Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.
Adhesion and invasion by *Escherichia coli* K1

A dissertation presented for the degree of Doctor of Philosophy, in the Faculty of Science, Trinity College Dublin

by

Robert Fagan

2006

Department of Microbiology,
Moyne Institute of Preventive Medicine,
Trinity College Dublin
I, Robert Fagan, declare that the work presented herein represents my own work, except where duly acknowledged in the text, and has not been previously presented for a higher degree at this or any other University. I agree that the Library of Trinity College Dublin may lend or copy this thesis upon request.

Robert Fagan
Summary

*Escherichia coli* is one of the principal causes of septicaemia and meningitis in neonates. Sepsis occurs in approximately one in every thousand live births, with up to 10% progressing to develop meningitis. A disproportionate number of these infections are caused by strains expressing the K1 capsule; up to 40% of septicaemia isolates and 80% of meningitis isolates. To date, most studies have focused on immune evasion and bacterial penetration of the blood-brain barrier. More than 15 *E. coli* proteins have been shown to contribute to these events, including: OmpA, YijP, IbeA and IbeB. However it is still unclear how the bacteria first colonise and then invade the epithelial surfaces prior to septicaemia.

A gene was identified and cloned from *E. coli* K1 strain RS218 which, when over-expressed in *E. coli* K-12, promotes adhesion to and invasion of cultured CHO and T84 epithelial cells. The protein encoded by this gene is 90% identical to the Hra1 protein from porcine *E. coli* strain O9:H10:K99 and is also highly homologous (63% identical) to Tia, an adhesin/invasin from ETEC. This protein, designated Hek (Hra1 from *E. coli* K1), is a 26 kDa outer-membrane protein and is predicted to form a β-barrel in the outer membrane consisting of 8 anti-parallel, amphipathic β-sheets with four short periplasmic turns and four longer external loops. In addition to an invasive phenotype, expression of Hek in an *E. coli* K-12 strain also confers a heat-resistant haemagglutination phenotype, similar to that observed with Hra1, and an auto-aggregation phenotype, similar to that reported for the Ag43, FimH and Cah proteins. Structure function analysis of the Hek protein indicates that a single surface exposed loop is critical for all three of these phenotypes. The invasion of CHO cells by Hek-expressing *E. coli* K-12 can also be inhibited in a dose-dependent manner with purified heparin but not with other glycosaminoglycans indicating that the receptor for the Hek protein on mammalian cell surfaces is likely to
be a heparinated proteoglycan. Expression of the Hek protein in the clinical isolate RS218 is variable and appears to be controlled at a transcriptional level by the global regulator protein IHF.

The expression of Hek is variable and this variability is independent of the native hek gene promoter. A region which is involved in the regulation of Hek expression was identified upstream of the hek gene on the E. coli K1 RS218 chromosome.

Mutants of E. coli K1 strain RS218 defective in production of two well characterised virulence factors, OmpA and IbeB, were also found to be less invasive in human colonic epithelial cell line. These mutants were subjected to proteomic analysis by 2D PAGE.

The Hek protein is involved in direct interactions between bacterial cells and is sufficient to promote invasion of cultured human epithelial cells. Taken together these data suggest a role for the Hek protein in the colonisation and invasion of the human gastrointestinal epithelium. Invasion of human cells often involves multiple bacterial factors but to date few proteins have been implicated in the initial stages of infection by E. coli K1. This study has identified possible roles for the Hek, OmpA and IbeB proteins.
Acknowledgements

First and foremost I'd like to thank Stephen, for great advice, truly terrible jokes and too many other things to list here. He took me on despite having witnessed my fourth year project and that says it all. Special thanks also to Dr. Yuri Riträcht for his vital contribution to my fourth year project.

Thanks especially to everyone in the PHI lab, the last few years would have been a lot harder without you all - Matthew, for lots of new phrases and innumerable entertaining desktops, Vivienne, for keeping me up to date on all the TV I never seem to get around to watching, and Gwen, for always offering to help with the plate counts! Thanks to everyone in the Moyne, current and past members alike. I'm not going to list you all here but to everyone I chatted with over coffee or a pint, thanks, it's been a great few years and I'll miss you all when I leave. To Tadhg, Libby, Jane and Paul for those nights in Mahaffey's when I needed to get of the lab and Judy, for listening to my rants when things went badly and reminding me that I'd finish eventually. Thanks to all the weekend regulars who made Saturdays in the lab less depressing. To Arlene and Davida who started with me four years ago, thanks for everything and good luck in the new jobs. To everyone in the CJD lab, you've been a constant source of advice and ideas since day one, thanks. A final big thanks to everyone in the prep room, especially Paddy, Joe and Ronan.

To my family, I couldn't have asked for more support or help than you've given me for the last eight years. Brian, I promise I'm finished now and I'll go get a real job, well sort of anyway. Andrew, you were always there with advice when I needed it, thanks. Jane, thanks for being my big sis and all those entertaining arguments. Finally, to my parents, I wouldn't be writing this if it wasn't for you. You've always supported me in everything I've done. Thank you.

And finally to Mercedes, you've been there for me since the start and I couldn't have made it this far without you. Your unconditional love and support were always there no matter what else was going on in my life. Thank you most of all.
Table of Contents

Title page I
Declarations II
Summary III
Acknowledgements V
Table of contents VI
Abbreviations XIII
List of Figures XIV
List of Tables XVI

Chapter 1 Escherichia coli: Mechanisms of survival and virulence 1
1.1 Escherichia coli the pathogen 2
1.1.1 Classification of Escherichia coli pathotypes 2
1.1.2 The intestinal E. coli pathotypes 4
1.1.2.1 Enteropathogenic E. coli - EPEC 4
1.1.2.2 Enterohaemorrhagic E. coli - EHEC 6
1.1.2.3 Enterotoxigenic E. coli - ETEC 8
1.1.2.4 Diffusely-adherent E. coli - DAEC 9
1.1.2.5 Enteroinvasive E. coli - EIIEC 10
1.1.2.6 Enteroinvasive E. coli - EIEC 11
1.1.3 The extraintestinal E. coli pathotypes 12
1.1.3.1 Uropathogenic E. coli - UPEC 12
1.1.3.2 Neonatal-meningitic E. coli - NMEC 14
1.2 Translocation, membrane insertion and structure of β-barrel proteins in E. coli 19
1.2.1 Mechanisms of inner membrane translocation in E. coli 19
1.2.2 The Sec translocation pathway 19
1.2.3 Periplasmic transit and insertion into the outer membrane 20
1.2.4 β-barrel protein structure 20
1.3 Scope of this thesis 23
<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Materials and Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>General Methods</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Bacterial strains and culture conditions</td>
</tr>
<tr>
<td>2.1.1.1</td>
<td>Bacterial strains</td>
</tr>
<tr>
<td>2.1.1.2</td>
<td>Bacterial growth media</td>
</tr>
<tr>
<td>2.1.1.3</td>
<td>Bacterial culture conditions</td>
</tr>
<tr>
<td>2.1.1.4</td>
<td>Antibiotics and media additions</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Eukaryotic cell lines and growth conditions</td>
</tr>
<tr>
<td>2.1.2.1</td>
<td>Eukaryotic cell lines</td>
</tr>
<tr>
<td>2.1.2.2</td>
<td>Cell line growth conditions</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Plasmids, bacteriophage and oligonucleotides</td>
</tr>
<tr>
<td>2.1.3.1</td>
<td>Plasmids</td>
</tr>
<tr>
<td>2.1.3.2</td>
<td>Oligonucleotides</td>
</tr>
<tr>
<td>2.2</td>
<td>Nucleic acid methodologies</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Transformation of <em>E. coli</em> strains</td>
</tr>
<tr>
<td>2.2.1.1</td>
<td>Transformation of <em>E. coli</em> K-12 strains using calcium chloride method</td>
</tr>
<tr>
<td>2.2.1.2</td>
<td>Transformation of <em>E. coli</em> strains by electroporation</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Transduction with bacteriophage P1vir <em>dam rev6</em></td>
</tr>
<tr>
<td>2.2.2.1</td>
<td>Preparation of a P1vir lysate</td>
</tr>
<tr>
<td>2.2.2.2</td>
<td>P1vir transduction</td>
</tr>
<tr>
<td>2.2.3</td>
<td>λ Red-mediated allele replacement</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>Generation of mutated alleles in vitro</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>λ Red-mediated allele replacement in <em>E. coli</em> K1</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Purification of plasmid and chromosomal DNA</td>
</tr>
<tr>
<td>2.2.4.1</td>
<td>Small-scale purification of plasmid DNA</td>
</tr>
<tr>
<td>2.2.4.2</td>
<td>Large-scale purification of plasmid DNA</td>
</tr>
<tr>
<td>2.2.4.3</td>
<td>Purification of total genomic DNA</td>
</tr>
<tr>
<td>2.2.5</td>
<td>In vitro manipulations of DNA</td>
</tr>
<tr>
<td>2.2.5.1</td>
<td>Restriction endonuclease digestion of DNA</td>
</tr>
<tr>
<td>2.2.5.2</td>
<td>Purification of DNA fragments</td>
</tr>
<tr>
<td>2.2.5.3</td>
<td>Ligation of DNA fragments</td>
</tr>
<tr>
<td>2.2.5.4</td>
<td>Agarose gel electrophoresis</td>
</tr>
</tbody>
</table>

