pili length was 0.539 ± 0.018 μm, which was not significantly different from pPM5631 alone (P= 0.68)(Figure 5.9, panel D; Table 5.1). The data suggests that pilus length may not be determined solely by CsF, since the specific
induction of CsfE could not reduce pilus length below that of the wild type. However, the data does show that CsfE is specifically involved in length regulation.

5.2.5 Over-Expression of CsfE

The CsfE protein was over-expressed from pPM5633, which contains the csfE gene behind the T7 promoter (Figure 3.15), to generate polyclonal antiserum to be used in fractionation analysis. pPM5633 was transformed into E. coli strain E2096 that contains the T7 RNA polymerase gene on pGP1-2 (Tabor and Richardson, 1985), with specific expression carried out as described in section 2.12.3.1. Bacterial cells were then fractionated into the following: whole cells (w/c), inclusion bodies (i/b), periplasms (peri), cytoplasms (cyto) and whole membranes (w/m) as described in section 2.14. Over-expressed CsfE was detected in the whole cell fraction and found to localise exclusively in the inclusion body fraction (Figure 5.10). This observation was utilised as a means to purify the CsfE protein from this fraction for polyclonal antiserum preparation.

5.2.6 Generation of Specific Antibodies Against CsfE and Cellular Location

Specific polyclonal antiserum against CsfE was raised in New Zealand white rabbits following an immunisation protocol essentially as described in section 2.16.1. Inclusion body extracts were prepared from a 100 ml culture of E. coli over-expressing CsfE from the T7 promoter (section 2.12.3.1), since over-expressed CsfE accumulates in this cellular fraction (Figure 5.10). A sample of the inclusion body extract in 2x SDS-sample buffer was run on SDS-15% PAGE, followed by Coomassie Brilliant Blue staining. A protein band corresponding to CsfE was then carefully excised and re-run on a second SDS-15% PAGE. Again the protein band corresponding to CsfE was excised following Coomassie Brilliant Blue staining, but this time homogenised in 500 μl of 1x PBS with an equal volume of Freund’s complete adjuvant for primary immunisation. These steps were repeated for subsequent immunisations, with Freund’s incomplete adjuvant used. The antiserum obtained
Figure 5.10 Over-expression and fractionation of CsfE from pPM5633.

Coomassie Brilliant Blue stained SDS-15% PAGE of *E. coli* K-12 harbouring pBS-SK⁺ or pPM5633. The 23 kDa CsfE protein is indicated by an asterix (*). Over-expression of CsfE from pPM5633 by temperature shift (30-42°C) to induce the λp_L promoter is as described in section 2.12.3.1. The bacteria were fractionated into whole cell (w/c); inclusion body (i/b); periplasm (peri); cytoplasm (cyto) and whole membrane (w/m) fractions as described in section 2.14. Molecular mass standards are indicated.
was absorbed extensively against whole cells of *E. coli* K-12 harbouring pBS-SK*+* and neutralised across inclusion body extracts of the same strain on nitrocellulose membranes (section 2.16.2 and 2.16.3). The specificity of the antiserum for CsfE was tested in western immunoblot analysis, and was found to recognise CsfE (data not shown).

The antiserum generated was used to localise CsfE in *E. coli* K-12 by fractionation experiments. *E. coli* K-12 harbouring pPM5631 or pGEM-7Zf*+* was fractionated into the following: whole cells (w/c), inclusion bodies (i/b), cytoplasms (cyto), periplasms (peri), whole membranes (w/m), outer membranes (o/m) and inner membranes (i/m), essentially as described in section 2.14. Western immunoblot analysis on the fractions using the rabbit anti-CsfE antiserum showed that CsfE localised in the outer membrane fraction (Figure 5.11A). There was no contamination between the outer and inner membranes since the outer membrane porins of *E. coli* (OmpC/OmpF) localised exclusively in the outer membrane fraction as judged by Coomassie Brilliant Blue staining (Figure 5.11B). Clearly, CsfE is expressed at low levels in the cell.

### 5.2.7 The Association of CsfE with CS5 Pili

Since CsfE has a specific role in controlling the length of CS5 pili, it was determined whether CsfE is assembled as a minor component of CS5 pili thereby terminating assembly, or if the association of CsfE in the outer membrane, presumably with the hypothesised outer membrane assembly protein CsfC, terminates assembly. Fractionation of *E. coli* K-12 bearing pPM5631 showed that CsfE was detected in the outer membranes (Figure 5.11), which indicates that CsfE is targeted either into and/or across the outer membrane.

To examine the association of CsfE with CS5 pili, caesium chloride purified CS5 pili were prepared according to section 2.17.3. Different concentrations of denatured purified pili prepared in 2x SDS-sample buffer ranging from 10-100 μg were loaded onto a SDS-15% PAGE including a positive control for the CsfE protein, namely inclusion body extracts from over-expression of CsfE from pPM5633 (Figure 5.10). Western immunoblot analysis using
Figure 5.11 Localisation of CsfE in *E. coli* K-12 harbouring pPM5631.

A. Western blot analysis was carried out on whole cell fractionations (equivalent to $-1 \times 10^8$ cells) derived from *E. coli* K-12 harbouring pPM5631 or pGEM-7ZT+ to determine the cellular location of CsfE using rabbit anti-native CS5 pili antiserum. Bacteria were fractionated according to section 2.14 into whole cell (w/c); inclusion body (i/b); supernatant (s/n); cytoplasm (cyto); periplasm (peri); whole membrane (w/m); outer membrane (o/m) and inner membrane (i/m) fractions. B. Coomassie Brilliant Blue stained inner and outer membrane fractions of pGEM-7ZT+ and pPM5631 indicating the position of the two major *E. coli* porins as a control for separation of these fractions. Molecular mass markers are indicated.
both rabbit anti-CsfE and rabbit anti-CS5 pili antisera showed that CsfE could not be detected in denatured CS5 pili up to 100 µg, whereas the major and minor pilin subunits CsfA and CsfD respectively, are readily detected (Figure 5.12A and 5.12B). No CsfE was detected in heat extracts representing $5 \times 10^9$ cells (Figure 5.12C), which has been shown to release over 100 µg of CS5 pili by Coomassie Brilliant Blue gel analysis (data not shown). CsfE was retained in the outer membrane from the same numbers of cells after heat extraction (Figure 5.12C).

**5.3 Summary and Conclusions**

This chapter examined the potential role of CsfE in CS5 pili biogenesis. As previously described in chapter 3, the csfE gene sequence has no recognised homology to any relevant protein sequences lodged into the database when using advanced BLAST at EMBL (P > 0.21). An in-frame deletion constructed in the csfE gene from pPM5631 did not significantly effect the amount of surface exposed pilus antigen detectable in heat extracts, thereby indicating CsfE is not required in the secretion or the assembly processes of the pilin subunits CsfA and CsfD. However, the csfE deletion did result in pili which were over three times longer when compared to the wild type strain, which showed that CsfE played a specific role in controlling the length of assembled CS5 pili. In support of this, when a wild type copy of the csfE gene was re-introduced into the csfE mutant, the pili were restored to a length which was longer than that observed in the wild type strain, but significantly shorter than in the csfE mutant.

Specific induction of CsE expression from the $P_{BAD}$ promoter in the csfE mutant could not restore pilus length to the wild type (0.55 µm), which suggests there are other factors controlling the length of CS5 pili. This was verified further when the same plasmid was introduced into the wild type strain and pilus length remained unchanged. For other analogous proteins such as PapH and MrpB, specific over-expression of these proteins resulted in a significant reduction in pilus length compared to the wild type (Båga et al., 1987; Li and Mobley, 1998). Therefore, the CsfE protein is not rate-limiting for CS5 pilus length. Chapter
Figure 5.12 Association of CsfE with CS5 pili by western immunoblot.

Western immunoblot analysis was performed on 10-100 µg of purified CS5 pili, along with a sample of an inclusion body fraction from pPM5633 over-expressing CsfE (section 5.2.4) using (A) anti-native CS5 pili antiserum (1:5000) or (B) anti-CsfE antiserum (1:1000). The dark arrow indicates CsfA protein, empty arrow CsfD (A), while the position of CsfE is also indicated by a dark arrow (B). C. Western immunoblot using rabbit anti-CsfE antiserum (1:1000) of both heat extract and outer membrane samples of E. coli K-12 with pPM5631 or pGEM-7Zf+. Samples represent 5x 10^9 cells, with CsfE indicated by an arrow. Molecular mass standards are indicated.
7 identifies a second protein, CsfF, which is also responsible for controlling pilus length, and is specifically shown to be rate-limiting in determining pilus length.

The csfE deletion also resulted in a drastic reduction in cell surface located pili when compared to the wild type. This implies that CsfE may also function to anchor the pilus to the cell surface, not unlike PapH. In support of this, in a csfE deletion, the relative quantity of CS5 pili in the supernatant sample was similar to the wild type, and a significant increase in pilin subunit accumulation was detected in the cell pellet fraction compared to the wild type. The increase in the cell pellet fraction is attributable to the cell-free longer pili re-associating with surface pili and centrifuging with whole bacteria, and is likely due to the hydrophobic nature of the pili. Alternatively, the increase observed in the csfE mutant is due to the increased shearing forces on the longer pili, resulting in an increase in cell-free pili able to re-associate with surface pili.

Over-expression and subsequent gel purification of CsfE was used as a means to generate specific polyclonal antiserum. The antiserum was found to localise CsfE in the outer membrane of E. coli K-12 harbouring pPM5631, which expresses cell surface CS5 pili. Moreover, CsfE was undetectable in crude heat extracts corresponding to over 100 μg of CS5 pili and was found to be retained in the outer membrane fraction of the heat extracted cells. This implies that (1) CsfE is not polymerised into the pilus fibre, or (2) the association is relatively weak and is readily disrupted by heat extraction. Furthermore, when up to 100 μg of purified CS5 pili was examined using the anti-CsfE antiserum, no CsfE was detected. More likely is that the incorporation of CsfE into the pilus fibre through the predicted outer membrane assembly protein CsfC (section 3.2.4.1) results in a stable binding complex with CsfC. This prohibits further pilins from associating with CsfC in the outer membrane and is therefore responsible for termination of pilus assembly.
The Role of CsfC in the Assembly of CS5 Pili

6.1 Introduction

In Gram negative bacteria, the assembly of individual pilin subunits into pilus structures on the cell surface requires specialised outer membrane assembly proteins, which are responsible for translocating the subunits from the periplasm across the outer membrane, in concert with the specific periplasmic chaperone proteins. The assembly of >27 distinct adhesive organelles from Gram negative bacteria requires both periplasmic and outer membrane assembly proteins (Hung et al., 1996). In the case of Pap pili, the term “usher” has been used to describe the outer membrane assembly protein PapC of this system, since PapC is able to distinguish between different pilin subunits to direct the ordered assembly of Pap pili, rather like a human usher at the cinema whose job requires a distinction between different “ticket holders” (Dodson et al., 1993). It is thought that other characterised outer membrane assembly proteins including CooC from CS1 pili (Sakellaris et al., 1996), CfaC from CFA/I pili (Jordi et al., 1992b), FimD from type I pili (Klemm and Christiansen, 1990), CstB from CS3 pili (Jalajakumari et al., 1989), FaeD from K88 pili (Mooi et al., 1986) and FanD from K99 pili (Mooi et al., 1986) act by a similar mechanism.

The topology of the FaeD usher from K88 pili indicates this protein to contain a large central domain containing 24 membrane-spanning segments and two large periplasmic regions at the N-terminal and C-terminal ends of the protein respectively (Valent et al., 1995). Each of the periplasmic segments contained two cysteine residues which are strongly conserved in other pili systems and may play a role in stabilising a specific conformation which is required for specific interaction with chaperone-subunit complexes prior to assembly across the outer
membrane (Valent et al., 1995). Such topological studies may provide a general structural model representative of ushers from other systems.

More detailed structural analyses on the PapC usher showed that this protein exists as a ring-shaped oligomeric complex composed of six subunits which forms a channel approximately 2 nm in diameter, through which pilin subunits are assembled (Thanassi et al., 1998b). However, PapC has a low pore-forming ability when compared to OmpF, even though the diameter of the pore is larger. This is highly suggestive of a gating mechanism, which specifically opens and closes the pore (Thanassi et al., 1998b). The translocation of subunits across the membrane occurs independently of cellular energy and is thermodynamically driven (Jacob-Dubuisson et al., 1994b).

The elucidation of how outer membrane ushers may work to translocate pilin subunits across the outer membrane came from studies on FimD from type I pili (Saulino et al., 1998). It was shown that the kinetic partitioning of chaperone-adhesin complexes to the usher determines the localisation of the adhesin in the pilus, and that the conformation of the usher was altered upon this binding, which rendered the usher in an assembly-competent state during pilus assembly (Saulino et al., 1998). The same system occurs in Pap pili biogenesis with the PapC usher rendered in an assembly-competent state upon binding of the PapG-PapD adhesion-chaperone complex.

This chapter describes the characterisation of the third and largest gene comprising the csf cluster, csfC which encodes an outer membrane assembly protein for CS5 pili.

6.2 Results

6.2.1 Construction of an In-Frame Deletion in the csfC Gene

A single XhoI site located within the csfC gene across nt 4227-4232 of the characterised sequence (Appendix I) was utilised for exonuclease III mediated deletions, since XhoI digestion produces 5’ overhangs which are susceptible to exonuclease III. Nested deletions were carried out according to section 2.9.8.3. In short, 10 μg of pPM5631 was
digested with XhoI, and the DNA fragment treated with exonuclease III which deletes bidirectionally from the 5’ overhangs on the cut plasmid, with samples taken every 15 sec for 2 min at 37°C. The reaction was stopped and the DNA was then treated with Klenow enzyme and S1 nuclease, ligated with T4 DNA ligase and transformed into E. coli K-12 strain DH5α. Transformants were selected which retained at least part of the csfC gene by utilising the specific csfC oligonucleotides #2977 and #2978 (Table 2.3) in a PCR reaction on boiled lysates from the transformants (section 2.9.8.2). Chosen csfE deletions were sequenced to ensure the deletion retained the correct reading frame, using oligonucleotides #2561, #2542 and #2974 (Table 2.3). One such deletion, which juxtaposed at 3643 and 5068 together, thereby deleting 1425 nt of the csfC gene and retaining the correct reading frame, was selected for further use and termed pPM5651 (Figure 6.1).

6.2.2 Construction of a csfC Complementing Plasmid

The wild type csfC gene was PCR amplified from pPM5631 using oligonucleotides #2977 and #2978, which contain PstI and XbaI sites respectively (Table 2.3). The fragment was then digested with the appropriate restriction enzymes and cloned into PstI and XbaI cut pWKS130 (Wang and Kushner, 1991) to give pPM5638, with the csfC gene orientated behind the inducible lac promoter (Figure 6.2). The plasmid was used in trans to complement the introduced in-frame deletion of csfC.

6.2.3 The Effect of a csfC Deletion on CS5 Pilus Biosynthesis

The csfC gene has previously been predicted to encode an outer membrane assembly protein, based on homology with CooC from CS1 pili (section 3.2.4.1). The deletion in the csfC gene was assessed for any phenotypic changes on CS5 pili biogenesis by examining heat extracts, periplasms, whole cells and supernatants for any alterations from the wild type in western immunoblots, along with immunogold electron microscopy, and haemagglutination experiments. The following sections describe the role of CsfC in CS5 pili biogenesis.
Figure 6.1 Construction of a *csfC* deletion mutant from pPM5631.

Plasmid pPM5631 was digested with *XhoI*, whose site is located across nt 4227-4232 within the *csfC* gene (orientated according to the arrow) and subjected to bi-directional deletions using exonuclease III. The DNA was then treated with Klenow enzyme, S1 nuclease and T4 DNA ligase to give pPM5651, which contains an in-frame deletion of the *csfC* gene, by the juxtapositioning of nt 3643 and 5068.
Cleave with XhoI, exonuclease III deletion

Ligation

modified csf region (ΔcsfC)
Figure 6.2 Construction of a complementing csfC plasmid.

The csfC gene was PCR amplified using oligonucleotides #2977 and #2978 (Table 2.3) which contain PstI and XbaI restriction sites respectively, from the template pPM5631. The amplified csfC gene fragment, which is orientated according to the direction of the arrow, was then digested with the appropriate restriction enzymes and ligated into PstI and XbaI cut pWKS130 (Wang and Kushner, 1991) to generate pPM5638 which contains the csfC gene behind the lac promoter.
PCR with oligonucleotides #2977, #2978

Ligation

Cleave with PstI and XbaI

Cleave with PstI and XbaI

Ligation

pPM5638

Kan^R

PstI

XbaI

lacZa

MCS

pWKS130

f1 origin

lacZa

MCS

Kan^R

pSC101 origin

PstI

XbaI

PstI

XbaI

csfC

csf region

pPM5631

Amp^R

SmaI

T2
6.2.3.1 Examination of CS5 Pilus Assembly

The constructed csfC mutant was initially assessed for any phenotypic effects on CS5 pili biogenesis by slide agglutinations (section 2.18). E. coli K-12 harbouring the csfC mutant plasmid (pPM5651) did not agglutinate with the rabbit anti-CS5 pili antiserum, whereas strong agglutination was observed in E. coli K-12 harbouring pPM5631 or pPM5651 complemented with pPM5638. This indicates that the introduced csfC mutant effects cell surface expression of CS5 pili.