VII
2.2.5.5 Purification of DNA samples from solution by ethanol precipitation

2.2.6 Polymerase chain reaction

2.2.6.1 Amplification of DNA by polymerase chain reaction

2.3 Analysis and Manipulation of Proteins

2.3.1 SDS-PAGE

2.3.1.1 Preparation of total cellular protein for SDS-PAGE analysis

2.3.1.2 Sarkosyl enrichment of outer-membrane proteins

2.3.1.3 Electrophoresis of protein samples

2.3.2 Production of an anti-Hek antiserum and purification of an MBP-Hek fusion protein

2.3.2.1 Hek antiserum

2.3.2.2 Pilot purification of MBP-Hek

2.3.2.3 Large-scale purification of MBP-Hek

2.3.3 Western immunoblotting

2.3.3.1 Electro-transfer of separated proteins

2.3.3.2 Detection of bound proteins

2.3.4 Flow-cytometry

2.3.5 2-D PAGE

2.3.5.1 Sample preparation

2.3.5.2 IPG strip loading

2.3.5.3 First dimension isoelectric focussing

2.3.5.4 Equilibration and second dimension SDS-PAGE

2.3.6 Enzyme assays for beta-galactosidase and alkaline phosphatase

2.4 Phenotypic assays

2.4.1 Haemagglutination assay

2.4.2 Autoaggregation assays

2.4.2.1 Quantification of the rate of autoaggregation

2.4.2.2 Competitive autoaggregation assays

2.4.3 Adhesion, cell association and invasion assays

2.4.3.1 Qualitative adhesion assays

2.4.3.2 Quantitative cell association and invasion assays
<table>
<thead>
<tr>
<th>Chapter 3</th>
<th>Phenotypic analysis of the Hek protein of Escherichia coli K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.1.1</td>
<td>E. coli K1 pathogenesis</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Bacterial adhesins</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Bacterial invasion</td>
</tr>
<tr>
<td>3.2</td>
<td>Results</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Identification of the hek gene in E. coli K1 strain RS218</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Cloning and expression of hek in E. coli K-12</td>
</tr>
<tr>
<td>3.2.2.1</td>
<td>Expression of hek from its native promoter</td>
</tr>
<tr>
<td>3.2.2.2</td>
<td>Construction of a low-copy Hek-expression plasmid</td>
</tr>
<tr>
<td>3.2.2.3</td>
<td>Construction of the IPTG-inducible pThek6 plasmid</td>
</tr>
<tr>
<td>3.2.2.4</td>
<td>Construction of a malE-hek fusion expression plasmid and</td>
</tr>
<tr>
<td></td>
<td>purification of the MBP-Hek fusion protein</td>
</tr>
<tr>
<td>3.2.3</td>
<td>E. coli K-12 expressing Hek interact with human erythrocytes</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Expression of Hek in E. coli K-12 promotes autoaggregation</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Qualitative adhesion assays on Hek-expressing E. coli using T84 colonic epithelial cells</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Quantitative cell-association and invasion assays on Hek-expressing E. coli</td>
</tr>
<tr>
<td>3.2.6.1</td>
<td>Cloning of the Tia invasin from ETEC</td>
</tr>
<tr>
<td>3.2.6.2</td>
<td>Interaction of Hek- or Tia-expressing E. coli with CHO-K1</td>
</tr>
<tr>
<td></td>
<td>epithelial cells</td>
</tr>
<tr>
<td>3.2.6.3</td>
<td>Invasion of human T84 cells by E. coli K-12 expressing Hek or Tia</td>
</tr>
<tr>
<td>3.2.6.4</td>
<td>Construction of an E. coli K1 hek mutant</td>
</tr>
<tr>
<td>3.2.6.5</td>
<td>Invasion of human T84 cells by an E. coli K1 hek mutant</td>
</tr>
<tr>
<td>3.2.6.6</td>
<td>Complementation of the hek mutation in strain RShek</td>
</tr>
<tr>
<td>3.3</td>
<td>Discussion</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Identification of the Hek-receptor and structure-function analysis of the Hek protein of <em>Escherichia coli</em> K1</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Proteoglycans</td>
</tr>
<tr>
<td>4.2</td>
<td>Results</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Identification of the Hek receptor</td>
</tr>
<tr>
<td>4.2.1.1</td>
<td>Invasion of proteoglycan deficient CHO-K1 cells by <em>E. coli</em> K-12 expressing Hek or Tia</td>
</tr>
<tr>
<td>4.2.1.2</td>
<td>Inhibition of invasion of CHO-K1 cells with purified glycosaminoglycans</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Prediction of the structure of the Hek protein</td>
</tr>
<tr>
<td>4.2.2.1</td>
<td>Analysis of the amino-terminal sequence of the mature Hek protein</td>
</tr>
<tr>
<td>4.2.2.2</td>
<td>Prediction of the secondary structure of the Hek protein</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Structure-function analysis of the Hek protein</td>
</tr>
<tr>
<td>4.2.3.1</td>
<td>Construction of <em>hek</em> loop-deletion mutants</td>
</tr>
<tr>
<td>4.2.3.2</td>
<td>Expression of Hek loop-deletion mutants</td>
</tr>
<tr>
<td>4.2.3.3</td>
<td>Haemagglutination and autoaggregation phenotypes of Hek loop-deletion mutants</td>
</tr>
<tr>
<td>4.2.3.4</td>
<td>Interaction of Hek loop-deletion mutants with mammalian epithelial cells</td>
</tr>
<tr>
<td>4.3</td>
<td>Discussion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5</th>
<th>Regulation of <em>hek</em> expression in <em>Escherichia coli</em> K1</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Phase variation</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Integration host factor</td>
</tr>
<tr>
<td>5.2</td>
<td>Results</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Plasmid nomenclature</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Analysis of Hek protein expression in <em>E. coli</em> K-12 and K1</td>
</tr>
<tr>
<td>5.2.2.1</td>
<td>Expression of Hek from its native promoter in high-copy</td>
</tr>
</tbody>
</table>
number