Immunogold electron microscopy was then performed on the strains using the rabbit anti-CS5 pili antiserum, as described in section 2.20. E. coli K-12 harbouring pPM5631 and E. coli K-12 with pPM5651 and pPM5638 produced peritrichous CS5 pili, while E. coli K-12 containing pGEM-7Zf² was devoid of CS5 pili (Figure 6.3, panel A-B, D). E. coli K-12 harbouring pPM5651 also did not assemble cell surface CS5 pili (Figure 6.3, panel C).

The csfC mutant strain was also tested for the ability to cause MRHA of human group A* erythrocytes (section 2.19). As expected, no haemagglutination ability was conferred by E. coli K-12 with pPM5651, while haemagglutination was restored when the csfC mutant was complemented with pPM5638 (Figure 6.4).

Western immunoblot analysis following SDS-15% PAGE using rabbit anti-CS5 pili antiserum was also performed on both heat extracts and whole cell supernatants (section 2.17.1) prepared from the strains. 40 µl of sample was mixed with an equal volume of 2x SDS-sample buffer prior to SDS-15% PAGE. E. coli K-12 harbouring pPM5651 showed an abolition of CsfA and CsfD in heat extracts or whole cell supernatants when compared to E. coli K-12 harbouring pPM5631 and E. coli K-12 with pPM5651 and pPM5638 (Figure 6.5). This indicates that CsfC is essential in the release and therefore the subsequent assembly of CsfA and CsfD across the outer membrane into pili.
Figure 6.3 Immunogold electron microscopy analysis.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pGEM-7Zf*, B. pPM5631, C. pPM5651, or D. pPM5651 + pPM5638. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
Figure 6.4 Requirement of CsfC for CS5 pilus mediated haemagglutination.

Samples of *E. coli* K-12 containing pPM5651 or pPM5651 + pPM5638 were prepared by gently resuspending the bacteria to an OD$_{600nm}$ = 4.0 in 1x PBS. 10 μl was mixed with an equal volume of human group A$^+$ erythrocytes in the presence of 0.1 M D(+)-mannose and gently rocked (section 2.19). Clear haemagglutination of the erythrocytes, which appear as darkened granules, only occurs when the *csfC* mutation is restored (pPM5651 + pPM5638).
Figure 6.5 Western analysis on heat extracts and whole cell supernatants.

Heat extracts and whole cell supernatants were prepared from the various strains (section 2.17.1). The western immunoblot was performed using rabbit anti-native CS5 pili antiserum (1:5000) on cell supernatants (A) or heat extracts (B) of E. coli K-12 harbouring either pGEM-7zf+, pPM5631, pPM5651 or pPM5651 + pPM5638. Samples represent $1 \times 10^9$ cells. The solid arrow indicates position of Csfa protein and the empty arrow Csfd. Molecular mass standards are indicated.
6.2.3.2 Periplasmic Analysis

Periplasmic extracts were prepared from the various strains, to ensure that a mutation in \textit{csfC} does not effect the delivery of major and minor subunits into the periplasm, but also the stability of the subunits within the periplasm.

Periplasmic extracts were prepared from \textit{E. coli} K-12 strains by resuspending 2x $10^{10}$ cells in 1 ml of 20\% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1. Cells were converted to sphaeroplasts with 100 \(\mu\)g/ml of lysozyme in 0.1 M EDTA, pH 7.3 and the resulting supernatant collected after centrifugation was the periplasmic fraction (section 2.14). Whole cell samples were prepared by adjusting the harvested bacteria to 1x $10^{10}$ cells per 500 \(\mu\)l. For both periplasmic and whole cell extracts, 40 \(\mu\)l of sample was mixed with an equal volume of 2x SDS-sample buffer and analysed by western immunoblots following SDS-15\% PAGE (section 2.12.2).

\textit{E. coli} K-12 harbouring pPM5651 did not adversely alter the accumulation of both CsfA and CsfD in the periplasm when compared to both \textit{E. coli} K-12 containing pPM5631 or pPM5651 complemented with pPM5638 (Figure 6.6). This suggests that CsfC is not required for stability of CsfA or CsfD in the periplasm of \textit{E. coli} K-12.

6.2.4 Localisation of CsfC in \textit{E. coli} K-12

Attempts at producing significant quantities of the CsfC protein for antiserum generation either alone or genetically fused to the maltose binding protein (MBP) failed. Therefore, to determine the cellular location of CsfC in \textit{E. coli} K-12, the CsfC protein was labelled with L-[\textsuperscript{35}S]-methionine following expression from pPM5636, in which the \textit{csfC} gene is orientated from the T7 promoter (Figure 3.13). Specific expression was carried out according to section 2.12.3.2. The cells were then fractionated into the individual cell components: whole cells, inclusion bodies, cytoplasms, periplasms, outer membranes and inner membranes according to section 2.14. A sample of each was analysed by SDS-15\% PAGE and the gel stained with Coomassie Brilliant Blue and destained with 5\% acetic acid.
Figure 6.6 Whole cell and periplasmic analysis of the csfC deletion mutant.

Western immunoblot analysis of A. Whole cell samples and B. periplasmic extracts (section 2.14) prepared from *E. coli* K-12 harbouring pGEM-7Zf+ or pPM5631 or pPM5651 or pPM5651 + pPM5638, using anti-native CS5 pili antiserum. The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Each sample represents 2x 10^9 cells (periplasm) and 1x 10^9 cells (whole cells). Molecular mass standards are indicated.
The L-[\textsuperscript{35}S]-methionine labelled CsfC protein was detected by autoradiography of the dried gel (section 2.13). CsfC was detected in the outer membranes of the fractionated cells, and also within whole cell, cytoplasm and periplasm fractions (Figure 6.7A). No CsfC was detected in the inner membrane fraction. The two major outer membrane porins of \textit{E. coli} (OmpC/OmpF) served as controls for outer membrane and inner membrane fraction contamination in Coomassie Brilliant Blue staining, and were found to localise almost exclusively in the outer membrane fraction (Figure 6.7B). Numerous degradation and/or internal CsfC methionine-initiated proteins can also be detected in the various fractions (Figure 6.7A). These results show that CsfC resides in the outer membrane of \textit{E. coli} K-12 cells.

6.2.5 Complementation Experiments with Homologous Outer Membrane Assembly Proteins

The results so far are highly suggestive that CsfC acts as an outer membrane assembly protein for CS5 pili. Therefore, it was determined whether the homologous outer membrane proteins CooC from CS1 pili and CfaC from CFA/I pili could specifically complement a csfC mutant to produce cell surface CS5 pili. Both cooC and cfaC genes have been shown to encode outer membrane assembly proteins in the CS1 and CFA/I systems (Jordi \textit{et al.}, 1992b; Sakellaris \textit{et al.}, 1996).

The cooC gene was specifically PCR amplified from the wild type CS1 expressing ETEC strain C9216-2 (a gift from J. R. Scott) using oligonucleotides #3062 and #3063 which incorporates \textit{PstI} and \textit{BamHI} sites respectively (Table 2.3). The fragment was then digested with the appropriate enzymes and cloned into \textit{PstI-BamHI} cut pWKS130 orientated from the \textit{lac} promoter. Numerous attempts were made at isolating recombinants in \textit{E. coli} K-12 strain DH5\textalpha, without success. This is likely due to the toxic effects on the cell of expression of the CooC outer membrane protein from the \textit{lac} promoter, without specific repression. Therefore, the ligated plasmid was transformed into the \textit{E. coli} K-12 strain TOPP 10 F' (Table 2.1), which contains the \textit{lac} repression gene \textit{lacI} under constitutive expression (Stratagene). One
Figure 6.7 Localisation of L-\[^{35}\text{S}\]-methionine labelled CsF in *E. coli* K-12.

The CsF protein was labelled with L-\[^{35}\text{S}\]-methionine following expression from pPM5636, in which the *csf* gene is orientated from the T7 promoter. A negative control plasmid (pBS-SK\(^{+}\)) was also included. Specific expression was carried out according to section 2.12.3.2. The cells were then fractionated into the individual cell components: whole cells (w/c), inclusion bodies (i/b), cytoplasms (cyto), periplasms (peri), outer membranes (o/m) and inner membranes (i/m) according to section 2.14 and run on SDS-15% PAGE followed by Coomassie Brilliant Blue staining and destaining in 5% acetic acid. The L-\[^{35}\text{S}\]-methionine labelled CsF protein was detected by autoradiography of the dried gel (section 2.13) and is indicated by an arrow (A). (B) Coomassie Brilliant Blue stained inner and outer membrane fractions of pBS-SK\(^{+}\) and pPM5636 indicating the position of the two major *E. coli* porins as a control for separation of these fractions. Molecular mass markers are indicated.
transformant which contained the cloned \textit{cooC} gene was retained and the plasmid named pPM5681 (Figure 6.8). The \textit{cooC} gene was sequenced by utilising oligonucleotides derived from the published \textit{cooC} gene sequence (Accession no.76908) and was found to coincide 100\% with the published sequence (data not shown). The plasmid pPM5681 was then used to complement the \textit{csfC} deletion in pPM5651 using \textit{E. coli} K-12 strain DH5\alpha.

Similarly the \textit{cfaC} gene was PCR amplified from the CFA/I expressing strain PE213 (Willshaw \textit{et al.}, 1982) by the specific oligonucleotides #3064 and #3065, which contain \textit{PstI} and \textit{XbaI} sites respectively. The fragment was digested with the appropriate enzymes and cloned into \textit{PstI-XbaI} cut pWK8130 orientated from the \textit{lac} promoter to give pPM5661 (Figure 6.9). The \textit{cfaC} gene was sequenced using designed oligonucleotides based on the published \textit{cfaC} sequence (Accession no. M55661), and the sequence coincided 100\% with the published sequence (data not shown). The plasmid pPM5661 was used to complement the \textit{csfC} deletion in pPM5651.

\textit{E. coli} K-12 strains harbouring pPM5651, pPM5651 and pPM5681 (\textit{cooC}), pPM5651 and pPM5661 (\textit{cfaC}), or pPM5651 and pPM5638 (\textit{csfC}) were grown on CFA agar with the appropriate antibiotics and IPTG to induce expression from the \textit{lac} promoter and examined for cell surface CS5 pili expression by immunogold electron microscopy (section 2.20). CooC and CfaC were unable to restore cell surface CS5 pili expression in a \textit{csfC} deletion mutant (Figure 6.10, panel B-C), while complementation with pPM5638 restored the expression of CS5 pili in a \textit{csfC} mutant (Figure 6.10, panel D).

SDS-15\% PAGE was also performed on supernatant extracts (equivalent to 1x 10^9 cells) prepared from the same strains grown on CFA agar overnight at 37\degree C in the presence of the appropriate antibiotics and IPTG, along with \textit{E. coli} K-12 harbouring pPM5631 for comparison. Cell supernatant extracts showed that CooC and CfaC were unable to mediate the release of either Csfa or Csfd across the outer membrane and into the surrounding milieu, whereas the wild type CS5 expressing system of \textit{E. coli} K-12 with pPM5631 or pPM5651 complemented with pPM5638, allowed for the efficient release of both pilins (Figure 6.11).
Figure 6.8 Construction of the cooC complementing plasmid pPM5681.

The cooC gene was PCR amplified from the chromosome of boiled whole cell lysates of ETEC strain C9216-2 using the cooC specific oligonucleotides #3062 and #3063 (Table 2.3), which contain PstI and BamHI sites respectively. The cooC fragment (which is orientated according to the arrow) was digested with these enzymes and cloned into PstI-BamHI cut pWKS130 (Wang and Kushner, 1991) then transformed into the E. coli K-12 strain TOPP 10 F' to generate pPM5681 which contains the cooC gene under the control of the inducible lac promoter.
C9216-2 *rns·coo* + CS1 pili

- Clear with *PstI, BamHI*
- Ligation

**pWKS130**

- **pSC101 origin**
- **MC3**
- **Kan**

**pPM5681**

- **Kan**
- **BamHI**
- **Inc**
- **PstI**
Figure 6.9 Construction of the cfaC complementing plasmid pPM5661.

The cfaC gene was PCR amplified from the virulence plasmid containing the cfa gene cluster present in boiled whole cell lysates of ETEC strain PE213. The cfaC specific oligonucleotides #3064 and #3065 (Table 2.3) which contain PstI and XbaI sites respectively, were used to amplify the cfaC gene which was then digested with these enzymes prior to cloning into PstI and XbaI cut pWKS130 (Wang and Kushner, 1991) and transformation into E. coli K-12 strain DH5α. One transformant obtained with the cfaC gene correctly orientated from the inducible lac promoter was retained and designated pPM5661.
Cleave with PstI, XbaI

virulence plasmid
chromosomal DNA

Cleave with PstI, XbaI

Ligation

Cleave with PstI, XbaI

virulence plasmid
chromosomal DNA

Cleave with PstI, XbaI

Ligation

Cleave with PstI, XbaI

virulence plasmid
chromosomal DNA

Cleave with PstI, XbaI

Ligation

Cleave with PstI, XbaI
Figure 6.10 Immunogold electron microscopy analysis of CooC/CfaC complementations.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pPM5651, B. pPM5651 + pPM5681 (*cooC*), C. pPM5651 + pPM5661 (*cfaC*), or D. pPM5651 + pPM5638 (*csfC*). The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
Figure 6.11 Western immunoblot analysis of CooC/CfaC complementations.

Western immunoblots were performed on whole cell supernatants prepared from the various strains. The western immunoblot was performed using rabbit anti-native CS5 pili antiserum (1:5000) on *E. coli* K-12 harbouring either pPM5631, pPM5651, pPM5651 + pPM5638, pPM5651 + pPM5661 or pPM5651 + pPM5681. Samples represent 1x 10⁹ cells. The solid arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
6.3 Summary and Conclusions

This chapter examined the role of CsfC in CS5 pilus biogenesis. The CsfC protein was previously predicted to encode an outer membrane assembly protein, based on secondary structural features consistent with outer membrane proteins and significant protein homology with the CooC outer membrane assembly protein from CS1 pili (section 3.2.4.1). Initial experiments focused on creating an in-frame deletion in the csfC gene to determine the affect on CS5 pilus biogenesis. The in-frame deletion of 1425 nt was sufficient to inactivate the CsfC protein, since there was complete abolition of CS5 pili assembled on the cell surface in these bacterial strains, as judged by slide agglutinations and immunogold electron microscopy. As expected, the deletion did not result in any downstream polar effects on other csf genes, since a wild type copy of the csfC gene provided in trans was able to restore CS5 pili expression on the cell surface.

Western immunoblot analysis revealed the complete absence of both CsfA and CsfD pilin subunits in cell supernatants or in heat extracts, which indicates that CsfC may play a role in the assembly of pilin subunits across the outer membrane. CsfC was shown to play no role in stabilising pilin subunits in the periplasms of the strains, since no deleterious effects on the stability of CsfA or CsfD was observed in periplasmic extracts from the csfC deletion strain.

The CsfC protein was specifically localised to the outer membranes of E. coli K-12 strains when this protein was expressed from the T7 promoter and specifically labelled with L-[\textsuperscript{35}S]-methionine. The localisation of CsfC in this fraction, along with the abolition of pilus assembly in a csfC mutant indicates the requirement of CsfC in the transport of the CS5 pilin subunits across the outer membrane. Therefore, CsfC is an outer membrane assembly protein of CS5 pili.

The CsfC protein also contains a pair of N-terminal cysteine residues capable of forming an intramolecular disulphide bond, and two pairs of C-terminal cysteine residues
capable of forming two intramolecular disulphide bonds. The approximate location and number of cysteine residues in CsfC is a conserved feature of the N- and C-terminal periplasmic segments among other outer membrane assembly proteins (Valent et al., 1995). The formation of disulphide bonds are predicted to stabilise a specific conformation which is required for the association of the protein with chaperone-subunit complexes (Valent et al., 1995).