5.2.2.2 Expression of Hek from the inducible \( P_{ac} \) promoter 90
5.2.2.3 Expression of Hek from its native promoter in low-copy number 90
5.2.2.4 Cloning of the \( hek \) regulatory region 91
5.2.2.5 Analysis of Hek expression from plasmids containing the extended upstream region 93
5.2.2.6 Correlation between the level of Hek expression and Hek biological activity 94
5.2.3 Analysis of Hek translation and export using a PhoA reporter fusion 94
5.2.3.1 Construction of high- and low-copy \( hek-phoA \) fusion plasmids 94
5.2.3.2 Expression of the Hek-PhoA fusion protein 95
5.2.3.3 Construction of an inducible \( hek-phoA \) fusion plasmid and expression in \( E. coli \) K1 96
5.2.3.4 Addition of regulatory region upstream of \( hek-phoA \) fusions 97
5.2.3.5 Quantitative analysis of PhoA activity from low copy constructs 97
5.2.4 Analysis of \( hek \) transcription using a \( lacZ \) transcriptional reporter 98
5.2.4.1 Construction of \( lacZ \) transcriptional fusion to the putative \( hek \) promoter 98
5.2.4.2 Transcriptional analysis of the putative \( hek \) promoter in \( E. coli \) K-12 98
5.2.4.3 Construction of a Lac- \( E. coli \) K1 RS218 derivative strain 99
5.2.4.4 Transcriptional analysis of the putative \( hek \) promoter in \( E. coli \) K1 strain RS\( \lambda \)ac 100
5.2.4.5 Cloning of the \( hek \) regulatory region upstream of a promoterless-\( lacZ \) in plasmid pQF50 101
5.2.4.6 Transcriptional analysis of the \( hek \) promoter in \( E. coli \) K-12 and K1 102
5.2.4.7 Identification of putative regulators of \( hek \) transcription 103
Chapter 6 Proteomic analyses of invasion deficient mutants of E. coli K1

6.1 Introduction
6.1.1 Proteomics
6.1.1.1 2-D PAGE
6.1.1.2 MALDI-TOF

6.2 Results
6.2.1 Construction of E. coli K1 ompA and ibeB mutants
6.2.1.1 Construction of an E. coli RS218 ibeB mutant
6.2.1.2 Invasion of human T84 cells by E. coli K1 RSibeB
6.2.1.3 Construction of an E. coli RS218 ompA mutant
6.2.1.4 Construction of a low copy OmpA-expressing plasmid
6.2.1.5 Invasion of human T84 cells by E. coli K1 RSompA
6.2.2 Proteomic analysis of E. coli K1
6.2.2.1 Proteomic analysis of E. coli K1 RSihf
6.2.2.2 Proteomic analysis of E. coli K1 ompA and ibeB mutants

6.3 Discussion

Chapter 7 General Discussion

References
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D PAGE</td>
<td>two-dimensional polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>CHAPS</td>
<td>3-[(3-cholamidopropyl)dimethylammonio] 1-propanesulfonate</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's modified Eagle's medium</td>
</tr>
<tr>
<td>DMF</td>
<td>dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>N,N-dimethylsulphoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FITC</td>
<td>fluorescein-5-isothiocyanate</td>
</tr>
<tr>
<td>GAG</td>
<td>glycosaminoglycan</td>
</tr>
<tr>
<td>GI</td>
<td>gastrointestinal</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IEF</td>
<td>isoelectric focussing</td>
</tr>
<tr>
<td>IPG</td>
<td>immobilised-pH gradient</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropylthiogalactoside</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>matrix assisted laser desorption ionisation - time of flight</td>
</tr>
<tr>
<td>MBP</td>
<td>maltose-binding protein</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple-cloning site</td>
</tr>
<tr>
<td>MHBP</td>
<td>mammalian heparin binding protein</td>
</tr>
<tr>
<td>ONPG</td>
<td>o-nitrophenyl β-D-galactopyranoside</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PIPES</td>
<td>piperazine-1,4-bis(2-ethanesulfonic acid)</td>
</tr>
<tr>
<td>pNPP</td>
<td>o-nitrophenyl phosphate</td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene difluoride</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecylsulphate - polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TE</td>
<td>tris-EDTA buffer</td>
</tr>
<tr>
<td>TBE</td>
<td>tris-borate-EDTA buffer</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-bromo-4-chloro-3-indoyl-β-D-galactoside</td>
</tr>
<tr>
<td>X-P</td>
<td>5-bromo-4-chloro-3-indoyl phosphate</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.1</td>
<td>Overview of pathogenesis by the intestinal <em>E. coli</em></td>
</tr>
<tr>
<td>1.2</td>
<td>Overview of pathogenesis by the extraintestinal <em>E. coli</em></td>
</tr>
<tr>
<td>1.3</td>
<td>Export of proteins via the Sec pathway</td>
</tr>
<tr>
<td>1.4</td>
<td>Three-dimensional structure of a β-barrel protein</td>
</tr>
<tr>
<td>3.1</td>
<td>Clustal alignment of the primary sequences of the <em>E. coli</em> Hek, Hral and Tia proteins</td>
</tr>
<tr>
<td>3.2</td>
<td>Construction of the pHEK6 plasmid</td>
</tr>
<tr>
<td>3.3</td>
<td>Expression of the Hek protein in <em>E. coli</em> K-12 from high and low copy plasmids</td>
</tr>
<tr>
<td>3.4</td>
<td>Expression of the Hek protein from an inducible P&lt;sub&gt;lac&lt;/sub&gt; promoter</td>
</tr>
<tr>
<td>3.5</td>
<td>Structure of pMhek1 and expression of an MBP-Hek fusion protein</td>
</tr>
<tr>
<td>3.6</td>
<td>Haemagglutination of <em>E. coli</em> K-12 expressing Hek</td>
</tr>
<tr>
<td>3.7</td>
<td>Hek-promoted autoaggregation</td>
</tr>
<tr>
<td>3.8</td>
<td>Analysis of Hek-promoted autoaggregation in mixed cultures</td>
</tr>
<tr>
<td>3.9</td>
<td>Qualitative analysis of Hek-promoted adherence</td>
</tr>
<tr>
<td>3.10</td>
<td>Interaction of <em>E. coli</em> K-12 expressing Hek or Tia with CHO-K1 cells</td>
</tr>
<tr>
<td>3.11</td>
<td>Interaction bacteria expressing Hek or Tia with human colonic epithelial cells</td>
</tr>
<tr>
<td>3.12</td>
<td>Construction of an <em>E. coli</em> K1 RS218 hek mutant</td>
</tr>
<tr>
<td>4.1</td>
<td>The contribution of GAGs to the invasion of CHO-K1 cells by <em>E. coli</em> K-12 expressing Hek or Tia</td>
</tr>
<tr>
<td>4.2</td>
<td>Inhibition of Hek-promoted invasion with exogenous glycosaminoglycans</td>
</tr>
<tr>
<td>4.3</td>
<td>MALDI-TOF analysis and amino-terminal sequencing of the mature Hek protein</td>
</tr>
<tr>
<td>4.4</td>
<td>Topological model of the Hek protein and loop deletion strategy</td>
</tr>
<tr>
<td>4.5</td>
<td>Expression of Hek loop-deletion mutants</td>
</tr>
<tr>
<td>4.6</td>
<td>Haemagglutination and autoaggregation by bacteria expressing Hek or loop-deletion derivatives</td>
</tr>
<tr>
<td>4.7</td>
<td>Interaction of Hek loop deletion mutants with CHO-K1 cells</td>
</tr>
<tr>
<td>5.1</td>
<td>Comparison of Hek protein expression in <em>E. coli</em> K-12 and K1</td>
</tr>
<tr>
<td>5.2</td>
<td>Western immunoblot analysis of Hek protein in <em>E. coli</em> K-12 and K1</td>
</tr>
<tr>
<td>5.3</td>
<td>Construction of the pH+ plasmid</td>
</tr>
<tr>
<td>5.4</td>
<td>Western immunoblot analysis of Hek expression in <em>E. coli</em> K-12 strain XL-1 and K1 strain RS218</td>
</tr>
<tr>
<td>5.5</td>
<td>Phenotypic analysis of <em>E. coli</em> K-12 strains expressing different amounts of Hek protein</td>
</tr>
<tr>
<td>5.6</td>
<td>Construction of hek-phoA fusion reporter plasmids</td>
</tr>
<tr>
<td>5.7</td>
<td>Variable PhoA phenotypes from high and low copy hek-phoA fusion plasmids</td>
</tr>
</tbody>
</table>