To examine the possibility of homologous outer membrane assembly proteins being able to functionally replace CsfC to assemble CS5 pili, CooC from CS1 pili and CfaC from CFA/I pili were chosen, based on their established homology to CsfC. In both cases, these proteins were unable to complement a csfC mutant to restore CS5 pili production on the cell surface. This indicates that the action of CsfC in translocating CS5 pilin subunits across the outer membrane is likely to be a specific one, based on the sequence and/or structural properties of the CsfC protein.
Chapter 7

Characterisation of the Homologous Proteins CsfB and CsfF

7.1 Introduction

The assembly of cell surface pili relies on the delivery of the individual pilin subunits from the periplasm to the outer membrane assembly protein for translocation onto the cell surface. This process is carried out by specific periplasmic chaperone proteins, which are involved in stabilising the subunits in an assembly competent conformation, protecting the subunits from periplasmic proteases, and delivering the subunits to the outer membrane assembly protein.

PapD from Pap pili is the prototypic member of a superfamily of periplasmic immunoglobulin-like chaperones which are involved in the assembly of over 26 diverse adhesive surface organelles in Gram negative bacteria (Hung et al., 1996). Other members include FaeE from K88 (Mooi et al., 1983), FanE from K99 (Bakker et al., 1991), CssC from CS6 (Knutton et al., 1989; Wolf et al., 1989) and CstA from CS3 pili (Jaajakumari et al., 1989). Although the chaperones FanE, FaeE and PapD are highly homologous they cannot functionally complement each other, which suggests that the interactions between the chaperones and the subunits in these systems is very specific (Bakker et al., 1991).

In contrast, CooB is a periplasmic-like protein from CS1 pili which, by homology, does not form part of the periplasmic superfamily of proteins, but nonetheless can form intermolecular complexes with the pilin subunits CooA and CooD and stabilise these proteins (Sakellaris et al., 1996; Voegele et al., 1997). Unlike other chaperones described, CooB also has a positive effect on the expression of the outer membrane protein CooC, since a cooB deletion results in an increase in the truncated form of CooC (Voegele et al., 1997).
A model for how periplasmic chaperones bind, stabilise and deliver pilin subunits to the outer membrane assembly protein have come from studies on PapD. This protein is known to function in the initial stages of biogenesis by facilitating the importation of subunits into the periplasm (Jones et al., 1997). In the absence of PapD, or if PapD is unable to bind the pilin subunits, the subunits misfold, aggregate and are targeted for degradation by the DegP protease (Jones et al., 1997). PapD recognises and binds to a highly conserved C-terminus motif present on the pilin subunits (Kuehn et al., 1993). The chaperone-subunit complexes are then targeted to the outer membrane assembly protein, where the chaperone is disassociated and the chaperone-subunit interactions are exchanged for subunit-subunit interactions which drive the assembly of the pilus organelle (Dodson et al., 1993).

This chapter describes the characterisation of the two homologous proteins from the csf gene cluster, namely CsfB and CsfF. This chapter shows that both of these proteins exhibit chaperone-like characteristics for the major subunit CsfA, in the case of CsfB and the two minor subunits CsfD and CsfE in the case of CsfF.

7.2 Results

7.2.1 Construction of an In-Frame Deletion in the csfF Gene

The csfF gene was targeted for deletion in pPM5631 by utilising a unique StuI site located across nt 6229-6234 (Appendix I). pPM5631 was digested with StuI which results in exonuclease III sensitive blunt DNA ends. Nested deletions were carried out essentially according to section 2.9.8.3. Approximately 10 μg of pPM5631 was digested with StuI, and exonuclease III was added and samples were taken every 15 sec for 1 min at 37°C. The reaction was stopped and the DNA was then treated with Klenow enzyme and S1 nuclease, ligated with T4 DNA ligase and transformed into E. coli K-12 strain DH5α. Transformants were selected which retained at least part of the csfF gene by utilising the specific csfF oligonucleotides #2971 and #2972 (Table 2.3) in a PCR reaction on boiled lysates from the transformants (section 2.9.8.2). Chosen csfF deletions were sequenced to ensure the deletion
retained the correct reading frame, using oligonucleotides #2971 and #2972 (section 2.11.2). One such deletion obtained resulted in the juxtapositioning of nt 6070 and nt 6397, thereby deleting 327 nt of the csfF gene and retaining the correct reading frame. This plasmid was selected for further use and termed pPM5655 (Figure 7.1).

7.2.2 Construction of a csfF Complementing Plasmid

To construct a complementing plasmid, the wild type csfF gene was PCR amplified from pPM5631 using oligonucleotides #2971 and #2972, which contain EcoRI and XbaI sites respectively (Table 2.3). The PCR fragment obtained was then digested with the appropriate restriction enzymes and cloned into EcoRI and XbaI cut pWKS130 (Wang and Kushner, 1991) to give pPM5659, with the csfF gene orientated behind the inducible lac promoter (Figure 7.2). The plasmid was used in trans to complement the introduced in-frame deletion of csfF in pPM5655.

7.2.3 The Effect of a csfF Deletion on Biogenesis of CS5 pili

To assess the any phenotypic effects of an in-frame deletion in the csfF gene on CS5 pilus assembly, slide agglutinations, haemagglutinations and immunogold electron microscopy were performed. Furthermore, the effect of a csfF deletion on CsfA and CsfD pilin subunit accumulation was assessed by western immunoblot analysis on periplasmic fractions, supernatant and heat extract samples.

7.2.3.1 Initial Phenotypic Characterisation

The outcome of a deletion in the csfF gene on CS5 pilus biogenesis was initially assessed by slide agglutinations using rabbit anti-CS5 pili antiserum (section 2.18). E. coli K-12 harbouring pGEM-7Zf+ or pPM5655 did not agglutinate with the antiserum used. Conversely, E. coli K-12 with pPM5631 or pPM5655 complemented with pPM5659 showed
Figure 7.1 Construction of a csfF deletion mutant from pPM5631.

Plasmid pPM5631 was digested with StuI, with a single site is located across nt 6229-6234 within the csfF gene (orientated according to the arrow) and subjected to bi-directional deletions using exonuclease III. The DNA was then treated with Klenow enzyme, S1 nuclease and T4 DNA ligase to give pPM5655, which contains an in-frame deletion of the csfF gene, by the juxtapositioning of nt 6070 and 6397.
Cleave with Stul, exonuclease III deletion.
Figure 7.2 Construction of a complementing csfF plasmid.

The csfF gene was PCR amplified using oligonucleotides #2971 and #2972 (Table 2.3) which contain EcoRI and XbaI restriction sites respectively, from the template pPM5631. The amplified csfF gene fragment, which is orientated according to the direction of the arrow, was then digested with the appropriate restriction enzymes and ligated into EcoRI and XbaI cut pWKS130 (Wang and Kushner, 1991) to generate pPM5659 which contains the csfF gene behind the lac promoter.
PCR with oligonucleotides #2971, #2972

Cleave with EcoRI and XbaI

Ligation

Cleave with EcoRI and XbaI
immediate agglutination at room temperature. This indicates the csfF deletion may alter cell surface assembly of CS5 pili.

This observation was also confirmed by performing MRHA of human group A⁺ erythrocytes (section 2.19) and immunogold electron microscopy (section 2.20). No haemagglutination of the erythrocytes was observed with E. coli K-12 harbouring pPM5655; however, haemagglutination of the erythrocytes was restored when E. coli K-12 harbouring pPM5655 was complemented with pPM5659 (Figure 7.3).

Examination of the same strains by immunogold electron microscopy revealed that E. coli K-12 harbouring pPM5631 or pPM5655 complemented with pPM5659 produced peritrichous CS5 pili (Figure 7.4, panel B and F) and E. coli K-12 harbouring pGEM-7Zf⁺ was devoid of CS5 pili (Figure 7.4, panel A). E. coli K-12 containing pPM5655 showed cell surface assembly of CS5 pili, but in far less numbers, while the pili appeared to be far greater in overall length than that observed in the wild type or complemented mutant (Figure 7.4, panel C-E). This phenotype is very similar to that observed in a csfE deletion strain (Figure 5.3, panel D-E), which suggests that both CsfE and CsfF may be directly responsible for controlling pilus length. The reduction in cell surface pili, along with the increased susceptibility to shearing forces on longer pili may explain the negative slide agglutination and haemagglutination results obtained for the csfF deletion mutant strain.

7.2.3.2 Periplasmic, Heat Extract and Cell Supernatant Analysis

The effect of an in-frame deletion on the stability of the CsfA and CsfD pilins was examined by preparing periplasmic extracts from E. coli K-12 harbouring pGEM-7Zf⁺, pPM5631, pPM5655 or pPM5655 complemented with pPM5659 along with whole cell extracts. Periplasmic extracts were prepared from E. coli K-12 strains by resuspending 2x 10¹⁰ cells in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1 in the presence of 100 μg/ml of lysozyme in 0.1 M EDTA, pH 7.3. The supernatant resulting after centrifugation was the periplasmic fraction (section 2.14). Whole cell samples were prepared by adjusting the
Figure 7.3 Requirement of CsfF for CS5 pilus mediated haemagglutination.

Samples of *E. coli* K-12 containing pPM5655 or pPM5655 + pPM5659 were prepared by gently resuspending the bacteria to an OD_{600nm} = 4.0 in 1x PBS. 10 μl was mixed with an equal volume of human group A⁺ erythrocytes in the presence of 0.1 M D(+) -mannose and gently rocked. Clear haemagglutination of the erythrocytes, which appear as darkened granules, only occurs when the *csfF* mutation is restored (pPM5655 + pPM5659).
Figure 7.4 Immunogold electron microscopy analysis.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pGEM-7Zf+, B. pPM5631, C-E. pPM5655, or F. pPM5655 + pPM5659. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
harvested bacteria to $1 \times 10^{10}$ cells in a volume of 500 µl. For both periplasmic and whole cell extracts, 40 µl of sample was mixed with an equal volume of 2x SDS-sample buffer and analysed by western immunoblots following SDS-15% PAGE.

The deletion of the csfF gene did not result in any significant difference in the periplasmic accumulation of both CsfA and CsfD pilins when compared to E. coli K-12 harbouring pPM5631 or pPM5655 complemented with pPM5659 by western analysis using anti-CS5 pili antiserum (Figure 7.5). An increase in the whole cell concentration of CsfA was evident in the csfF deletion strain (E. coli K-12 with pPM5655) when compared to E. coli K-12 harbouring pPM5631 or pPM5655 complemented with pPM5659 (Figure 7.5). This is similar to that observed for whole cells of the csfE deletion strain (Figure 5.6) and is likely a reflection of the hydrophobic nature of CS5 pili, whereby cell-free longer pili re-associate with surface pili and are centrifuged with the whole cells. Interestingly, no parallel increase in the accumulation of CsfD in the whole cells was detected.

Heat extracts and cell supernatants were also prepared from the same strains (section 2.17.1), which were grown to confluence overnight on CFA agar at 37°C with the appropriate antibiotics and IPTG as required. A 40 µl of each sample was mixed with an equal volume of 2x SDS-sample buffer prior to SDS-15% PAGE and western immunoblot analysis using rabbit anti-CS5 pili antiserum. Examination of heat extracts derived from the strains revealed that E. coli K-12 harbouring pPM5655 produces pili with apparently little or no CsfD minor pilin, when compared to E. coli K-12 with pPM5631 or pPM5655 complemented with pPM5659 (Figure 7.6A). The amount of CsfA detected remained relatively constant. In the supernatant extracts, once again no CsfD was detected with E. coli K-12 harbouring pPM5655, while the expression of CsfA appears to be slightly less than E. coli K-12 containing pPM5631 (Figure 7.6B). This data indicates that CsfF may also play a role in delivering CsfD across the outer membrane during CS5 biogenesis, but is not necessarily required for stabilising CsfD in the periplasm, since CsfD levels remain unchanged in this fraction in a csfF deletion mutant.
Figure 7.5 Whole cell and periplasmic analysis of the csfF deletion mutant.

Western immunoblot analysis of A. Whole cell samples and B. Periplasmic extracts (section 2.14) prepared from *E. coli* K-12 harbouring either pGEM-7Zf*, pPM5631, pPM5655, or pPM5655 + pPM5659, using anti-native CS5 pili antiserum. The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Each sample represents $1 \times 10^9$ cells (whole cells) and $2 \times 10^9$ cells (periplasm). Molecular mass standards are indicated.
A

B

kDa

pGEM-7Zf+
pPM5631
pPM5655
pPM5655 + pPM5659

43
30
20.1
Figure 7.6 Western analysis on heat extracts and whole cell supernatants.

Heat extracts and whole cell supernatants were prepared from the various strains (section 2.17.1). The western immunoblot was performed using rabbit anti-native CS5 pili antiserum (1:5000) on heat extracts (A) or cell supernatants (B) of *E. coli* K-12 harbouring either pGEM-7Zf⁺, pPM5631, pPM5655 or pPM5655 + pPM5659. Samples represent $1 \times 10^9$ cells. The solid arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
7.2.4 Specific Modulation of csfF Expression

The phenotype of a deletion mutant in the csfF gene is identical to that of a csfE deletion, with the CS5 pili produced appearing longer than that observed in the wild type strain. Therefore, the effect of the specific induction of csfF expression on the length of CS5 pili by immunogold electron microscopy was examined followed by specific pilus measurements in terms of length and cell surface numbers.

The csfF gene was excised from pPM5659 by digestion with EcoRI and XbaI and the fragment was cloned directly into EcoRI-XbaI cut pBAD18-Kan (Guzman et al., 1995) to give pPM5672 in which the csfF gene is orientated from the inducible P_{BAD} promoter of the araBAD operon (Figure 7.7). The plasmid pPM5672 was introduced in trans into E. coli K-12 harbouring either pPM5631 or pPM5655 and the strains grown on CFA agar in the presence of the appropriate antibiotics and two different concentrations of arabinose, namely 0.002% and 0.01%. As previously described in section 5.2.4.2, the presence of both plasmids in the same strain leads to a reduction in the numbers of cell surface CS5 pili expressed, but does not significantly alter the length of the pili. E. coli K-12 harbouring pPM5655 and pPM5659 was also examined for any potential effects on the length of CS5 pili in the presence or absence of IPTG.

The mean number of surface located pili was calculated on fifty randomly selected bacteria ± SEM, while the mean length of pili was calculated by recording the length of five pili on ten randomly selected bacteria (n=50) ± SEM. Since CS5 pili bundle, the mean number and length is representative of discrete bundles of pili. The mean length of CS5 pili in E. coli K-12 with pPM5655 was \(1.54 \pm 0.091 \mu m\) with \(14.08 \pm 0.893\) pili per cell \((P<0.0001)\)(Table 7.1; Figure 7.8, panel A) compared to pPM5631 with a mean length of \(0.563 \pm 0.038 \mu m\) with \(43.21 \pm 1.101\) pili per cell (Table 7.1; Figure 7.8, panel B). Therefore, a deletion in the csfF gene results in far fewer cell surface associated CS5 pili, but the pili are approximately three times longer compared to the wild type.
Figure 7.7  Construction of pPM5672 containing csfF from the P_{BAD} promoter.

The csfF gene was excised from pPM5659 as an EcoRI-XbaI fragment and cloned directly into EcoRI and XbaI cut pBAD18-Kan (Guzman et al., 1995) to create pPM5672, which contains the csfF gene under the influence of the P_{BAD} promoter derived from the araBAD (arabinose) operon.
Table 7.1 Effect of Csff on CS5 pilus length.

<table>
<thead>
<tr>
<th>E. coli K-12 containing</th>
<th># pili per cell&lt;sup&gt;a&lt;/sup&gt;</th>
<th>length of pili (μm)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPM5631</td>
<td>43.21 ± 1.101</td>
<td>0.563 ± 0.038</td>
</tr>
<tr>
<td>pPM5631 + pPM5672 + 0.002% arabinose</td>
<td>0.480 ± 0.104&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N. D</td>
</tr>
<tr>
<td>pPM5655</td>
<td>14.08 ± 0.893&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.540 ± 0.091&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5655 + pPM5659 - IPTG</td>
<td>17.26 ± 0.767&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.703 ± 0.025&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5655 + pPM5659 + IPTG</td>
<td>19.30 ± 0.703&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.771 ± 0.023&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5655 + pPM5672 + 0.002% arabinose</td>
<td>4.96 ± 0.536&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.247 ± 0.015&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPM5654 + pPM5672 + 0.01% arabinose</td>
<td>N. D</td>
<td>N. D</td>
</tr>
</tbody>
</table>

<sup>a</sup> Mean number of pili on 50 randomly selected bacteria ± SEM.

<sup>b</sup> Mean length of 5 pili on 10 randomly selected bacteria (n=50) ± SEM.

<sup>c</sup> P< 0.001 compared to wild type (pPM5631) by using the unpaired 2 tailed t test.

N.D Not Determined.
Figure 7.8 Visual CS5 pili production following CsfF modulation by immunogold electron microscopy.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pPM5655, B. pPM5631, C. pPM5631 + pPM5672 (0.002% arabinose), D. pPM5655 + pPM5659 (- IPTG), E. pPM5655 + pPM5659 (+ IPTG), F. pPM5655 + pPM5672 (0.002% arabinose). The different bacterial strains were first reacted with rabbit antiserum raised against native CS5 pili (1:10), then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
**E. coli** K-12 harbouring both pPM5655 and the compatible low copy number plasmid pPM5659 grown in the presence or absence of IPTG reduced the mean length of the pili to 0.771 ± 0.023 and 0.703 ± 0.025 μm, respectively, but was still significantly longer than that observed in the wild type (P < 0.0001) (Table 7.1; Figure 7.8, panel D-E). The mean number of pili per cell was 19.30 ± 0.703 and 17.26 ± 0.767 in the presence or absence of IPTG (Table 7.1). The presence of IPTG in the growth medium did not result in significant changes in pili length or numbers than without IPTG present.

**E. coli** K-12 harbouring pPM5655 complemented with pPM5672 grown in the presence of 0.002% arabinose resulted in a significant decrease in the mean length of pili to 0.247 ± 0.015 μm (Table 7.1; Figure 7.8, panel F), while the relative number of cell associated pili also was significantly lower (4.96 ± 0.536) if compared to **E. coli** K-12 with pPM5631 and pBAD18-Kan (11.82 ± 1.09). When the same strain was grown in the presence of 0.01% arabinose, no surface associated pili was evident (data not shown). Moreover, when **E. coli** K-12 harbouring pPM5631 was complemented with pPM5672 in the presence of 0.002% arabinose, the bacteria were devoid of pili (0.480 ± 0.104) and therefore no mean pilus length was recorded (Table 7.1; Figure 7.8, panel C).

Taken together, these results indicate a direct association between the level of expression of CsfF in the periplasm and the mean length of the pilus.

### 7.2.5 Construction of an In-Frame Deletion in the csfB Gene

An in-frame deletion was also constructed in the csfB gene, since this gene bears considerable homology to the csfF gene, thus the two proteins encoded by these genes are likely to play a similar role in CS5 pilus biogenesis. A double PstI site located within the csfB gene at nt 2318-2323 and nt 2327-4232 of the characterised sequence (Appendix I) was utilised for exonuclease III mediated deletions. A PstI restriction produces 3' overhangs which are resistant to exonuclease III, therefore to overcome this the 3' overhangs were blunt ended by utilising the 3'-5' exonuclease activity of the Klenow enzyme. Blunt end DNA
fragments are susceptible to exonuclease III. Nested deletions were carried out essentially according to section 2.9.8.3. Approximately 10 μg of pPM5631 was digested with PstI, and the DNA fragment treated with Klenow enzyme (1 U/μg DNA) without dNTPs for 10 min at 37°C followed by inactivation of the enzyme at 65°C for a further 15 min. Exonuclease III was added which deletes bi-directionally from the blunt ends on the cut plasmid. Samples were taken every 15 sec for 1 min at 37°C. The reaction was stopped and the DNA was then treated with Klenow enzyme and S1 nuclease, ligated with T4 DNA ligase and transformed into *E. coli* K-12 strain DH5α. Transformants were selected which retained at least part of the *csfB* gene by utilising the specific *csfB* oligonucleotides #2944 and #2945 (Table 2.3) in a PCR reaction on boiled lysates from the transformants (section 2.9.8.2).

Chosen *csfB* deletions were sequenced to ensure the deletion retained the correct reading frame, using oligonucleotides #2944 and #2945 (section 2.11.2). One such deletion obtained resulted in the juxtapositioning of nt 2157 and 2602, thereby deleting 445 nt of the *csfB* gene and retaining the correct reading frame. This plasmid was selected for further use and termed pPM5662 (Figure 7.9).

### 7.2.6 Construction of a *csfB* Complementing Plasmid

The wild type *csfB* gene was PCR amplified from pPM5631 using oligonucleotides #2944 and #2945, which contain *Eco*RI and *Xba*I sites respectively (Table 2.3). The PCR fragment obtained was then digested with the appropriate restriction enzymes and cloned into *Eco*RI and *Xba*I cut pWKS130 (Wang and Kushner, 1991) to give pPM5665, with the *csfB* gene orientated behind the inducible *lac* promoter (Figure 7.10). The plasmid was used in *trans* to complement the introduced in-frame deletion of *csfB*.

### 7.2.7 The Role of CsfB in CS5 Pilus Biogenesis

The following sections describe the proposed role of CsfB in CS5 pilus biogenesis by characterising the phenotype of a *csfB* deletion versus the wild type or complemented mutant
Figure 7.9 Construction of a csfB deletion mutant from pPM5631.

Plasmid pPM5631 was initially digested with PstI, with a double site located across nt 2318-2323 and nt 2327-4232 within the csfB gene (orientated according to the arrow). The 3' overhangs were blunt ended with Klenow enzyme and then subjected to bi-directional deletions using exonuclease III. The DNA was then treated with Klenow enzyme, S1 nuclease and T4 DNA ligase to give pPM5662, which contains an in-frame deletion of the csfB gene, by the juxtapositioning of nt 2157 and nt 2602.
Cleave with *Pst*I, exonuclease III deletion.

Ligation

modified \(csf\) region (\(\Delta csfB\))
Figure 7.10 Construction of the complementing csfB plasmid pPM5665.

The csfB gene was PCR amplified using oligonucleotides #2944 and #2945 (Table 2.3) which contain EcoRI and XbaI restriction sites respectively, from the template pPM5631. The amplified csfB gene fragment, which is orientated according to the direction of the arrow, was then digested with the appropriate restriction enzymes and ligated into EcoRI and XbaI cut pWKS130 (Wang and Kushner, 1991) to generate pPM5665 which contains the csfB gene orientated from the lac promoter.
PCR with oligonucleotides #2944, #2945

Cleave with EcoRI and XbaI

Cleave with EcoRI and XbaI

Ligation

Kan^R

XbaI

EcoRI

Kan^R

XbaI

EcoRI

lacZ^a

CSJf

region

pPM5631

csf region

Smal

pSC101 origin

T7

T3

pWKS130

csf region

lacZ^a

MCS

XbaI

EcoRI

EcoRI
using western blot analysis on whole cells, periplasmic preparations, pili preparations and cell supernatants, along with immunogold electron microscopy.

7.2.7.1 Initial Characterisation

_E. coli_ K-12 harbouring either pGEM-7Zf^+, pPM5662, pPM5662 complemented with pPM5665 or pPM5631 were grown on CFA agar with the appropriate antibiotics and IPTG where required and slide agglutinations performed using rabbit anti-CS5 pili antiserum (section 2.18). _E. coli_ K-12 harbouring pPM5631 or pPM5662 complemented with pPM5665 agglutinated, whereas strains harbouring pPM5662 or pGEM7-Zf^+ did not agglutinate. When the same strains were examined by immunogold electron microscopy (section 2.20), _E. coli_ K-12 containing either pPM5631 or pPM5662 complemented with pPM5665 produced peritrichous CS5 pili (Figure 7.11, panel B and D), while _E. coli_ K-12 strains containing pGEM-7Zf^+ or pPM5662 were devoid of CS5 pili (Figure 7.11, panel A and C). The _csfB_ mutant strain was also tested for the ability to cause MRHA of human group A^+ erythrocytes (section 2.19). As expected from the immunogold electron microscopy data, no haemagglutination ability was conferred by _E. coli_ K-12 with pPM5662, while haemagglutination of the human erythrocytes was restored when the _csfB_ mutant was complemented with pPM5665 (Figure 7.12). These results indicate the importance of CsfB for cell surface assembly of CS5 pili.

Western immunoblot analysis following SDS-15% PAGE using rabbit anti-CS5 pili antiserum was also performed on both heat extracts and cell supernatants (section 2.17.1) prepared from the strains, which were grown to confluence overnight on CFA agar at 37°C. 40 μl of sample was mixed with an equal volume of 2x SDS-sample buffer prior to SDS-15% PAGE. _E. coli_ K-12 harbouring pPM5651 showed an abolition of CsfA, but not CsfD in heat extracts and cell supernatants when compared to _E. coli_ K-12 harbouring pPM5631 and _E. coli_ K-12 with pPM5662 and pPM5665 (Figure 7.13). Therefore, the absence of the CsfB
Figure 7.11 Immunogold electron microscopy analysis.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pGEM-7Zf+, B. pPM5631, C. pPM5662, or D. pPM5662 + pPM5665. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
Figure 7.12 Requirement of CsfB for CS5 pilus mediated haemagglutination.

Samples of *E. coli* K-12 containing pPM5662 or pPM5662 + pPM5665 were prepared by gently resuspending the bacteria to an OD<sub>600nm</sub> = 4.0 in 1x PBS. 10 µl was mixed with an equal volume of human group A<sup>+</sup> erythrocytes in the presence of 0.1 M D(+)-mannose and gently rocked. Clear haemagglutination of the erythrocytes, which appear as darkened granules, only occurs when the csfB mutation is restored (pPM5662 + pPM5665).
Figure 7.13 Western analysis on heat extracts and whole cell supernatants.

Heat extracts and whole cell supernatants were prepared from the various strains (section 2.17.1). The western immunoblot was performed using rabbit anti-native CS5 pili antiserum (1:5000) on cell supernatants (A) or heat extracts (B) of *E. coli* K-12 harbouring either pGEM-7Zf+, pPM5631, pPM5662 or pPM5662 + pPM5665. Samples represent 1x 10^9 cells. The solid arrow indicates position of CsfA protein and the empty arrow CsfD. Molecular mass standards are indicated.
protein in the cells is affecting the release of CsfA, but not CsfD across the outer membrane, since CsfD is readily detected in both heat extracts and cell supernatants.

7.2.7.2 The Effect of a csfB Deletion on the Periplasmic Stability of the CS5 Pilins

Periplasmic extracts were prepared from *E. coli* K-12 strains by resuspending 2x $10^{10}$ cells in 1 ml of 20% (w/v) sucrose, 30 mM Tris-HCl, pH 8.1. Cells were converted to sphaerooplasts with 100 µg/ml of lysozyme in 0.1 M EDTA, pH 7.3 and the resulting supernatant collected after centrifugation was the periplasmic fraction (section 2.14). Whole cell samples were prepared by adjusting the harvested bacteria to 1x $10^{10}$ cells in a volume of 500 µl. For both periplasmic and whole cell extracts, 40 µl of sample was mixed with an equal volume of 2x SDS-sample buffer and analysed by western immunoblots following SDS-15% PAGE. *E. coli* K-12 harbouring pPM5662 resulted in no detectable CsfA protein in both whole cell and concentrated periplasmic preparations when compared to *E. coli* K-12 with pPM5631 and *E. coli* K-12 harbouring pPM5662 complemented with pPM5665 (Figure 7.14). There was also a corresponding decrease in CsfD in the periplasm and whole cell sample. This data clearly indicates that CsfB is specifically required for CsfA stability in the periplasm, and may also play some role in stabilising CsfD, since a decrease in expression of this pilin was also observed in the csfB mutant.

7.2.8 Localisation of CsfB in *E. coli* K-12

The cellular location of CsfB in *E. coli* K-12 was determined by radioactive labelling of the protein and fractionation experiments. The CsfB protein was labelled with L-[35S]-methionine following expression from pPM5605, in which the csfB gene is orientated from the T7 promoter (Figure 3.12). Specific expression was carried out according to section 2.12.3.2. The cells were then fractionated into the individual cell components: whole cells (w/c), inclusion bodies (i/b), cytoplasms (cyto), periplasms (peri), outer membranes (o/m) and inner membranes (i/m) according to section 2.14. A sample of each was analysed by SDS-
Figure 7.14 Whole cell and periplasmic analysis of the csfB deletion mutant.

Western immunoblot analysis of A. Whole cell samples and B. Periplasmic extracts (section 2.14) prepared from E. coli K-12 harbouring either pGEM-7Zf+, pPM5631, pPM5662, or pPM5662 + pPM5665, using anti-native CS5 pili antiserum (1:5000). The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Each sample represents 1x 10^6 cells (whole cells) and 2x 10^6 cells (periplasm). Molecular mass standards are indicated.
15% PAGE and the gel stained with Coomassie Brilliant Blue and destained with 5% acetic acid. The L-[\textsuperscript{35}S]-methionine labelled CsfB protein was detected by autoradiography of the dried gel (section 2.13). CsfB was found to localise in the periplasmic fraction, and was not associated with either the outer or inner membrane fractions (Figure 7.15).

### 7.2.9 Over-expression of CsfB and Antiserum Preparation

The csfB gene was PCR amplified from pPM5631 using oligonucleotides #3110 and #3111 which contain BamHI and KpnI sites respectively (Table 2.3). The csfB fragment was then digested with the same restriction enzymes and cloned into BamHI and KpnI cut pQE-31 such that the csfB gene is cloned in-frame with an N-terminal (His)$_6$-Tag, to give plasmid pPM5679 (Figure 7.16). The plasmid was then isolated from E. coli K-12 strain DH5\(\alpha\) and transformed into strain BL21 (DE3) harbouring pREP4 (Table 2.1) for specific over-expression of the CsfB protein, along with the negative control plasmid pQE-31. Specific over-expression was carried out according to section 2.12.3.3. Briefly, inoculated bacterial cultures were grown in the presence of Amp (100 \(\mu\)g/ml) and Kan (25 \(\mu\)g/ml) to mid exponential phase of growth (OD$_{600nm}$ = 0.6) and IPTG added to a final concentration of 1 mM for 3 hr at 37°C with vigorous aeration. A 50 ml sample of induced bacterial cells was then fractionated into whole cells (w/c), periplasms (peri), cytoplasms (cyto) and inclusion bodies (i/b) according to section 2.14. A whole cell sample of uninduced cells was also included as a control. The over-expressed CsfB protein was detectable in the whole cell sample containing pPM5679 and was found to localise in the inclusion body fraction (Figure 7.17).

A sample of the inclusion body fraction (1 ml) was subjected to metal exchange chromatography using Ni-NTA agarose (Qiagen) in an attempt to purify the (His)$_6$-tagged CsfB protein to vaccinate a rabbit for polyclonal antiserum preparation. The metal exchange chromatography was carried out according to section 2.12.3.3. Although some CsfB could be detected in the supernatant from the nickel-NTA washes, the majority of the CsfB protein was
Figure 7.15 Localisation of L-[\textsuperscript{35}S]-methionine labelled CsfB in *E. coli* K-12.

The CsfB protein was labelled with L-[\textsuperscript{35}S]-methionine following expression from pPM5605, in which the *csfB* gene is orientated from the T7 promoter. A negative control plasmid (pBS-SK\textsuperscript{+}) was also included. Specific expression was carried out according to section 2.12.3.2. The cells were then fractionated into the individual cell components: whole cells (w/c), cytoplasms (cyto), periplasms (peri), inclusion bodies (i/b), inner membranes (i/m) and outer membranes (o/m) according to section 2.14, and run on SDS-15% PAGE followed by Coomassie Brilliant Blue staining and destaining in 5% acetic acid. The L-[\textsuperscript{35}S]-methionine labelled CsfB protein was detected by autoradiography of the dried gel (section 2.13) and is indicated by an arrow. Molecular mass markers are indicated.
The csfB gene was PCR-amplified from pPM5631 using the specific oligonucleotides #3110 and #3111 (Table 2.3) which contain BamHI and KpnI sites respectively, and cloned into BamHI-KpnI digested pQE-31 (Qiagen) to generate pPM5679, in which the csfB gene is fused in-frame to a sequence encoding for 6x Histidine residues.
PCR with oligonucleotides #3110, #3111

Cleave with BamHI, KpnI

Ligation

Cleave with BamHI, KpnI
Figure 7.17 Over-expression and fractionation of (His)$_6$-tagged CsfB from pPM5679.

Coomassie Brilliant Blue stained SDS-15% PAGE of *E. coli* K-12 strain BL21 (DE3) harbouring pQE-31 or pPM5679. Over-expression of (His)$_6$-tagged CsfB from pPM5679 by the addition of 1 mM IPTG was carried out according to section 2.12.3.3. The bacteria were fractionated as described in section 2.14 into whole cells (w/c), inclusion bodies (i/b), periplasms (peri) and cytoplasms (cyto). The CsfB protein is indicated by an asterix (*) and is localised in the i/b fraction. Molecular mass standards are indicated.
present in the elution steps from the nickel-NTA agarose using both buffer D and buffer E (Figure 7.18). Numerous contaminating upper and lower bands were consistently observed in the elution steps, even when the agarose was pre-incubated with increasing concentrations of imidazole, therefore it was decided to gel excise a suitable quantity of the CsfB protein (approximately 200-300 µg) from a preparative gel for immunising a rabbit. The gel excision and immunisation schedule was carried out according to section 2.16.1. The antiserum generated was specifically absorbed against whole cells of live *E. coli* strain BL21 (DE3) containing pQE-31 and neutralised across membranes containing cell extracts of the same strain according to sections 2.16.2 and 2.16.3.