XIV
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.8</td>
<td>Variable PhoA phenotypes from an IPTG-inducible <em>hek-phoA</em> fusion with increasing IPTG concentration</td>
<td>96</td>
</tr>
<tr>
<td>5.9</td>
<td>Analysis of transcription from the putative <em>hek</em> promoter</td>
<td>100</td>
</tr>
<tr>
<td>5.10</td>
<td>Construction of disrupted <em>lac</em> operon <em>in vitro</em></td>
<td>100</td>
</tr>
<tr>
<td>5.11</td>
<td>Construction of <em>hek</em> promoter-<em>lacZ</em> transcriptional fusions</td>
<td>102</td>
</tr>
<tr>
<td>5.12</td>
<td>Transcriptional analysis of the putative <em>hek</em> promoter in <em>E. coli</em> K-12 and K1</td>
<td>102</td>
</tr>
<tr>
<td>5.13</td>
<td>IHF and transcription of the <em>hek</em> gene</td>
<td>104</td>
</tr>
<tr>
<td>5.14</td>
<td>Two promoter model of <em>hek</em> transcription and translation</td>
<td>110</td>
</tr>
<tr>
<td>5.11</td>
<td>Construction of disrupted <em>lac</em> operon in vitro</td>
<td>100</td>
</tr>
<tr>
<td>5.12</td>
<td>Transcriptional analysis of the putative <em>hek</em> promoter in <em>E. coli</em> K-12 and K1</td>
<td>102</td>
</tr>
<tr>
<td>5.13</td>
<td>IHF and transcription of the <em>hek</em> gene</td>
<td>104</td>
</tr>
<tr>
<td>5.14</td>
<td>Two promoter model of <em>hek</em> transcription and translation</td>
<td>110</td>
</tr>
<tr>
<td>6.1</td>
<td>Mutation of the <em>E. coli</em> K1 <em>ibeB</em> gene and the interaction between RSibeB and T84 cells</td>
<td>116</td>
</tr>
<tr>
<td>6.2</td>
<td>Mutation of the <em>E. coli</em> K1 <em>ompA</em> gene</td>
<td>118</td>
</tr>
<tr>
<td>6.3</td>
<td>Invasion of human T84 cells by an <em>E. coli</em> K1 <em>ompA</em> mutant</td>
<td>118</td>
</tr>
<tr>
<td>6.4</td>
<td>Proteomic analysis of <em>E. coli</em> RSihf</td>
<td>120</td>
</tr>
<tr>
<td>6.5</td>
<td>Proteomic analysis of <em>E. coli</em> RSibeB</td>
<td>120</td>
</tr>
<tr>
<td>6.6</td>
<td>Proteomic analysis of <em>E. coli</em> RSompA</td>
<td>120</td>
</tr>
<tr>
<td>7.1</td>
<td>Model of GI colonisation and invasion by <em>E. coli</em> K1</td>
<td>132</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>After page:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td><em>E. coli</em> strains used in this study</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>Plasmids and phage used in this study</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Oligonucleotide primers used in this study</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>BLASTN searches of incomplete <em>E. coli</em> RS218 genome sequence</td>
<td>54</td>
</tr>
<tr>
<td>3.2</td>
<td>Quantitative analysis of Hek-promoted autoaggregation in <em>E. coli</em> K-12</td>
<td>58</td>
</tr>
<tr>
<td>5.1</td>
<td>Plasmid nomenclature for Chapter 5</td>
<td>90</td>
</tr>
</tbody>
</table>
Chapter 1  *Escherichia coli*: Mechanisms of survival and virulence
1.1 *Escherichia coli* the pathogen

1.1.1 Classification of *Escherichia coli* pathotypes

*Escherichia coli* is a common commensal and pathogen in both human and animal hosts. The majority of *E. coli* strains are highly adapted to a commensal lifestyle in the mammalian gut but in the quest to exploit new niches and modes of transmission several groups of *E. coli* have developed or acquired abilities allowing them to cause disease in their host. These pathogenic *E. coli* lineages can be roughly classified according to the site of disease into the intestinal and the extraintestinal pathogens. The intestinal pathogenic *E. coli* are those responsible for diarrhoeal disease and the extraintestinal strains cause infections of the urinary tract or sepsis and meningitis in neonates. Diverse *E. coli* strains that cause similar disease symptoms using a selection of common virulence factors are grouped together into pathotypes (123). Not all members of a given pathotype necessarily possess a full set of the same virulence attributes but all are capable of producing disease with similar symptoms in the host. Eight human-specific *E. coli* pathotypes have been described to date on the basis of their virulence profiles and specific disease characteristics: six pathotypes are involved in diarrhoeal disease with a further two pathotypes encompassing the urinary tract and neonatal meningitic strains. The six intestinal pathotypes are enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), diffusely-adherent *E. coli* (DAEC), enteroaggregative *E. coli* (EAEC) and enteroinvasive *E. coli* (EIEC) while the extraintestinal pathogenic strains are grouped as uropathogenic *E. coli* (UPEC) and neonatal-meningitic *E. coli* (NMEC). Several of these pathotypes are also involved in animal disease and a further animal-specific pathotype, avian pathogenic *E. coli* (APEC), that causes extraintestinal infections in poultry has also been described. Based on genomic sequence analysis *Shigella* spp. may also be considered as an *E. coli* pathotype (207), which has completely adapted to its intracellular niche. For the sake of discussion here *Shigella* will be grouped with the EIEC pathotype to which it is most closely related. The mechanisms of virulence of each of the eight human *E. coli* pathotypes will be described in some detail later in this chapter.
Pathogenic *E. coli* may also be distinguished from one another on the basis of their ability to invade host cells. The eight *E. coli* pathotypes can therefore be separated into the mucosal pathogens that interact with the host via adhesins and secreted effectors and the invasive pathogens, which are capable of gaining access to the intracellular environment. The first group includes the EPEC, EHEC, DAEC and EAEC pathotypes and the latter includes the EIEC, NMEC, ETEC and UPEC pathotypes, although the importance of invasion in the virulence of ETEC and UPEC is still open to debate. Conventional thinking holds that the EIEC group includes the only true intracellular pathogenic *E. coli* strains, defined as those capable of intracellular survival and replication. This assumption is challenged by data showing efficient invasion of several human tissue types by strains of the ETEC, UPEC and NMEC groups (30, 74, 113, 159) and intracellular replication of an NMEC strain in professional macrophage (249). The intracellular milieu presents an important new niche for *E. coli*, providing a significant advantage to bacteria in terms of immune evasion but also presenting many new challenges that must be overcome to allow successful infection including the innate cellular defences and relative scarcity of some.

Several distinct lineages of *E. coli* have seemingly independently made the transition from a commensal to pathogenic lifestyle. Each pathogenic lineage has acquired specific virulence traits that have allowed them to exploit new host niches and cause disease in both human and animal hosts. These virulence factors have frequently been acquired on plasmids and other mobile genetic elements such as pathogenicity islands, transposons and bacteriophage. As a result of these acquisitions the genomes of pathogenic *E. coli* strains can be up to 1 Mb larger than non-pathogenic laboratory strains (15). However, the transition from commensal to pathogen has been marked by more than just acquisitions of novel DNA, microarray analysis of pathogenic and commensal strains has also revealed considerable regions of commensal-specific DNA, regions presumably deleted during the evolution of the pathogenic strains (57). Most surprisingly however, was the finding that the different pathogenic lineages differed as much from each other as from commensal strains, for example EHEC strain EDL933 and UPEC strain CFT073 were found to share only 39.2 % of their protein coding sequences (273). These enormous differences have made *E. coli* a highly successful
and diverse pathogen, diverse in terms of both virulence mechanisms and disease phenotypes. Individual *E. coli* isolates have commonly been differentiated using a complex serotyping system based on the dominant lipopolysaccharide (O), flagellar (H) and capsular (K) antigens. There are currently 170 O-, 80 K- and 56 H-antigens recognised, which, in the many possible combinations, allow in excess of 100,000 *E. coli* serotypes (186). The K designation is now often discarded when describing an *E. coli* serotype, for example the commonly known EPEC O157:H7 serotype, except in situations where the K antigen is itself closely associated with disease as in the case of K1 capsular antigen and neonatal disease in humans (see section 1.1.3.2). It is important to note however, that the prototypic and universally used *E. coli* K-12 laboratory strains do not in fact belong to the K12 serotyping group, rather the name refers to the strain collection label which was assigned to the original isolate in the Stanford University culture collection in 1922 (145).

### 1.1.2 The intestinal *E. coli* pathotypes

#### 1.1.2.1 Enteropathogenic *E. coli* - EPEC

EPEC was the first *E. coli* pathotype to be described (26) and is an important cause of diarrhoeal disease in developing countries. Infection with EPEC is characterised by the formation of distinctive micro-colonies on cultured HEp-2 cells (226) and the formation of attaching and effacing (A/E) lesions on intestinal epithelial cells (116) (Fig. 1.1). These A/E lesions are formed following intimate attachment of the bacteria to the epithelial surface and result from the accumulation of polymerised actin beneath the bacteria. The microvilli forming the intestinal brush-border are effaced following actin mobilisation and pedestal-like structures often form directly beneath the attached bacteria (178). All of the bacterial factors required for this distinctive disease phenotype are encoded on the 35.5 kb locus for enterocyte attachment (LEE) pathogenicity island (252). The LEE region bears all the hallmarks of a horizontally acquired element with a GC content of only 38 % (compared with the *E. coli* average of 50 %) (163) and an insertion point associated with the selC tRNA gene, a characteristic of many other pathogenicity islands (22). The LEE island includes some 41 genes and encodes a type III secretion system, a number of secreted effector molecules (Esp
Figure 1.1 Overview of pathogenesis by the intestinal *E. coli*. The interactions between the six intestinal *E. coli* pathotypes and the human GI epithelium are outlined briefly above.
proteins and Tir) and several other virulence-associated proteins (such as Intimin) (252).

Following initial adherence, the LEE-encoded type III secretion system exports a number of effector molecules into the host cell, including the Tir protein. The type III secretion system consists of a multi-protein complex resembling a needle that extends from the bacterial surface. The EspA protein is secreted to the end of the needle and forms a hollow extension that contacts the host cell membrane. EspB and D are then translocated and are predicted to form a pore in the host plasma membrane completing the link between bacterium and host cell. Following translocation, the Tir protein inserts into the host cell's plasma membrane and acts as a receptor for another LEE-encoded protein, the intimin adhesin, on the surface of the loosely attached bacteria. In effect the bacteria produce and export their own receptor (128) and it is this interaction between Tir and intimin that results in intimate attachment of the bacteria. In addition to Tir, the intimin adhesin also appears to recognise a second receptor although the identity of the second receptor is a matter of some debate with the mammalian β1-integrin being the current leading contender (77). The translocated Tir protein is also capable of interacting with the host cytoskeleton and associated signalling molecules and it is these interactions that result in the large rearrangements of the host cell's cytoskeleton. Tir first interacts with Nck, an adaptor protein, which in turn binds to the amino-terminal portion of the Wiskott-Aldrich syndrome protein (N-WASP) and the actin-related protein complex (Arp2/3). It is this Arp2/3 complex that initiates actin nucleation and ultimately results in pedestal formation beneath the bound bacterium. Several other effectors are also translocated into the host cell along with Tir including the EspF, EspG, EspH, EspZ and Map proteins (51). EspF induces apoptosis (41) and promotes migration of the tight-junction protein occludin resulting in a potential breakdown in barrier function across the epithelium (164). The Map protein stimulates filopodia formation in a Cdc42 dependent manner and disrupts mitochondrial membrane potential. EPEC also produce a number of other toxins including EspC which results in increased ion secretion in host cells. EspC is a member of the serine protease autotransporter family, so-called because one portion of the protein forms a β-barrel membrane pore that facilitates export of the toxin domain. The net effect of all these EPEC virulence mechanisms is an induced inflammatory response (following
NF-κB activation via the MAP kinase pathway), increased intestinal permeability due to tight-junction disruption, loss of the absorptive microvilli due to actin mobilisation and increased ion secretion due to toxins such as EspC. All of these, and perhaps other factors, contribute to the severe diarrhoea that results from infection with EPEC.

The bundle-forming pili (BFP) have been implicated in the initial adherence of EPEC to intestinal cells (86) and the associated formation of characteristic micro-colonies on the cell surface. However, EPEC strains lacking the bfp operon are still highly virulent. BLAST searches of the EPEC strain E2348/69 genome sequence revealed 9 regions with homology to putative fimbrial gene clusters and 10 putative non-fimbrial adhesins (254). The importance of these putative adhesins in adhesion to epithelial cells is now an area of considerable interest. The LifA protein has been found to play a role in adhesion in the absence of BFP (9) and EspA, which forms the needle-like connection between the LEE type III secretion system and the host cell, has also been implicated in the initial attachment to cultured epithelial cells (34).

Interestingly the LEE pathogenicity island has been integrated into the bacterial gene regulation circuitry. The genes encoding the LEE type III secretion system, comprising three distinct operons LEE1, 2 and 3, are under the control of the EAF plasmid-encoded Per (plasmid-encoded regulator) protein, a member of the AraC family of transcriptional regulators (90). Only LEE1 is directly regulated by Per, while the other two operons are regulated by Ler (LEE-encoded regulator), a second regulatory protein encoded within LEE1 (166). The Per protein was also found to regulate the expression of the bfp operon (also encoded on the EAF plasmid). In this way EPEC regulate and coordinate the expression of their diverse and often horizontally acquired virulence mechanisms. The expression of Ler is also regulated by the global regulators IHF and Fis (79, 88), further tying LEE-encoded virulence into the existing bacterial regulatory network.