### 7.2.10 Localisation of CsfB in *E. coli* K-12 expressing CS5 pili

The rabbit anti-CsfB antiserum was used to localise CsfB in *E. coli* K-12 harbouring pPM5631 to determine whether the presence of the entire CS5 encoding region results in an altered localisation for the CsfB protein. The CooB chaperone from CS1 pili has been shown to associate directly with the assembly protein CooC in the outer membrane, and therefore is not strictly periplasmically located (Voegele *et al.*, 1997). Expression of the CsfB protein alone in *E. coli* K-12 has been shown by L-[³⁵S]-methionine labelling to localise exclusively in the periplasmic fraction (Figure 7.15).

*E. coli* K-12 harbouring pPM5631 was fractionated according to section 2.14. Western analysis of periplasms (peri), whole membranes (w/m), inner membranes (i/m) and outer membranes (o/m) showed that CsfB could only be detected in the periplasmic fraction and was not associated with the membranes in the presence of the complete *csf* gene cluster (Figure 7.19A). However, the potential negative effects on protein-protein interactions in the outer membrane during the fractionation process cannot be discounted.

To determine whether CsfB may constitute a minor component of CS5 pili, western analysis using the rabbit anti-CsfB antiserum was performed on different quantities of caesium purified CS5 pili (section 2.17.3) ranging from 10-150 µg. No CsfB was detected in
Figure 7.18 Attempted purification of His-tagged CsfB across Ni-NTA agarose.

A sample of the solubilised inclusion body fractions containing CsfB were incubated with pre-equilibrated Ni-NTA agarose and purification carried out according to section 2.12.3.3. Lane 1: first wash extracts; lane 2: second wash extracts; lane 3: buffer D elution extracts; lane 4: buffer E elution extracts. The CsfB protein is indicated by an arrow. Molecular mass standards are indicated.
Figure 7.19 Localisation of Csfb in *E. coli* K-12 Harbouring pPM5631 and Association with CS5 Pili.

A. *E. coli* K-12 harbouring pGEM-7Zf⁺ or pPM5631 were fractionated into periplasm (peri), whole membrane (w/m), outer membrane (o/m) and inner membrane (i/m) fractions as described in section 2.14. Western immunoblot analysis was performed using rabbit anti-Csfb pili antiserum (1:1000) on samples representing 2x 10⁹ cells. B. Western immunoblot analysis of 10-150 µg of purified CS5 pili, along with a sample of an inclusion body fraction from pPM5679 over-expressing Csfb, using anti-Csfb antiserum (1:1000). Molecular mass standards are indicated.
the CS5 pili samples, which indicates that CsfB is unlikely to form part of the mature pilus structure (Figure 7.19B). Indeed Coomassie Brilliant Blue staining following SDS-15% PAGE of up to 250 μg of denatured purified CS5 pili produces only two protein bands corresponding to CsfA and CsfD.

7.2.11 Construction of a Double Deletion in the csfB and csfF Genes

Given the significant protein and structural homology identified between CsfB and CsfF (Figure 3.5 and 3.6) it was determined what effect a deletion in both genes has on CS5 pili biogenesis, thereby establishing whether functional redundancy may exist in the CS5 pilus system.

A double deletion in the csfB and csfF genes was constructed by utilising the csfF deletion plasmid pPM5655 (section 7.2.1) to construct a further deletion in the csfB gene. The generation of a specific in-frame deletion in csfB was carried out according to section 7.2.5, utilising the plasmid pPM5655 for exonuclease III digestion. Transformants were selected which retained at least part of the csfB gene by utilising the specific csfB oligonucleotides #2944 and #2945 (Table 2.3) in a PCR reaction on boiled lysates from the transformants (section 2.9.8.2). Chosen csfB deletions were sequenced to ensure the deletion retained the correct reading frame, using oligonucleotides #2944 and #2945 (section 2.11.2). One such deletion obtained resulted in the juxtapositioning of nt 2197 and 2591, thereby deleting 394 nt of the csfB gene and retaining the correct reading frame. This plasmid was selected for further use and termed pPM5690 (Figure 7.20).

7.2.12 Phenotypic Characteristics of the Double Deletion Mutant

The phenotype of the csfB/csfF double deletion was initially assessed by slide agglutination experiments using anti-CS5 pili antiserum with E. coli K-12 strain DH5α harbouring pPM5631 or pPM5690. As would be expected based on the results obtained for single deletion mutants in the csfB or csfF genes, E. coli K-12 containing pPM5690 did not
Figure 7.20 Construction of a csfB-csfF deletion mutant from pPM5655.

The csfF deletion plasmid pPM5655 was initially digested with PstI, with a double site located within the csfB gene (orientated according to the arrow). The 3' overhangs were blunt ended with Klenow enzyme and then subjected to bi-directional deletions using exonuclease III. The DNA was then treated with Klenow enzyme, S1 nuclease and T4 DNA ligase to give pPM5690, which contains an in-frame deletion of the csfB gene, by the juxtapositioning of nt 2197 and nt 2591, along with the previously constructed csfF deletion.
Cleave with *PstI*, exonuclease III deletion

Ligation

modified *csf* region ($\Delta{csfB}$, $\Delta{csfF}$)
agglutinate, whilst strong agglutination occurred for *E. coli* K-12 with pPM5631. When the same strains were examined visually by immunogold electron microscopy, *E. coli* K-12 with pPM5631 showed peritrichous surface CS5 pili (Figure 7.21, panel A), while *E. coli* K-12 with pPM5690 was devoid of cell surface CS5 pili (Figure 7.21, panel B).

7.2.13 Periplasmic Analysis

Periplasmic fractions were prepared from *E. coli* K-12 harbouring pPM5690 or pPM5631, but also *E. coli* K-12 containing pPM5690 complemented in *trans* with a wild type copy of *csfB* (pPM5665) or *csfF* (pPM5659) according to section 2.14.

No CsfA pilin subunits were detected in periplasms prepared from *E. coli* K-12 with pPM5690, when compared to the wild type system of *E. coli* K-12 with pPM5631 (Figure 7.22). This is in agreement with a single deletion in *csfB*, which has been shown to abolish CsfA subunit accumulation in the periplasm (Figure 7.14). However, unlike the single deletion mutants in *csfB* or *csfF*, where CsfD levels are either relatively constant in the case of a *csfF* deletion (Figure 7.5) or slightly reduced for a deletion in *csfB* (Figure 7.14), when both genes are deleted the level of expression of CsfD is greatly reduced, and a CsfD degradation product also accumulates (Figure 7.22). No such degradation product is observed in the negative control strain (*E. coli* K-12 with pGEM-7Zf+) or in *E. coli* K-12 harbouring pPM5631.

7.2.14 Transport of Pilin Subunits Across the Outer Membrane

The transport of CsfA and CsfD pilin subunits across the outer membranes was analysed by examining outer membrane extracts prepared from *E. coli* K-12 harbouring either pPM5690 or pPM5631 according to section 2.14, along with whole cell supernatant extracts.

Examination of outer membrane extracts by western immunoblot analysis showed that no CsfD or CsfA could be detected in the outer membranes of *E. coli* K-12 harbouring pPM5690 when compared to *E. coli* K-12 containing pPM5631 (Figure 7.23A). Furthermore,
Figure 7.21 Immunogold electron microscopy analysis of the csfB-csfF deletion mutant.

Immunogold electron micrographs of 1% uranyl acetate stained *E. coli* K-12 harbouring A. pPM5631, B. pPM5690. The different bacterial strains were reacted first with rabbit antiserum raised against native CS5 pili (1:10) and then with protein A gold conjugated to 20 nm diameter gold particles (1:40). Bar represents 500 nm.
Figure 7.22 Periplasmic analysis of the *csfB-csfF* double deletion mutant.

Western immunoblot analysis of periplasmic extracts (section 2.14) prepared from *E. coli* K-12 harbouring pGEM-7Zf+, pPM5631, or pPM5690 using rabbit anti-native CS5 pili antiserum (1:5000). The solid arrow indicates the position of the CsfA protein, the empty arrow indicates the CsfD protein and the grey arrow shows the CsfD degradation product. A CsfD degradation product is also clearly visibly in the pPM5690 lane. Each sample represents 2x 10⁹ cells. Molecular mass standards are indicated.
Figure 7.23 Delivery and release of pilin subunits in a csfB-csfF deletion strain.

Western immunoblot analysis of (A) outer membranes and (B) supernatants prepared from E. coli K-12 harbouring pPM5631 or pPM5690 using rabbit anti-CS5 pili antiserum (1:5000). Cells were fractionated according to section 2.14. The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. The supernatant sample represents 1x 10⁹ cells. Molecular mass standards are indicated.
no CsfD or CsfA could be detected in supernatant extracts prepared from *E. coli* K-12 harbouring pPM5690, when compared to *E. coli* K-12 harbouring pPM5631 by western immunoblot analysis using anti-CS5 pili antiserum (Figure 7.23B). This data further confirms the observation that a *csfB* mutant does not produce CsfA, but also indicates that CsfD cannot be delivered to and subsequently across the outer membranes in the absence of both CsfB and CsfF proteins.

### 7.2.15 Complementation of pPM5690 with *csfB* or *csfF* in *trans*

The double deletion in pPM5690 was complemented with either *csfB* or *csfF* provided in *trans* from plasmids pPM5665 (Figure 7.10) or pPM5659 (Figure 7.2) respectively to determine the effects on accumulation of CsfA or CsfD in the periplasm. Periplasmic extracts were prepared from the strains, which were grown in the presence of suitable antibiotics and IPTG according to section 2.14. *E. coli* K-12 harbouring pPM5690 complemented with pPM5665 restored the periplasmic accumulation of CsfA, which was absent in *E. coli* K-12 with pPM5690 (Figure 7.24). Again, this is consistent with the hypothesis that CsfB is a periplasmic chaperone for CsfA. The level of restoration achieved was similar to that observed in *E. coli* K-12 with pPM5631. When *E. coli* K-12 harbouring pPM5690 was complemented in *trans* with pPM5659, CsfD accumulation was restored in the periplasm (Figure 7.24), albeit to a lesser extent than the wild type with no CsfA accumulation. This data shows that CsfF is required for CsfD stability in the periplasm in the absence of CsfB.

### 7.3 Summary and Conclusions

This chapter described the characterisation of the CsfB and CsfF proteins from the CS5 gene cluster, which have been shown to share significant protein sequence homology (30% identity, 51% similarity), but also structural homology (Figure 3.6). Initially, in-frame deletions were constructed in each of the *csfB* and *csfF* genes, which deleted 455 nt and 327 nt from each gene respectively, and was sufficient to confer an altered phenotype when
Periplasmic analysis of specific complementations of pPM5690 with pPM5665 (csfB) or pPM5659 (csfF) in trans.

Western immunoblot analysis of periplasmic extracts (section 2.14) prepared from E. coli K-12 harbouring either pGEM-7Zf, pPM5631, pPM5690, pPM5690 + pPM5659 or pPM5690 + pPM5665, using rabbit anti-native CS5 pili antiserum (1:5000). The solid arrow indicates the position of the CsfA protein, and the empty arrow indicates the CsfD protein. Each sample represents $2 \times 10^9$ cells. Molecular mass standards are indicated.
compared to the wild type in each case. As expected with the in-frame deletions, no downstream polar effects were observed, since each deletion was complemented in trans with a complete copy of the gene to restore the wild type phenotype.

A deletion in the csfB gene was found to completely abolish CsfA subunit accumulation in the periplasm, and therefore bacteria with this deletion did not produce detectable cell surface CS5 pili. However, the deletion in CsfB did not effect the transport of the minor pilin subunit CsfD into and across the outer membrane, since CsfD subunits were detected in both outer membrane fractions and supernatant samples from the csfB deletion strain. CsfB is therefore a major subunit-specific chaperone. Localisation studies from radio-labelled CsfB protein and using anti-CsfB antiserum on fractions derived from the wild type strain confirmed the periplasmic location of CsfB, and showed CsfB did not form part of the mature pilus structure.

A deletion constructed in the csfF gene resulted in two distinct changes in the pilus structure. Firstly, the pili produced in this strain were approximately three times longer than that observed in the wild type, and secondly, no CsfD was detected in the longer pili structures, or indeed released into the supernatant. Specific modulation of CsfF expression from P_BAD in a csfF deletion strain resulted in significantly shorter and less numerous pili, and when CsfF was introduced into the wild type strain, cell surface pilus expression was abolished. This shows that CsfF is rate-limiting for control of pilus length, and unlike CsfE, increasing expression of CsfF results in significantly shorter pili when compared to the wild type. Based on the homology to the CsfB chaperone, CsfF is therefore predicted to function as a specific chaperone for delivering CsfE to the outer membrane assembly protein CsfC, to terminate pilus elongation.

The elongated pili produced in a csfF deletion differed in composition from the normal wild type pili and the elongated pili observed in a csfE deletion, since no CsfD minor subunits could be detected. However, CsfD was detected in levels comparable to the wild type in the periplasm of a csfF deletion strain. Therefore, the absence of csfF does not effect the
accumulation of CsfD in the periplasm, but certainly abrogates delivery of CsfD across the outer membrane. This implies that CsfF is a specific chaperone for CsfD, but may not required for the stability of CsfD. Clearly some CsfD must be assembled onto the cell surface, since CsfD has been previously shown to be absolutely required for initiation of CS5 pilus biogenesis (chapter 4). However, the level of CsfD is undetectable under the conditions used.

To explain this observation, a double in-frame deletion in the csfB and csfF genes was constructed, thereby establishing directly whether some functional redundancy may exist between CsfB and CsfF. This double deletion mutant resulted in the complete absence of CsfA subunits in the periplasm (this was expected based on the results for a single deletion in the csfB gene), but more importantly a significant decrease in the accumulation of CsfD was evident. Furthermore, a CsfD degradation product accumulated. When a wild type copy of csfF was provided in trans, this resulted in the re-appearance of stable CsfD in the periplasm. Likewise when csfB was re-introduced in trans, CsfA accumulation in the periplasm was restored. This indicated that CsfF also functions as a specific chaperone for CsfD, and is predicted to function as a chaperone for CsfE.

A protein alignment between CsfD and CsfE was performed to examine whether the two proteins share any similar or identical amino acids stretches that may serve as recognition domains for the CsfF chaperone. Although very dissimilar, a putative recognition domain for the CsfF chaperone is indicated in Figure 7.25. When the secondary structure of the two aligned proteins was examined using PHDSec (Rost et al., 1994) at the Pôle Bio-Informatique Lyonnais, no structural similarity was observed. Therefore, the recognition site for CsfF binding to CsfD and CsfE remains highly speculative.

This is the first description of a dual chaperone system in any human ETEC pilus system, in which one chaperone is directly responsible for stabilising and delivering the major subunit protein and the other is specifically responsible for stabilising and delivering minor subunits to the outer membrane. Only one other pilus biogenesis system, 987P, which is an
Figure 7.25 Comparison of amino acid sequences of CsfD and CsfE.

Identical amino acids are indicated by an asterix (*), conserved residues with a dot (.), gaps (-) have been introduced to optimise alignment. The analysis was performed with CLUSTAL W (Thompson et al., 1994). A possible recognition domain of the CsfD and CsfE proteins is indicated by the shaded box region.
| CsfE | MK-------------------IKKFKLVFYMVIFYVISLQEVLSASTSVITN-------MKKQPIVTLGFFSMSMVQAATTVTSEFEITINKTIEKTYTISSTDSTMTYTDVSGGLYKIS
| CsfD | **                      **                      **
| CsfE | ------------------------NGQTITLTLP--VRATITADSILRDTILVKPLSSLYD
| CsfD | DQYSDANVNRNYGNHQPGLLLRNNSTVINIMKGVNLGHTFTVQGYANSAVVPNPQKYF
| CsfE | VVTWDS-------ENNRFKNHEFLVRVIKET-AVPISFEVINNDQYTCSYNK-------PDRMS
| CsfD | TVRSNNGCSSVSSAYLGNASYTLYEIRSSNDVTRNCSSQHQUYTHPMPNSQQVNVTGIYR
| CsfE | PLPTDIAIANSDYKYSVWSGQGYVDMG-------------------------------
| CsfD | DFYLDIGRLQSDAEYRKAPPDITGTFAGEVLKNVQDSGTYPTYNTKGETKKPYES
| CsfE | ------------------------KGRAATVNDSHWLSVVNGVDR--
| CsfD | VTLPTVDNIFDTRTIQGQNLVIFVINGHFTPYNSSLQISLNGFLQSENVSSA
| CsfE | YLDLTLINFPDMTYPQOLLNRGGLCRGSITMLLSNKL------------------
| CsfD | TIPYSLMTIERSRSLATNGGGLNVTINNLESDGYSIOGRFNAFLDKNTAVGTGDY
| CsfE | ------------------------
| CsfD | ADTLTAIFQISLL
important colonisation factor on porcine ETEC strains, contains two chaperones with similar functions (Edwards et al., 1996). FasB is the periplasmic chaperone for the major pilin subunit FasA and FasC similarly chaperones the 987P adhesin FasG (Edwards et al., 1996).
Chapter 8

General Discussion

8.1 Introduction

The CS5 pilus is an important colonisation factor in the pathogenesis of human ETEC infection. CS5 is a component of CFA/IV-expressing ETEC strains, which also includes the antigenically distinct CS6 antigen and the CS4 pilus (Thomas et al., 1985). CFA/IV expressing strains have been shown to be the most commonly encountered group of colonisation factors in field studies from Bangladesh, central Africa and Egypt (McConnell et al., 1991; Peruski et al., 1999; Qadri et al., 2000a). Despite the obvious importance of CFA/IV in world-wide ETEC infections, very limited molecular studies on CS4, CS6 and in particular CS5 have been carried out.