1.1.2.2 Enterohaemorrhagic E. coli - EHEC

The EHEC pathotype was first described following outbreaks of severe gastrointestinal infections associated with bloody diarrhoea (211). Following this initial observation these organisms were subsequently found associated with sporadic cases of haemolytic
uremic syndrome (HUS), a very serious condition involving acute renal failure (125). The cause of these extreme disease phenotypes is a cytotoxin that was first characterised in *Shigella dysenteriae* and designated Shiga toxin (Stx). To date most study has focussed on EHEC of the 0157:H7 serotype, which has caused large outbreaks in many western countries (125, 261). EHEC 0157:H7 appear to have evolved from an EPEC strain (serotype 055) that acquired the genes for these Shiga toxins on a lysogenic bacteriophage (208) and as a result they also contain the LEE pathogenicity island and can induce A/E lesion formation on cultured epithelial cells (Fig. 1.1) (115). In addition to the 0157:H7 serotype several other EHEC serotypes have also emerged as important human pathogens including the 026 and 0111 serotypes (123). Other *E. coli* serotypes have also acquired Stx but do not contain LEE and are generally not associated with human disease. EHEC therefore form a subgroup within the larger group of Shiga toxin-producing *E. coli* (STEC).

The Shiga toxins produced by EHEC strains fall into two distinct groups: Stx1, which is identical to that found in *Shigella dysenteriae*, and Stx2, which is only 55% identical to Stx1 at the amino acid level. Both forms of Stx adopt a basic ABj structure with a single A toxin domain linked to five identical B subunits that target and dock the toxin to the host cell membrane. The holotoxin is then endocytosed and the enzymatically active A subunit cleaves a single adenine residue from the 28S rRNA of the eukaryotic ribosome disabling cellular protein synthesis. In the gastrointestinal tract the additional damage caused by this toxin is responsible for the severe, and often bloody, nature of EHEC-induced diarrhoea. However, the most severe outcome of EHEC infection results when this toxin is transported to the kidneys via the blood stream. The direct action of Stx on the glomerular endothelial cells, clogging the renal vasculature with cell debris, and the induced inflammatory response appear to combine to cause the acute renal failure associated with HUS (150, 151).

In addition to the LEE pathogenicity island and the Stx bacteriophage EHEC 0157:H7 and several other serotypes also harbour a large virulence plasmid p0157 (29), which encodes an RTX toxin, similar to the UPEC haemolysin (section 1.1.3.1), the EspC serine protease and StcE. StcE has been shown to cleave the complement pathway C1
esterase inhibitor, perhaps contributing to tissue damage by removing a key control mechanism in the complement cascade (143).

1.1.2.3 Enterotoxigenic E. coli - ETEC

ETEC are important pathogens of children in developing countries and immunologically naïve tourists causing watery diarrhoea. ETEC strains are characterised by the production of ST and/or LT enterotoxin. ETEC efficiently colonise the mucosa of the small intestine and induce diarrhoea by the production of these enterotoxins (Fig. 1.1). Colonisation is mediated by extensive families of colonisation factor antigens (CFA), coli surface antigens (CS) and putative colonisation factors (PCF) (82). Each of these families encompasses a wide antigenically-distinct family of surface molecules. However, disease is more commonly associated with only a limited number of antigenic types, for example, out of at least 21 distinct colonisation factors, up to 75 % of human pathogenic ETEC strains produce CFA/I, CFA/II or CFA/IV (276).

Following colonisation, enterotoxins are produced resulting in the characteristic diarrhoea. The heat-labile enterotoxin (LT) is closely related to Vibrio cholerae enterotoxin (242) and two forms LT-I and LT-II may be produced by some ETEC strains although only LT-I is associated with human disease. LT-I is an AB5 toxin like Stx of EHEC and the enzymatic A domain has ADP-ribosyltransferase activity, transferring an ADP-ribosyl group from NAD to the α-subunit of the stimulatory G protein resulting in irreversible activation of adenylate cyclase. The resulting increase in intracellular cAMP levels activates cAMP-dependent kinases which in turn activate the cystic fibrosis membrane conductance regulator (CFTR) chloride channel. Chloride ion secretion inhibits NaCl uptake and results in water diffusion into the intestinal lumen and diarrhoea (231). ETEC strains can also/alternatively produce a heat-stable enterotoxin (ST). As with LT, ST also comes in two forms STα and STβ although to date only STα has been linked to human disease (178). STα is a very short 18 - 19 residue peptide with three internal disulphide bridges (281). STα binds to and stimulates the membrane spanning guanylate cyclase resulting in increase cGMP production (167) again resulting in chloride secretion.
Despite the large number of possible colonisation factors described in ETEC strains some virulent strains do not express any identifiable colonisation factor. In addition to this, invasion of cultured epithelial cells has been observed with some ETEC strains, including the prototypic H10407 strain. The search for additional adhesion- or invasion-associated factors revealed two chromosomal loci, \textit{tia} and \textit{tib}, that were sufficient to promote adhesion and invasion in non-pathogenic laboratory strains of \textit{E. coli} (67, 74). Mutations in either locus in the wild-type H10407 strain also resulted in impaired adhesion and invasion. The \textit{tia} gene encodes an approximately 25 kDa outer membrane protein that is predicted to adopt an 8-stranded \(\beta\)-barrel conformation (see section 1.2) in the membrane with four surface-exposed loops. Evidence suggests that the Tia protein mediates attachment to epithelial cells via interactions with the heparan moieties on mammalian cell proteoglycans (73). The heparan-binding activity of Tia was found to reside in the second of the four surface-exposed loops (73, 154), although the evidence presented in these papers is disputed in this thesis (see Chapter 4, section 4.3). The second ETEC invasin, TibA, encoded by \textit{tib} is a large 104 kDa glycoprotein and a member of the autotransporter family of proteins (67, 149). The TibA protein has also been implicated in autoaggregation and biofilm formation (236) although, once again, the importance of these phenotypes to ETEC pathogenesis is not known.

### 1.1.2.4 Diffusely-adherent \textit{E. coli} - DAEC

Very little is known about the diffusely adherent \textit{E. coli} pathotype although it has been associated with disease in older infants (225). The distinguishing feature of this pathotype is a characteristic scattered pattern of adhesion on cultured HEp-2 cells which differentiated these strains from EPEC (forms micro-colonies) or EAEC (forms “stacked-brick” biofilm, section 1.1.2.5). The mechanisms of DAEC virulence have yet to be fully elucidated but one important virulence factor has been identified. The F1845 fimbriae, found in at least 75% of DAEC strains (123), have been implicated in the characteristic adherence pattern (20). These fimbriae are homologous the Dr fimbriae of UPEC (section 1.1.3.1) and were found to induce the formation of long cellular projections that curl around and protect adherent bacteria (Fig. 1.1) (36, 280). The Dr fimbriae in UPEC specifically bind to the decay accelerating factor (DAF), whose normal function is to protect the cell from the complement cascade, and the F1845 fimbriae have also been shown to interact with this receptor (17). It has been proposed
that it is the interaction with DAF and the clustering of this receptor at the point of adhesion which induces signal transduction events involving PI-3 kinase resulting in the formation of the cellular extensions (193). A second adhesin has also been identified in DAEC strains and designated AIDA. AIDA is a plasmid-encoded autotransporter protein comprised of a carboxy-terminal outer membrane β-barrel domain through which an adhesive amino terminal domain, termed AIDA-I, is translocated. The AIDA protein has been shown to mediate diffuse adherence to HeLa cells (14) and has also been implicated in autoaggregation of bacteria and biofilm formation (235). However, it is difficult to see how these two seemingly contradictory activities can be reconciled. Recent evidence also suggests that AIDA may be expressed by only a small minority of DAEC strains (14).

1.1.2.5  Enteroaggregative E. coli - EAEC

EAEC strains are associated with diarrhoea in both children and adults and have been isolated in outbreaks worldwide (180). The pathotype is defined by a characteristic aggregative adherence pattern on cultured HEp-2 cells. EAEC bacteria efficiently adhere both to the cultured cells and to one another resulting in the formation of a distinctive “stacked-brick” biofilm on the cell surface (Fig. 1.1) (179). However, the link between this adhesion phenotype and pathogenesis is tenuous and this definition may encompass both pathogenic and non-pathogenic strains (178). In some EAEC strains the aggregative adherence phenotype has been attributed to one of a number of related fimbrial adhesins designated aggregative adherence fimbriae (AAF) (43). The AAF fimbriae belong to the Dr family of fimbrial adhesins, which are also found in UPEC and DAEC strains. To date, four distinct AFF variants have been described although each AAF is only found in a small minority of strains (180). This suggests that other adhesins are involved in aggregative adherence in some or all EAEC strains and a number of outer membrane proteins have emerged as candidates for further investigation (52, 265).

Adhesion of EAEC to cultured epithelial cells is accompanied by considerable cell damage, including vacuolisation and loss of cytoplasmic contents (1, 177). A number of toxins have been identified in EAEC strains, including EAST1, ShET1, Pic and Pet. The plasmid-encoded EAST1 toxin resembles the prototypic E. coli heat-stable
enterotoxin (STa) from ETEC (223), but is produced by both pathogenic and non-pathogenic strains and has yet to be linked conclusively to human disease (224). EAEC also produce a pair of chromosomally-encoded toxins, encoded on opposite strands of the same locus. One is Shigella entrotoxin 1 (ShET1), which is also found in Shigella flexneri, and the other is Pic, a large autotransporter protease with mucinase activity (98). The final characterised toxin, Pet, is a 108 kDa plasmid-encoded autotransporter protein with enterotoxin activity (99, 181).

Invasion of human epithelial cells by EAEC strains has also been reported (1, 13) but, as with ETEC, the significance of this phenomenon to EAEC virulence is unknown. One striking aspect of EAEC pathogenesis is the extreme duration of diarrhoeal disease. The average duration can be as much as 17 days (19). The resilient biofilms formed by EAEC strains may account for this persistence of infection as biofilms have been shown to be highly resistant to clearing by antibiotic treatment and host immune responses (40, 279). Another surprising aspect of EAEC pathogenesis is the heterogeneity of EAEC strains. Many putative EAEC virulence factors have been described but none is found in all (or even a significant majority) of the grouped EAEC strains. It may be that diverse sets of virulence determinants result in the single pattern of EAEC virulence phenotypes or perhaps a set of common virulence determinants do exist in EAEC but have yet to be identified.

1.1.2.6 Enteroinvasive E. coli - EIEC

Any consideration of the EIEC pathotype must necessarily involve a discussion of the virulence of the Shigella spp. as the two are very closely related (207, 270). In recent times it has become clear that Shigella may be considered as a highly adapted pathogenic grouping of E. coli. EIEC may simply be considered as the less pathogenic extreme in a pathotype encompassing EIEC and the four human pathogenic Shigella spp.: S. boydii, S. dysenteriae, S. flexneri and S. sonnei. The details of EIEC pathogenesis appear to be essentially identical to that of Shigella. EIEC adhere to and invade intestinal mucosa, probably via phagocytic M-cells (218) or disrupted tight-junctions (217), induce apoptosis in professional macrophage (284) and invade epithelial cells via the basolateral membrane (172). Once inside an epithelial cell the bacteria escape from the endocytic vacuole, replicate intracellularly and spread
horizontal to neighbouring cells (Fig. 1.1). The bacteria promote the polymerisation of an actin tail at one pole of the bacterial cell that propels the bacterium through the host cytoplasm to the plasma membrane producing a membrane protrusion into the neighbouring cell. Endocytosis by the neighbouring cell follows, producing a double-membrane vacuole containing the bacteria. The bacteria can also escape from this double-membrane vacuole. The genes required for this complex virulence phenotype are found on a large 213 kb virulence plasmid, pWR100, which is found in EIEC and all Shigella spp. (221). Interestingly, this plasmid appears to be a mosaic of at least four original plasmids and also contains a very high number of insertion sequence elements (28). Invasion is mediated by the pWR100 _mxii/spa_ gene products that form a type III secretion system (160). The type III secretion apparatus translocates a number of effector molecules into the host cell, including the _Ipa_ and _Ipg_ proteins. _IpaB_ and _C_ form the translocation pore in the host cell membrane and _IpaA_, _C_ and _D_ interact with cytoskeleton components and Rho GTPases to allow rearrangement of cellular actin and the formation of membrane extensions surrounding the bacteria (182, 256). The endocytic vacuole is then lysed by _IpaB_ (104) and the 120 kDa membrane protein _IcsA_ promotes actin polymerisation at one pole of the bacterial cell via interactions with N-WASP and Arp2/3 (65). Following endocytosis by a neighbouring cell the double-membrane vacuole is lysed by _IpaB_ and _C_ (188) and the cycle of replication and horizontal spread continues. In this manner the bacteria can effectively avoid exposure to the hosts immune response. Much of the epithelial damage associated with the symptoms of Shigella/EIEC infection appears to be the result of the host’s immune response rather than direct effects of the intracellular bacteria (195).