The DNA region required for biogenesis of CS5 pili has been cloned (Heuzenroeder et al., 1989), and the sequence of the major subunit determined (Clark et al., 1992). The major subunit has a predicted size for the mature protein of 18.6 kDa, although estimates from SDS-PAGE indicated the subunit had a size of 23 kDa (Clark et al., 1992).

At the commencement of this work, very little was known at the molecular level regarding the genes involved in the biogenesis of the CS5 pilus and indeed how the final structure was assembled. There were also contradictory studies on the morphology of the CS5 pilus itself. Initially, the CS5 pilus was reported to be a rigid structure of approximately 6-7 nm (Thomas et al., 1985), which was later clarified as a semi-rigid fibrillar structure (Heuzenroeder et al., 1989). There was also one report which indicated that the CS5 pilus is 5-6 nm in diameter and consists of two fine fibrils arranged in a double helical structure (Knutton et al., 1989).
The work presented in this thesis is specifically involved in addressing these unknowns in order to add to the current body of knowledge regarding pilus assembly mechanisms in pathogenic bacteria. This may eventually lead to a rational basis for the design of improved vaccines against ETEC-mediated infections.

8.2 The DNA Region Encoding for Biosynthesis of CS5 Pili

Nucleotide sequencing of the region required for biosynthesis of CS5 pili revealed a further five *csf* (*coli-surface-factor-five*) genes downstream of the previously described major subunit gene, now denoted *csfA*. The *csfC* gene was predicted to encode an outer membrane assembly protein by virtue of its homology to the CooC and CfaC assembly proteins from CS1 and CFA/I pili, respectively (section 3.2.4.1)(Froehlich et al., 1994; Jordi et al., 1992b). The uniqueness of the CS5 encoding region was evident in the lack of significant homology to sequences lodged into the databases for the remaining genes *csfB, csfE, csfF* and *csfD*. However, *csfB* was very similar to *csfF*, sharing both significant protein identity along with notable structural homology (section 3.2.4.2). Indeed, alignment of these two proteins with the PapD chaperone from Pap (Lindberg et al., 1989) showed at least some conservation among the ten invariant residues associated with the >26 members of the PapD immunoglobulin-like chaperone family (Hung et al., 1996), although overall CsfB and CsfF were very dissimilar to PapD (section 3.2.4.2).

At the protein sequence level, CsfA was similar to CsfD (section 3.2.4.3) and mirrored the relative percentage identity between major and minor pilins in other systems such as CS1, Pap or type I pili (Girardeau and Bertin, 1995; Sakellaris and Scott, 1998). CsfE was not significantly homologous to the other proteins involved in CS5 biogenesis or any sequences submitted to the databases as assessed by BLAST or advanced BLAST searches at NCBI.

Expression studies utilising L-[³⁵S]-methionine showed that each of the *csf* genes were able to encode a detectable protein product (section 3.2.6.2). Furthermore, all of the *csf* genes possessed a characteristic signal peptidase I (Sec)-dependent signal sequence cleavage site,
allowing these proteins to be translocated into the periplasm of the bacteria (von Heijne, 1985).

The organisation of the csf genes indicated the region was likely to constitute an operon. The csfB and csfC genes as well as csfF and csfD overlap, which is suggestive of translational coupling (McCarthy and Gualerzi, 1990). Furthermore, two sets of potential $\sigma^{70}$ consensus promoter sequences were identified upstream of csfA, but no such sequences were identified preceding the other csf genes. Primer extension analysis of the csfA transcript would determine the exact transcriptional start site, while similar analyses on the other csf genes may establish whether they are expressed from a mono-cistronic message, indicative of an internal promoter, or whether they are expressed as part of a much larger poly-cistronic message originating upstream of csfA. Northern analysis would also identify the size and number of RNA transcripts.

As with other ETEC pilus systems, the csf gene cluster was flanked by a number of IS elements (IS1 and IS30) or IS-related sequences including IS911, IS66 and orfB from Tn21 (section 3.2.2). The csf gene cluster is one of the first identified human ETEC pilus gene clusters to be flanked by an intact IS element, namely IS1. A complete IS2 element is predicted to lie downstream of the coo gene cluster from CS1 pili, along with sequences related to IS150 and IS629 (Froehlich et al., 1994). The CFA/I, CS2 and CS3 operons are flanked by sequences related to a number of different IS elements including IS1, IS2, IS3, IS91 and IS150 (Froehlich et al., 1995; Hamers et al., 1989; Jalajakumari et al., 1989; Jordi et al., 1992b).

The complete IS1 element upstream from csfA requires further investigation. It is unlikely that the IS1 transposase is able to mediate transposition of the csf cluster from any of the inverted repeats (IRs) from the downstream IS elements IS911 and IS30, since they belong to different IS families (Mahillon and Chandler, 1998). However, it is possible that further downstream of the characterised sequence, a second IS1 or IS1-like sequence exists thereby producing a complex transposon which may then permit the entire region to undergo transposition.
A defective IS30 element was also identified downstream of the csf gene cluster. The five nucleotide insertion into the reading frame of the IS30 transposase, which is predicted to result in the premature termination of translation at the stop codon TGA, may be overcome by a -1 ribosomal frame-shifting event shortly before this stop codon. Ribosomal frame-shifting is important in the expression of at least two IS element transposases, namely IS911 (Polard et al., 1991; Prère et al., 1990) and ISI (Sekine and Ohtsubo, 1989) and may serve a regulatory function to limit the synthesis of functional transposase and consequently transposition. However, the motifs required to promote ribosomal frame-shifting in E. coli (5'-A-AAA-AAC/G-3') are absent from this region (Weiss et al., 1990; Weiss et al., 1989).

The presence of IS or IS-like elements flanking the csf gene cluster indicates that this region may have originated from another organism via transposition. In support of this, the % G+C content of the cluster is only 37.1% versus 50.8 % for E. coli (Blattner et al., 1997). Furthermore, the known rare or modulator codons of E. coli are used at a much greater frequency in the translation of the csf genes when compared to genes from E. coli itself (Table 3.2).

8.3 Regulation

The regulation of the csf region remains largely unknown. CS5 pili expression in the wild type strain PE423 can be greatly enhanced when bile salts are added to the CFA agar. Conversely, the presence of NaCl to the growth medium or growth at 30°C abolishes expression, suggesting CS5 production is subject to thermo-regulation like other ETEC pili. In contrast, expression of CS5 pili from the recombinant clone in E. coli K-12 was not enhanced by bile salts and pilus production was greatly reduced, although not abolished at 30°C. As with the wild type, the presence of NaCl abolished pili expression in the recombinant bacteria. So how is expression of CS5 pili regulated? Expression of CS1 and CS2 pili is dependent on the plasmid encoded regulatory gene, rms (Caron et al., 1989) which is a member of the AraC family of regulators and highly homologous to cfaD/cfaR from CFA/I.
Both of these regulators are not encoded on contiguous regions of DNA with the structural genes for synthesis and assembly of pili. The DNA binding protein H-NS is thought to negatively regulate expression of CS1 by binding to a silencer region of transcription, which is offset by Rns (Murphree et al., 1997).

The csvR gene is proposed as a plasmid-encoded regulatory gene controlling CS5 pilus synthesis. It was originally cloned from E. coli O167:H5, a strain which produces both CS5 pili and CS6 (Willshaw et al., 1991). The csvR gene is 87% homologous with cfaD/cfaR from CFA/I and is able to functionally replace cfaD/cfaR to mediate CFA/I expression (de Haan et al., 1991). However, this study failed to determine whether csvR is involved in the expression of CS5 pili, probably because the investigators did not possess a suitable CS5 pilus expression system. Therefore, the link between csvR and CS5 pilus expression remains highly speculative.

An IS1 element was also identified 1 kb downstream of the csvR gene (Willshaw et al., 1991). Given this information, and knowing that upstream of the csf cluster is a complete IS1 element, an attempt was made in this study to link the csvR gene to the csf gene cluster by PCR. No linkage was obtained in either the wild type strain, or a cosmid clone containing unknown DNA upstream of IS1, suggesting that the observed csvR gene and IS1 element may be located elsewhere on the virulence plasmid, or possibly on an entirely separate and distinct virulence plasmid altogether (data not shown).

The csvR gene was also PCR-amplified, cloned and sequenced from PE423 and found to contain a carboxy-terminal truncation of 34 amino acids when compared to the published sequence (Willshaw et al., 1991). Although it is unknown whether this apparent truncation affects the function of the CsvR protein, it may indicate that the csvR gene is defective in PE423. The expression of CsvR from either low or medium copy number plasmids did not alter CS5 pilus expression (data not shown). Future studies would aim to distinguish between the wild type CsvR protein and the truncated protein by their activity on CS5-related promoter sequences linked to a reporter gene such as lacZ (β-galactosidase). Such studies would then allow a better
understanding of the interaction between CsvR and CS5 pilus expression, or establish whether an interaction exists at all.

8.4 Gene Function

A summary of the function of each of the csf genes characterised is given in Figure 8.1. The following sections describe in more detail the role of these proteins, based on the evidence obtained during the course of this work.

8.4.1 The Major Pilin Subunit Csfa

Csfa has been previously shown to encode a 23 kDa protein by SDS-PAGE, with sequence homology to the F41 subunit from porcine ETEC (Clark et al., 1992). The bulk of the CS5 pilus structure is composed of Csfa subunits. The properties of the Csfa protein suggest it is very hydrophobic, since the mean hydrophobicity index for this protein is 0.32 and CS5 pili are readily precipitated in less than 5% ammonium sulphate (Heuzenroeder et al., 1989).

Only very limited work was carried out on the Csfa major subunit. A specific deletion in the csfa gene was constructed and the remaining genes downstream of csfa were expressed from the lac promoter to determine the effects on pilus biogenesis of this deletion (section 4.2.6). In the absence of Csfa, no cell surface pili was identified on E. coli K-12 harbouring the remaining genes, and no MRHA activity was conferred. This clearly identified the role of the major subunit in producing a surface-located pilus structure with adhesive properties.

8.4.2 The Minor Pilin Subunit Csfd

A large carboxy-terminal deletion in the csfd gene, which is the last gene constituting the csf cluster, resulted in the complete abolition of cell surface CS5 pilus assembly (section 4.2.3.1). Furthermore, there were no detectable Csfa subunits identified in the supernatants of the csfd deletion strain, suggesting that Csfa remains sequestered in the periplasm and is not translocated across the outer membrane. Therefore, Csfd is essential in the initiation of CS5
Figure 8.1 Schematic representation of the csf gene cluster and the proposed function of each of the six genes.

The csf genes are indicated by boxes and the proposed function of the encoded protein shown. The csfC gene encodes the outer membrane usher (chapter 6), csfD encodes an integral minor pilin subunit necessary for pilus initiation (chapter 4), csfE encodes a minor pilin subunit that regulates the length of the pilus (chapter 5), csfB was shown to encode a periplasmic chaperone specific for the CsfA major subunit (chapter 7), and csfF is predicted to encode a periplasmic chaperone specific for CsfE and CsfD (chapter 7). The designation of the major pilin subunit for csfA has been reported previously (Clark et al., 1992). A single arrow indicates the direction of transcription, with the proposed stem-loop attenuator structure also shown between csfA and csfB (Clark et al., 1992).
MAJOR PILIN SUBUNIT

OUTER MEMBRANE USHER

CHAIN LENGTH REGULATOR

MINOR PILIN SUBUNIT

CHAPERONE FOR CHAIN LENGTH REGULATOR AND MINOR PILIN SUBUNIT

CHAPERONE FOR MAJOR SUBUNIT
pilus biogenesis. CsfD did not show any characteristics of a chaperone, since the \(csfD\) deletion strain did not result in a significant alteration in CsfA subunit accumulation in the periplasm (section 4.2.3.2). Specific studies also showed that CsfD is an integral pilus protein, and may constitute up to 2% of the total pilus structure (section 4.2.3.5).

Sequence alignment between CsfD and the major subunit CsfA showed stretches of conserved or identical residues. The similarity of the two pilins may explain why the association of CsfD within the pilus is quite high, due to the stability of the interaction between CsfA and CsfD. In contrast, the CsfE minor pilin is more similar to CsfD than CsfA and probably reflects the fact that these two pilins are likely to be chaperoned by the same protein, CsfF (chapter 7). The characterised CsfA, CsfD and CsfE pilins of CS5 lack many of the characteristics of the Pap pilins (Lindberg et al., 1986; Soto et al., 1998). CsfA lacks the penultimate tyrosine residue, the disulphide bond and the invariant glycine residue, while CsfD and CsfE have two cysteine residues capable of disulphide bond formation, but lack all other characteristics. These findings are similar to the proteins involved in the production of CS1 pili. The major pilin subunit, CooA, was found to lack the penultimate tyrosine residue, the disulphide bond and the invariant glycine residue while the minor pilin CooD has two cysteine residues, but also lacks all other characteristics (Sakellaris et al., 1996).

Attempts at localising CsfD within the CS5 pilus by immunogold electron microscopy failed (section 4.2.4.1). Therefore, high resolution electron microscopy was carried out on rotary shadowed pili. This served two purposes, firstly to ascertain the morphology of the pilus to a high degree of certainty and secondly to determine whether the CS5 pilus was a composite structure, composed of a 2 nm diameter fibrillar tip, not unlike the Pap or type I pilus (Jones et al., 1995; Kuehn et al., 1992).

The results showed that the entire CS5 pilus was a thin, flexible fibrillar structure of approximately 2 nm in diameter (section 4.2.5). Two forms were identified, a tightly packed helical form and a dominant slightly unravelled open helical form. The open form probably results from the mechanical stress in the preparation of the pili for electron microscopy.
Significant understanding of pilus unravelling in Pap suggests that these pili are formed by the tight winding a much thinner structure, which, when mechanically stressed, results in an extended conformation with an open helical structure (Bullitt and Makowski, 1995). Type I pili also exhibit the same phenomenon in the presence of glycerol (Abraham et al., 1992).

In the majority of cases, no distinct tip regions were identified on the CS5 pilus; however, several pili did appear to contain a thinner tip region rather like a small grappling hook which may contain CsfD, but such pili appeared short and fractured. This probably implies the observed tip region is a pilus fracture point. The preferred location for CsfD is at the distal end of the pilus (since CsfD is required for initiation) and then is incorporated throughout the pilus as a minor component, presumably at more or less regular intervals depending on the stoichiometric ratio between it and CsfA in the periplasm. There are two lines of reasoning for this. Firstly, in other systems such as K88 or K99, where the entire pilus is a thin flexible fibrillar structure approximately 2 nm in diameter, no distinct tip structure has been identified. The minor pilins FanF, FanH and FanG from K99 have been shown to be incorporated at the distal end of the fibrillar structure, but also along the shaft (Simons et al., 1990; Simons et al., 1991). Likewise, the minor pilins FaeF, FaeH and presumably FaeI and FaeJ are located along the K88 fibrillae shaft (Bakker et al., 1992). In K88 and K99, the major subunit (FaeG and FanC, respectively) is the adhesin, thus the role of the minor subunits is limited to initiation or elongation of the fibrillar structure (Jacobs et al., 1987a; Jacobs et al., 1987b).

Secondly, the flexible architecture of CS5 probably allows a direct interaction with cell surface receptors without the need for further tip structures on the distal end. In other systems such as Pap and type I pili, the shaft is a rigid structure of 6 nm. The presence of the flexible 2 nm fibrillae on the distal end gives the adhesin on the tip maximal steric freedom to interact with host receptors (Jones et al., 1995; Kuehn et al., 1992).

Is CsfD or CsfA the adhesin for CS5 pili? Mutagenesis studies alone cannot answer this question, since a deletion in each gene abolishes pilus expression. It is tempting to speculate that CsfA is the adhesin, based on studies from the morphologically similar K88 and K99
However, antiserum against CsfD was shown to inhibit MRHA of human erythrocytes by CS5 pili to a dilution of 1:32, although the steric effects of the antibody binding to the pilus and masking the actual adhesin (CsfA) cannot be ruled out. Site-directed mutagenesis on amino acid residues essential for receptor binding but not assembly may ultimately answer this question. Mutation of the positively charged amino acids Lys-132 and Arg-136 of the FanC major subunit of K99 resulted in surface assembled fibrillae which had completely lost adhesive capability. This indicated that these positive amino acids interact directly with the negatively charged sialic acid residue of the receptor molecules (Jacobs et al., 1987b). Similarly, in K88 the substitution of Phe-150 with Ser-150 in the FaeG major subunit resulted in a dramatic decrease in the adhesive capacity of the fibrillae. This also indicates this residue plays an essential role in the interaction of the FaeG adhesin with receptor molecules present on eukaryotic cells (Jacobs et al., 1987a). Studies by Sakellaris et al. (1999) on CS1 and CFA/I showed conclusively that the minor pilin from each pilus type (CooD and CfaE, respectively) is responsible for receptor binding, by site-directed mutagenesis on arginyl residues present in both major and minor pilin subunits. Such mutagenesis did not alter surface assembly of the pili, but did result in the abolition of receptor binding only when the minor pilins were altered (Sakellaris et al., 1999a).

If CsfD is not the adhesin, then what is its role in biogenesis? Clearly, CsfD plays an indispensable role in the initiation of biosynthesis of CS5 pili. The presence of CsfD in the CS5 fibrillar structure may also be important for the conformation and observed flexibility of these structures, thereby permitting efficient receptor binding by the major subunit, if indeed this subunit is the adhesin. However, this may also lead to fragile fracture-sensitive points along the fibrillar structure. This aspect has been investigated in type I pili. It was shown that type I pili devoid of the minor pilins FimF, FimG and FimH, therefore only containing the FimA major subunit, were far more rigid than the corresponding wild type pilus (Krogfelt and Klemm, 1988). Furthermore, the apparent strength of the pure FimA pili would also explain that the
incorporation of minor pilins leads to an increase in fragile points along the pili (Krogfelt and Klemm, 1988).

Conversely, if CsfD is the adhesin, then the incorporation of CsfD at the tip and along the pilus shaft may give the pilus the flexibility to bind host cell receptors, but also lead to an increase of breakage sites at the point of CsfD incorporation. This would create new active binding sites, allowing for a greater number of adhesin-receptor complexes and also bring the bacteria much closer to the eukaryotic cell. Studies on the receptor binding capacity of fragmented type I pili have highlighted the possible importance of broken pili and specific receptor interaction (Ponniah et al., 1991). The fragmentation of type I pili by sequential freeze-thawing resulted in an increase in the receptor binding properties of the pili, which indicated the presence of cryptic adhesin molecules present along the shaft of the pilus. The adhesins are buried within the structure and are non-functional under normal circumstances (Ponniah et al., 1991).

8.4.3 The Minor Pilin Subunit CsfE

A specific in-frame deletion was constructed in the cseE gene, and this deletion resulted in pili which were approximately three times longer than the pili observed in the wild type (section 5.2.3.1). On this basis alone, the data suggested that CsfE plays a specific role in determining pilus length, despite the fact that CsfE bears no significant homology to any other pilin proteins involved in length regulation. Indeed, when antiserum was raised against over-expressed CsfE, this protein was found to localise in the outer membrane of the cells, and did not appear to be part of the final pilus structure (section 5.2.7). Clearly CsfE is not the sole protein involved in pilus length determination. The evidence to support this observation was the fact that trans complementation of the cseE deletion strain with an inducible wild type copy of the gene, did not shorten the mean length of the pili below that observed in the wild type. Furthermore, when the same expression system was carried out in the wild type strain, the pili were not significantly shorter (section 5.2.4.3). Therefore, CsfE is not rate-limiting in
controlling pilus length. The relative amount of CsfF, the predicted chaperone for CsfE, was found to dictate the length of the pilus structure (section 7.2.4).

The difficulty in this study was determining a link between CsfE and the anchoring of CS5 pili to the cell surface. In other systems such as Pap, the PapH protein was shown to regulate the length of the pilus, and was predicted to anchor the pilus to the cell surface, since an increased amount of pili was detected in the supernatant of a papH mutant (Båga et al., 1987). Although this may be the case, the increased shearing forces on longer pili must be taken into account when harvesting the bacteria. The longer a pilus becomes, then presumably a corresponding increase in shearing forces on the pilus would result. Other analogous proteins to PapH involved in length regulation include FimG from type I pili (Russell and Orndorff, 1992), MrpB from MR/P pili of Proteus mirabilis (Li and Mobley, 1998) and CstG of CS3 pili (Jalajakumari et al., 1989). Length regulation in K88 and K99 differs from other systems, since no specific terminator of assembly has yet been identified for each. It is thought that the FanH (K99) and FaeF/FaeH (K88) minor pilins are directly responsible for elongation of the fibrillar structure, and this is responsible for regulating the length (Bakker et al., 1992; Simons et al., 1991).

In the csfE mutant, there was far fewer cell-associated pili compared to the wild type, however there was not a corresponding increase in pili detected the cell supernatants of this strain. This is likely due to the hydrophobic nature of CS5 pili. It was found that the longer detached pili tended to re-associate with surface pili such that when bacterial samples were centrifuged, a significant percentage of these pili become cell-associated. This was shown by a noticeable increase in CsfA and CsfD detected in whole cells of a csfE mutant (section 5.2.3.2). The reduction in the cell surface numbers of CS5 pili in a csfE deletion mutant may therefore be the result of increased shearing forces on the pili, or that CsfE is required to anchor the pilus to the bacterial envelope.

One suitable experiment to ascertain whether CsfE truly anchors CS5 pili to the cell surface would be to significantly increase the periplasmic accumulation of the CsfA major
subunit in the wild type strain, such that the rate of accumulation of CsfA into the pilus would increase. Presumably this would also increase the length of the pilus. If the length of the resulting pilus is approximately the same as that observed in the csfE deletion mutant, then this would allow for a much more direct comparison on the role of csfE for pilus anchoring. By virtue of their similar lengths, both types of pili would be susceptible to the same amount of shearing force, and therefore any increase in the supernatant accumulation of pili in a csfE mutant would be due to a decrease in cell-associated pili.

It is hypothesised that the incorporation of CsfE with the outer membrane usher CsfC results in a stable binding complex, which cannot be displaced by further chaperone-subunit complexes, thereby leading to termination of assembly. This event may also provide a stable platform for association of the pilus with the cell envelope. The low level of expression of CsfE in the cell suggests that the polymerisation of CsfE with other pilins would be a rare event.

8.4.4 The Usher CsfC

Initially, based on sequence data alone, the csfC gene was predicated to encode an outer membrane protein for two reasons. Firstly, secondary structure analysis using PHDSec (Rost et al., 1994) predicted a high β-sheet and low α-helical content for this protein, along with a high percentage of charged amino acids, without long regions of hydrophobic amino acids. Such features are commonly identified in other known outer membrane proteins (Nikaido and Vaara, 1985). Secondly, the CsfC protein sequence showed significant homology to the known outer membrane assembly protein CooC from CS1 pili (Froehlich et al., 1994), and some weak homology with the corresponding proteins CfaC and CotC from CFA/I and CS2 pili, respectively (section 3.2.4.1)(Froehlich et al., 1995; Jordi et al., 1992b).

The csfC gene was specifically shown to encode an outer membrane protein by radio-labelling the translated protein using the T7 polymerase expression system and fractionating the bacterial cells into the individual components (section 6.2.4). Repeated attempts at over-expressing CsfC for producing antiserum failed. This is not surprising, due to the toxic effect on
the cells of over-expressing outer membrane proteins. For instance, attempts at over-expressing the PapC usher resulted in deleterious effects on the bacterial cells (Norgren et al., 1987).

Deletion mutagenesis on the csfC gene showed that it was absolutely required for cell surface expression of CS5 pili. In the absence of CsfC, no pilin subunits were released across the outer membrane (section 6.2.3.1). Furthermore, CsfC was not required to stabilise pilin subunits in the periplasm; therefore CsfC does not act as a specific chaperone during biogenesis (section 6.2.3.2). Hence, CsfC expressly acts as the sole means for the translocation of pilin subunits across the outer membrane for cell surface localisation.

The specificity of CsfC for the CS5 pilins was revealed from complementation studies on the csfC deletion mutant in which the genes encoding for cooC and cfaC were introduced in trans. Both CooC and CfaC were unable to translocate or assemble the CS5 pilin subunits CsfA or CsfD (section 6.2.5). This showed that the interaction between CsfC and outer membrane-targeted chaperone-subunit complexes consisting of CsfA-CsfB, CsfD-CsfF or CsfE-CsfF (chapter 7) is a specific one, based on the structural and/or sequence characteristics of CsfC.

The data obtained from gene deletion experiments involving csfA and csfD also indicate that CsfC may play an usher role analogous to PapC from Pap. It has been shown that the PapC protein is not a porin-like molecule forming a channel in the membrane through which pilus subunits are passed, as first proposed (Norgren et al., 1987). Instead, PapC plays an active role in determining the order of pilus subunit passage, since major subunit-chaperone (PapA-PapD) complexes are not thought to be targeted to empty PapC sites (Dodson et al., 1993). Instead, these complexes have affinity for the PapC site already occupied by a tip fibrillum (consisting of PapG, PapF and PapE) terminated with PapK (Jacob-Dubuisson et al., 1993).

This may explain why in the absence of CsfD, no CsfA is translocated across the outer membrane via CsfC and assembled into pilus structures, because a specific interaction between CsfD and CsfC in the outer membrane is first required. This was further supported from studies in which a specific deletion in the csfA gene was constructed. In this case, CsfD minor pilin subunits were readily detected in the supernatant of the csfA deletion strain, although no pili
structures were identified on the cell surface (section 4.2.6). Similar studies examining the translocation of 987P pilin subunits through the outer membrane indicated that the assembly protein FasD also plays an indispensable role in determining the order of subunit passage (Cao et al., 1995; Schifferli and Alrutz, 1994). Gene deletion experiments showed that the FasG adhesin is exported and assembled first, followed by a second minor pilin FasF and then the major subunit FasA (Cao et al., 1995). It is thought that FasF acts as a linker molecule joining the adhesin to the structural subunit at the distal end of the pilus. An association between the CooD minor pilin and the CooC usher is a pre-requisite for assembly of the CS1 pilus, since a deletion in cooD prevents release of the CooA major subunit, while a cooA deletion does not inhibit the translocation of CooD across the outer membrane (Sakellaris et al., 1999b).

It is hypothesised that the specific binding of CsfD-CsfF complexes to CsfC in the outer membrane results in an altered conformation of CsfC, thereby rendering it in an assembly-competent state for delivery of CsfA-CsfB complexes or further CsfD-CsfF complexes. Such changes in the usher upon binding of adhesin-chaperone complexes in Pap and type I pili have been shown experimentally (Saulino et al., 1998), and this, along with the altered binding specificities of the various chaperone-subunit complexes to the usher, allows for the ordered assembly of the pilus structure to occur.

8.4.5 The Major and Minor Chaperones CsfB and CsfF

Based on sequence and structural similarities, the CsfB and CsfF proteins were predicted to have similar functions in CS5 pilus biogenesis (section 3.2.4.2). Indeed, experiments carried out showed that CsfB and CsfF were likely to function as major subunit-specific and minor subunit-specific chaperones, respectively. Furthermore, these chaperones did not show any discernible sequence or structural homology with the prototypic member of the immunoglobulin-like chaperones, PapD. However, several invariant residues associated with this family were aligned with residues present in CsfB and CsfF (section 3.2.4.2). Only the 987P-pilus system from porcine ETEC has been shown to consist of two chaperones with
separate specificity towards major and minor pilin subunits (Edwards et al., 1996). The FasB chaperone has specific binding and stabilising capability for the FasA major subunit only, while the FasC chaperone has specific binding and stabilising activity for the distally-located adhesin FasG. A chaperone-like protein required for export, FasE, was required for optimal export of FasG and possibly the other Fas pilins FasF and FasA (Edwards et al., 1996).

An in-frame deletion in the \textit{csfB} gene resulted in the complete absence of periplasmic accumulation of CsfA, but did not significantly affect the minor subunit CsfD (section 7.2.7.2). Therefore, strains harbouring the introduced deletion in the \textit{csfB} gene did not produce detectable cell surface pili. However, CsfD was still detected in the supernatants of these strains, which again provided further evidence that CsfD is the first pilin assembled on the cell surface and that CsfB does not chaperone CsfD (section 7.2.7.1). Expression and fractionation studies using radio-labelled CsfB showed that this protein was present exclusively in the periplasm (section 7.2.8). The generation of antiserum against CsfB was also able to detect CsfB exclusively in the periplasm, and CsfB could not be detected in purified CS5 pili preparations (section 7.2.10).

When an in-frame deletion in the \textit{csfF} gene was constructed, the resulting phenotype was almost identical to the in-frame deletion constructed from \textit{csfE}, in that the pili were over three times longer than the wild type (section 7.2.3.1). Specific induction of expression of the CsfF protein from a plasmid introduced in \textit{trans} in the \textit{csfF} mutant was able to reduce the length of the pilus below wild type levels. It was also shown to abolish cell surface pilus expression when the same plasmid was introduced into the wild type strains (section 7.2.4). This implied that CsfF also regulates the length of the pilus and that unlike CsfE, this protein is rate-limiting for controlling pilus length. Given the similarity of CsfF with CsfB, and that CsfF acts to chaperone CsfD (see below), CsfF is predicted to specifically chaperone CsfE in the periplasm, and deliver this protein to the outer membrane to terminate pilus assembly. Attempts at co-purifying CsfE-CsfF complexes from the periplasm using anti-CsfE antiserum failed. This is probably a reflection of the antiserum being unable to recognise native CsfE bound to CsfF. Another approach in which CsfF was (His)$_6$-tagged at the carboxy terminus also was unable to isolate
periplasmic CsfE-CsfF complexes across nickel-NTA agarose. This suggested the (His)₆-tagging at the carboxy terminus may effect subunit recognition domains. Future experiments could be aimed at generating specific antiserum against native CsfF in the periplasm, and using this antiserum to immunoprecipitate CsfE-CsfF complexes, to verify the implied function of CsfF. This would also serve to confirm a strict periplasmic location for CsfF.

The cell surface- or supernatant-associated pili from the csff mutant contained no detectable levels of the CsfD minor subunit (section 7.2.3.2). Interestingly, the periplasmic accumulation of CsfD remained unchanged in the csff mutant (section 7.2.3.2). Based on these results, it was inferred that CsfF might deliver CsfD to the outer membrane for assembly, but that it is not necessary for its stability. However, clearly some CsfD is assembled, given that pilus initiation cannot occur without CsfD.

To explain this apparent discrepancy in the data, a double in-frame deletion was constructed in both csfB and csff to determine whether, in the absence of CsfF, CsfB is able to stabilise CsfD in the periplasm (section 7.2.11). The double deletion resulted in the absence of CsfA and almost complete degradation of CsfD in the periplasm of the strain harbouring this construct. A breakdown product of CsfD also accumulated (section 7.2.13). Moreover, when the double deletion mutant was trans-complemented with csfB alone, only CsfA accumulation was restored and conversely when csff was re-introduced, only CsfD in the periplasm was restored (section 7.2.15). This clearly indicated that CsfF is responsible for stabilising CsfD in the periplasm. So why is CsfD detected in the periplasm, but not the final pilus structure in absence of CsfF only? The presence of CsfD in the periplasm in a csff mutant can be explained in terms of the relative copy number of the plasmid system used and altered binding specificity of CsfD in the absence of CsfF. In the absence of CsfF, CsfD may weakly interact with CsfB alone or CsfB-CsfA complexes in the periplasm. It is unclear how pilus initiation occurs in this case, but it is possible that a CsfB-CsfA-CsfD or indeed CsfD aggregates alone may be able to associate with CsfC in the outer membrane, resulting in subunit translocation and the targeting of further complexes to the usher. Clearly complexes of the CsfA-CsfB type
are favoured for assembly in this case, since CsfD is not detected. When the double deletion was complemented with csfB and csfF on low copy number plasmids in trans, the decrease in these proteins relative to the subunits originating from the high copy number plasmids could lead to the chaperones only binding the pilins for which they have the greatest affinity. This may explain why only CsfA accumulates when CsfB is provided in trans and likewise CsfD predominates when CsfF is provided. Since no CsfA is retained in a single csfB mutant, it is unlikely that CsfF shows any binding specificity for CsfA to help stabilise this protein in the periplasm.

The effects of a csfF deletion may therefore be considered two-fold. Firstly, there is a decrease in the incorporation of CsfD into the pilus fibre, as the efficiency of delivery is drastically reduced, and secondly the CsfE protein is either degraded or remains undelivered to the outer membrane resulting in the observed extended pilus morphologies.

### 8.5 A Comparison Between CS5 and Other E. coli Pilus Operons

The best-characterised E. coli pilus operons of K88, K99, 987P, type I and Pap contain eight to eleven genes, which encode for both regulatory proteins and proteins required for the assembly of the pilus structure. In contrast, the known operons encoding for colonisation factors of human ETEC strains are far simpler in their structure. The operons of CFA/I, CS1 and CS2 consists of only four genes, which encode for the major subunit, the periplasmic chaperone, the usher and the minor subunit which functions as the adhesin in the case of CFA/I and CS1 (Froehlich et al., 1994; Froehlich et al., 1995; Jordi et al., 1992b; Sakellaris et al., 1999a). The CS6 operon consists of four genes with a similar structure except two major subunits are co-expressed on the cell surface with no minor subunit present (Cassels and Wolf, 1995; Gaastra and Svennerholm, 1996). Likewise, the CS14 operon has a relatively simple structure, consisting of a chaperone, an usher and two genes encoding for co-expressed major subunits, not unlike CS6 (Gaastra and Svennerholm, 1996; McConnell et al., 1989a). The CS3 operon contains four contiguously located genes encoding for the major subunit,
minor subunit, chaperone and usher, but contained within the usher is four further predicted ORFs with unknown function (Jalajakumari et al., 1989).

The csf gene cluster of CS5 is more complicated than the other known human ETEC pilus operons, consisting of six contiguous genes with each possessing an indispensable function in the biogenesis of the mature pilus. An alignment of the csf genes with several of the well characterised pilus systems including Pap, type I, K88, K99, 987P and CS1 shows that the spatial arrangement of the csf genes is also quite different (Figure 8.2). The arrangement of the CsfC usher and CsfB chaperone resembles the type I system rather than the arrangement of the chaperones and ushers in the Pap, K88 or K99 systems. Perhaps the most striking feature of the csf cluster is its relationship to that of 987P from porcine ETEC (Schifferli et al., 1991a). Both systems utilise two chaperones, one for minor subunits and one for major subunits, and have a similar genetic arrangement (Edwards et al., 1996). Indeed, if the fasH regulatory gene is not considered, both contain the same number of genes encoding for structural and assembly proteins (Figure 8.2). Furthermore, a large stem-loop structure ($\Delta G = -21.1$) is located between the fasA major subunit gene and the fasB chaperone gene (Edwards et al., 1996). This is identical in organisational terms to the stem-loop structure located between csfA and csfB, which is predicted to act as an attenuator sequence to reduce expressing of downstream genes (Clark et al., 1992).

### 8.6 A Model for Biosynthesis of CS5 Pili

Based on the results obtained in this study, a proposed model for biogenesis of CS5 pili is presented in Figure 8.3. Since each of the proteins has a characteristic N-terminal signal sequence recognition site, the nascently translated Csf proteins are translocated across the inner membrane and into the periplasm via the Sec pathway. Upon translocation of the pilin subunits into the periplasm, they are bound and stabilised by the periplasmic chaperone proteins CsfB and CsfF. CsfB is responsible for binding and stabilising the major subunit CsfA, while CsfF functions to stabilise CsfD and is predicted to bind and stabilise CsfE. Presumably, in the
Figure 8.2 Genetic organisation of the csf gene cluster compared to other well known E. coli pili operons.

The genes are indicated by single letter designations within the boxes and correspond to the genetic nomenclature given for each operon. The different shading patterns in the boxes indicate the function of the polypeptide encoded by the gene. Arrows indicate the direction of transcription of each pilus operon. The gene which regulates CS1 pilus expression, ms, is not located on a contiguous region of DNA and is not shown.
Figure 8.3 Model of CS5 pilus assembly.

All of the Csf proteins are translocated across the inner membrane via the Sec pathway. The major pilin subunit CsfA is then bound by the CsfB chaperone and the minor pilin subunits CsfD and CsfE are bound by the CsfF chaperone. This serves to protect the pilin subunits from mis-folding, premature aggregation or degradation by periplasmic proteases. Pilus biogenesis is initiated by CsfD-CsfF complexes binding to the outer membrane assembly protein CsfC, which results in the translocation of CsfD across the outer membrane, and leaving CsfC with an altered conformation, in an assembly-competent state. Pilus elongation occurs with multiple CsfA-CsfB interactions with CsfC in the outer membrane, driving the assembly of CsfA into the growing pilus, along with further CsfD-CsfF interactions. The rate of incorporation of CsfA versus CsfD is thought to depend on the stoichiometric ratio of the two pilins in the periplasm. CsfD is thought to add flexibility to the CS5 structure (magnified region). Pilus termination occurs when CsfE-CsfF complexes are targeted to CsfC, which is predicted to result in the irreversible association of CsfE with CsfC, thereby preventing further polymerisation of the pilus.
absence of CsfB or CsfF the subunits either remain sequestered in the cytoplasmic membrane and are removed by the cell, or the subunits mis-fold and/or aggregate resulting in their proteolytic degradation by the DegP protease. The ability of chaperones to facilitate import of the subunits into the periplasm, and the role of DegP, was established in studies involving the PapD chaperone from the Pap system (Jones et al., 1997).

Pilus biogenesis is initiated by the delivery of CsfD-CsfF complexes to the outer membrane protein CsfC. This results in the translocation of CsfD subunits across the outer membrane and leaves the CsfC protein in an assembly-competent state so that it is now able to recognise CsfA-CsfB complexes, but also further CsfD-CsfF complexes. Evidence to support this selectivity of CsfC for chaperone-subunit complexes includes the absence of detectable CsfA subunits in the supernatant of a csfD mutant, and conversely detectable CsfD subunits in the supernatant of a csfA mutant strain. The decision to incorporate CsfD instead of CsfA is likely to be a consequence of the stoichiometric ratio between the two subunits in the periplasm. Since CsfA is expressed at much higher levels in the periplasm relative to CsfD, this explains why the bulk of the pilus is composed of CsfA, with only 2% attributable to CsfD. The presence of CsfD along the shaft may serve to add flexibility to the fibrillar structure of CSS, but at the same time increase sensitivity to fracturing.

Termination of pilus assembly occurs by the incorporation of CsfE with CsfC and the polymerising pilus in the outer membrane via delivery of CsfE by its chaperone CsfF. The CsfE-CsfF complexes are predicted to be at very low concentrations in the periplasm compared to the other chaperone-subunit complexes described. The association of CsfE with CsfC in the outer membrane is predicted to prevent further polymerisation of the pilus structure, by virtue of the association between CsfC and CsfE in the outer membrane preventing CsfC from translocating further subunits. It is hypothesised that the incorporation of CsfE results in a stabilised pilus structure on the cell surface, thereby acting as an anchor for the pilus to the outer membrane.
8.7 Concluding Remarks

This study has sought to provide an insight into the biogenesis of the important colonisation factor CS5. In particular, this study focussed on identifying the genes involved in biogenesis, the function of the genes, and how the mature CS5 pilus structure is formed on the cell surface as a result of the different interactions between proteins encoded by the genes. Furthermore, this study has also established the natural morphology of the CS5 pilus, which was identified as a thin flexible fibrillar structure approximately 2 nm in diameter.

The CS5 pilus-encoding region is unlike any other human ETEC pilus system so far identified. In this respect, CS5 pili may be representative of a distinct class of pili since: (1) four of the six csf genes lack any significant homology to proteins involved in pilus biogenesis from other systems and (2) the identification of two chaperones, one for major subunits and one for minor subunits, has not been previously described in any human ETEC pilus system. It is tempting to speculate an animal origin for CS5, given the major subunit is homologous to the corresponding subunit in F41 pili, the two chaperone system is similar to that utilised by 987P pili, and finally CS5 is morphologically similar to K88 and K99.

The fundamental knowledge gained from this study may be further utilised in the determination of the specific adhesin (CsfA or CsfD) responsible for binding to cell surface receptors on the brush border membranes of the intestinal epithelium. This may aid in the development of a more suitable oral vaccine against ETEC infection, which incorporates the adhesin of CS5 pili.
References


Marron M. B. and Smyth C. J. (1995) Molecular analysis of the *cso* operon of enterotoxigenic Escherichia coli reveals that CsoA is the adhesin of CS1 fimbriae and that the accessory genes are interchangeable with those of the *cfa* operon. *Microbiology* 141, 2849-2859.


APPENDIX I

Nucleotide sequence of the CS5 region from Enterotoxigenic Escherichia coli O115:H40.

Open reading frames (ORFs) are indicated by the gene name followed by >, termination codons are denoted by *. ORFs thought to belong exclusively involved in CS5 pili biosynthesis are denoted as csf (coli-surface-factor-five). A potential set of promoters is indicated with the consensus promoter sequences in bold preceding csfA. Nucleotides 1377-2324 have been previously sequenced, along with the identification of the proposed stem-loop structure preceding csfB (Clark et al., 1992; Accession no. X63411). The position of potential ribosome binding site (rbs) sequences are indicated in bold above the overlined sequence. Solid triangles indicate the predicted signal sequence cleavage site of the newly synthesised protein product. Boxed sequences indicate inverted repeats (IR) for each of the IS elements with left and right ends denoted by an L and R. For IS1, highlighted amino acids represent conservative substitutions from the published sequence (Ohtsubo and Ohtsubo, 1978). A 5 nt insertion into the IS30 transposase sequence (5' GCTCA 3') is indicated in bold. Relevant restriction enzyme sites are also indicated. The DNA sequence presented has been deposited into the EMBL/Genbank databases under the Accession no. AJ224079.


Corrigenda

Abstract
Page 2, line 11: “fibrillae” is amended to read “fibrillum”

Abbreviations
DTT- “dietiothreitol” is amended to read “dithiothreitol”
LA- “luria agar” is amended to read “Luria agar”
SAP- “shrimp alkaline phosphotase” is amended to read “shrimp alkaline phosphatase”

Chapter 1
Page 2, line 9: “and can” is changed to “that can”
Page 6, 2nd bottom line: “plasmid-encoding” is changed to “plasmid encoding”
Page 8, line 20-21: remove comma following mutant
Page 11, line 4: should read “is mediated”
Page 11, line 9: “consists” is amended to read “possesses”
Page 15, line 5: “found on pigs” is amended to read “found on piglet ETEC”
Page 17, line 17: insert comma after FanH
Page 22, line 5: delete “in CS1”
Page 23, line 9: delete “was”
Page 23, line 11: “but also stabilise” is changed to “and also to stabilise”
Page 26, line 8: “against” is changed to “from”
Page 26, line 18: “Moylan” to “Boylan”
Page 29, line 21: “Production of monoclonal antibodies” is amended to read “Monoclonal antibodies”
Page 29, line 23: “identified in” is amended to read “found in”
Page 31, line 14: “identifying” is amended to read “identified”
Page 32, line 10: “is” is amended to read “to be”
Page 33, line 14: “antigenic homology” is amended to read “antigenic determinants”
Page 35, 3rd bottom line: “servoar” is amended to read “serovar”
Page 38, line 1: “sialoglycoconjugates” is amended to read “sialoglycoconjugates”
Page 38, line 4: “sulphhydryal” is amended to read “sulphydryl”
Page 42, line 18: “O2” is amended to read “oxygen”
Page 42, 4th bottom line: “Gala(1-4)Gal” is amended to read “Galα(1-4)Gal”
Page 42, 2nd bottom line: “globopent-aosylceramide” is amended to read “globopent-aosylceramide”
Page 42, 2nd bottom line: “Forsman” is amended to read “Forsmann”

Page 43, line 6: “retains” is amended to read “retain”

Page 44, line 10: “did not effect” is amended to read “did not affect”

Page 45, line 17: insert comma after “interactions”

Page 46, line 11-12: should read “Periplasmic pilin intermediates complexed to PapD have been examined (Striker et al., 1994).”

Page 48, line 7: “DsbA dependent” is amended to read “DsbA-dependent”

Page 49, line 17: “aggregation in the periplasm” is amended to read “aggregation in the periplasm”

Chapter 2

Table 2.1: all gene letter and allele numbers should be italicised in E. coli K-12 strains used

Page 73, 3rd bottom line: “media” is amended to read “medium”

Page 73, 2nd bottom line: “media” is amended to read “medium”

Page 73, line 10-11: should read Morona et al. (1995)

Page 73, line 12: remove comma after “rotor”

Page 75, line 19: “Antisera” is amended to read “Antiserum”

Page 76, line 3: “Antiserum” is amended to read “Antiserum”

Chapter 3

Page 81, line 4: remove comma after “subunit”

Page 81, line 9: replace “Enterotoxigenic” with “enterotoxigenic”

Page 81, line 16: “phosphotase” should read “phosphatase”

Figure 3.3 “homology” is replaced with “of DNA sequence identity”

Figure 3.3, line 5: “Tn21” should read “Tn21”

Page 83, line 3: “Tn21” should read “Tn21”

Table 3.1, footnote e: remove comma after “af”

Page 85, line 8: “of csfC initiating,” is amended to read “of csfC, initiating”

Page 85, line 14: “the cluster termed csfF,” should read “the cluster termed, csfF”

Page 92, 3rd bottom line: “Tn21” should read “Tn21”
Page 93, line 10: “located” is amended to read “be located”

Chapter 4

Figure 4.5, last line: insert comma after “granules”
Figure 4.5, last line: “restored” is amended to read “complemented in trans”
Page 101, line 8: “neutralised across” is changed to “with immobilised”
Page 102, line 12: “that there is” is amended to read “that there are”
Page 102, line 16: “presented is” is amended to read “presented are”
Page 102, line 17: “was examined” is amended to “were examined”
Page 103, line 7: “CS5 pili is” is amended to “CS5 pili are”
Figure 4.13: insert “a restriction site for which” between “ApaI” and “which”
Figure 4.13, line 5: “and cloned” is amended to “and the modified csf region cloned”
Page 104, line 1: “CS5 gene sequence” is amended to read “CS5 operon sequence”
Page 104, line 5: “Antiserum” is amended to read “Antiserum”

Chapter 5

Page 107, line 17: “cell surface, however,” is amended to read “cell surface; however,”
Page 110, line 3: “This data” is amended to “These data”
Page 110, line 5: “delete the word “of”
Figure 5.4, last line: “is restored” is amended to read “is complemented in trans”
Page 111, line 3: “in csfE effects” is amended to read “in csfE affects”
Figure 5.8, line 3: replace “pPM5657” with “pPM5684”
Page 114, last line: replace “suggests” with “suggest”
Page 116, line 2: replace “neutralised across” with “and with immobilised inclusion”

Chapter 6

Page 122, line 6: “effects” is replaced with “affects”
Figure 6.4, line 6: “is restored” is amended to read “is complemented in trans”
Page 126, line 6: “effect” is replaced with “effect”
Page 126, 5th bottom line: insert comma after “mutant”
Page 127, line 4: “are predicted” is replaced with “is predicted”

Chapter 7

Figure 7.1, line 2: replace “is” to “for which”
Figure 7.3, last line: “is restored” is amended to read “is complemented in trans”
Page 132, 4th bottom line: “This data” is amended to read “These data”
Page 134, line 13: “was evident” is amended to read “were evident”
Page 134, line 21: “the csfF gene, thus” is amended to read “the csfF gene; thus,”
Page 134, 2nd bottom line: exonuclease III, therefore” should read “exonuclease III; therefore,”
Figure 7.12, last line: “is restored” is amended to read “is complemented in trans”
Page 137, line 14: “This data” is replaced with “These data”
Page 139, line 4: “imidazole, therefore, it was” is replaced with “imidazole; therefore, it was”
Page 139, line 8: “neutralised across” is amended to read “with”
Figure 7.22, line 6: replace “visibly” with “visible”

Chapter 8
Page 147, 3rd last line: “the csf gene were” is amended to read “the csf genes was”
Page 149, line 6: insert comma after “Ohtsubo, 1989)”
Page 149, 2nd last line: “CS2 pili is” is amended to read “CS2 pili are”
Page 153, line 18: “is the adhesin, thus” is amended to read “is the adhesin; thus,”
Page 154, last line: “explain that” is amended to “explain how”
Page 156, line 17: “there was” is amended to read “there were”
Page 159, last line: replace “consist” with “possess”
Page 163, line 1: replace “is” with “are”

References
Page 170, Collier et al.: remove full stops from journal title
Page 170, Dalrymple et al.: “Embo J” is replaced with “EMBO J”
Page 179, Jacobs et al., 1987b: “EMBO J” is amended to read “EMBO J”
Page 180, Jones et al., 1990: remove hyphens from abbreviated journal title
Page 182, Levine et al., 1986: “Molecular” is amended to read “Molecular”
Page 191, Sakellaris et al., 1999b: “J Bact” is amended to read “J Bacteriol”