1.1.3 The extraintestinal _E. coli_ pathotypes

1.1.3.1 Uropathogenic _E. coli_ - UPEC

_E. coli_ is the most common cause of urinary tract infections (UTIs) in humans. The strains that cause these infections, classed as UPEC, form a highly diverse and often clonally distinct group of organisms. UTIs are among the most common infections in western countries, especially in the female population, with up to 60 % of women suffering a UTI during their life-time (76). Infections range in severity from
Figure 1.2 Overview of pathogenesis by the extraintestinal *E. coli*. The invasion of bladder epithelial cells by UPEC is outlined at the top. Below is a brief outline of NMEC pathogenesis from colonisation of the GI epithelium to invasion of the brain microvascular endothelial cells (BMECs) which form the blood-brain barrier. The most important known virulence factors are highlighted.
uncomplicated bladder infections to severe infections of the kidney, which often progress to bacteraemia (237), and often recur in the weeks following an initial infection (75). As a result of the high incidence of UTIs UPEC has been the subject of intensive research for many years. However, the different strains, which together form the UPEC group, are highly heterogeneous and no single set of virulence factors has been identified that typify virulent UPEC (119). Despite this extreme variability among UPEC strains a number of important virulence factors have been identified (117) and in recent years a model of the pathogenesis of UPEC infection has emerged (Fig. 1.2) (25). A number of fimbriae are expressed by UPEC, including type 1, P (Pap), S, M and Dr fimbriae (117), although type 1 and P fimbriae appear to play the most important role. UPEC strains also produce several toxins such as cytotoxic necrotising factor (CNF1), haemolysin and Sat (an autotransporter protease) (7, 71, 93). Infection with UPEC is a multistage process that begins with the colonisation of the normally sterile urinary tract, probably mediated by type 1 and P fimbriae. Colonisation of the bladder epithelium is mediated by type 1 fimbriae, which are expressed by the majority of UPEC strains (142), via adherence to mannose moieties on uroplankin receptors of transitional epithelial cells (173). At least some UPEC strains can then invade the bladder epithelial cells. Invasion is dependent on FimH, the type 1 fimbrial-tip adhesin, and involves complex cellular signal cascades involving focal adhesin kinase phosphorylation, P1-3 kinase activation (159) and Rho-family GTPases (158). Following invasion, the bacteria replicate and form biofilm-like communities inside the cell (5). The members of these intracellular communities consist of slow-growing coccoid bacteria that can fill the entire available cell volume. It has been suggested that these biofilms can then enter a quiescent state and provide a reservoir for recurrent infection (5, 174), however, recurrent infections often involve a different UPEC serotype (123). UPEC from within these intracellular communities can become motile again and exit the epithelial cell to re-colonise the urinary tract or spread the infection. Interestingly the emergent bacteria adopt a highly elongated cellular morphology as they exit the infected bladder cell and these elongated bacteria are extremely resistant to attack by polymorphonuclear leukocytes (PMNs) (121). In strains where type 1 fimbriae are continually expressed, infection is usually restricted to the bladder (35). However, in strains were type 1 fimbriae retain their phase-variability afimbriate bacteria can detach from the bladder epithelium and ascend to the kidney where
colonisation is mediated by P fimbriae attachment to digalactoside receptors on the kidney epithelium (139). The kidney epithelium can then be damaged by bacterial toxins (257), including haemolysin and Sat, and the induced immune response, often resulting in glomerular damage. In some instances, the damage can result in a breach of the epithelium leading to bacteraemia and sepsis (268). Some UPEC strains have also been shown to invade epithelial cells independent of type 1 fimbriae. UPEC also produce Dr fimbriae (also expressed by DAEC) which have been shown to promote internalisation of bacteria (89), following interactions with CD55 (also known as DAF) and/or α5β1 integrin (196, 233). The exact mechanism of this invasion pathway is unclear but is believed to involve unstable microtubules and be independent of actin rearrangements (92). The CNF1 toxin is also believed to play a role in UPEC invasion. CNF1 is a large, 113 kDa, toxin with deamidase activity. CNF1 acts on Rho GTPases, converting specific glutamine residues to glutamate, resulting in constitutive activation and ultimately degradation by the ubiquitin pathway. CNF1 affects bacterial invasion either by inducing cell death and allowing access to underlying tissue (169) or by directly increasing internalisation by promoting membrane ruffling and filopodia formation following depletion of the Rho GTPase, Rac (61). Despite these myriad virulence pathways the extreme heterogeneity of the UPEC group means that its pathogenesis is still not completely understood. Further study of UPEC will undoubtedly uncover additional important virulence factors.

1.1.3.2 Neonatal-meningitic E. coli - NMEC

E. coli is the principle gram-negative causative agent of sepsis and meningitis in neonates and is second only to the group B Streptococci overall. The strains that cause these devastating infections are collectively known as NMEC and among NMEC strains those expressing the K1 capsular antigen are predominant. E. coli K1 accounts for up to 40 % of E. coli sepsis and 80 % of E. coli meningitic isolates (138, 212) and generally results in a much poorer clinical outcome than other NMEC strains (162). In many respects NMEC strains are similar to those classified as UPEC and indeed share many of the same virulence factors. E. coli expressing the K1 capsule are also common in urinary tract infections (118) and the highly virulent K1-expressing NMEC may have evolved from a K1-expressing UPEC strain. NMEC strains are more likely to express
type 1, P and S fimbriae than non-virulent or diarrhoeagenic *E. coli* strains (138) and contain many of the same pathogenicity islands as UPEC strains (58, 109).

NMEC pathogenesis is a complex process involving colonisation of the newborn during or soon after birth, penetration of the epithelium and haematogenous spread resulting in sepsis. Adhesion to and invasion of the brain microvascular endothelial cells, which form the blood-brain barrier, then follows (Fig. 1.2). Penetration of the BBB is accompanied by the severe clinical symptoms of meningitis, including inflammation of the meninges, migration of white blood cells into the cerebrospinal fluid (CSF) and increased intracranial pressure ultimately resulting in neural damage and often death (16). In many cases the onset of clinical symptoms occurs too late to allow successful chemotherapeutic intervention. As a result, the mortality rate in cases of NMEC infection remains high despite improvements in antimicrobial treatment and supportive care (32, 50, 91).

NMEC is probably acquired at birth from the mother’s genital tract, an assumption supported by the similarity between UPEC and NMEC, although subsequent contamination from environmental sources or other individuals cannot be discounted. The theory of acquisition from the mother at birth is supported by the finding that *E. coli* K1 positive stool samples from newborns occur much more frequently if the mother is *E. coli* K1 positive (77 % compared with 20 %) (220). *E. coli* K1 has been shown to invade bladder epithelial cells (165) and this has been suggested as the route of infection by NMEC. However, it is generally accepted that the primary site of colonisation and invasion is the gastrointestinal (GI) tract, an assumption supported by the presence of *E. coli* K1 in neonatal stool samples (220) and the ability of NMEC isolates to invade the well characterised T84 and Caco-2 GI epithelial cell lines (30). In addition to invading these cell lines, *E. coli* K1 strain RS218 was also found to efficiently cross polarised epithelial monolayers in transcytosis assays (30). These findings also agree with the model of NMEC pathogenesis described above which requires penetration of the host epithelium prior to haematogenous spread. One study has focussed on the colonisation of the GI tract using a 12-day old infant rat model of NMEC disease. Martindale *et al* employed signature-tagged transposon mutagenesis and identified 16 genes required for colonisation of the GI tract (157). The 16 genes
identified included 3 transcriptional regulators, 4 involved in metabolism, 2 membrane transporters, 2 involved in LPS biosynthesis, \( \text{timH} \), encoding the type 1 fimbrial tip adhesin, and a further four genes of unknown function, \( \text{dgcA-D} \). The function of these four genes has yet to be elucidated although \( \text{dgcD} \) was found to be specific to \( \text{E. coli} \) K1, \( \text{dgcA} \) has a homologue in \( \text{E. coli} \) K-12 and \( \text{dgcB} \) has homologues in UPEC and EAEC strains. No further work has been carried out specifically studying GI colonisation and to date no gene has been implicated in the invasion of intestinal epithelial cells following colonisation.

Following penetration of the epithelial barrier NMEC must avoid immune clearance and disseminate in the bloodstream in order for an infection to progress to sepsis or meningitis. Survival in the bloodstream has been the subject of considerable research for a number of years and picture is now beginning to emerge of the strategies employed by NMEC to survive and even thrive in this hostile environment. The earliest studies revealed that both LPS and the K1 capsule were vital for resistance to survival in the bloodstream. LPS is thought to mediate resistance to complement-mediated lysis (198) and the K1 capsule is poorly antigenic and may act as a mask for other surface antigens (126). The K1 capsule also mediates complement resistance and has anti-opsonisation properties (244). A role for the highly abundant outer membrane protein, OmpA, in resistance to complement has also been described (271). The mechanism of OmpA-mediated complement resistance was recently elucidated and involves an interaction between the amino-terminal portion of OmpA and C4bp, a regulator of the complement cascade, which inhibits activation of the cascade (200). NMEC was long thought to survive as planktonic bacteria in the bloodstream, protected from the immune response by LPS, the K1 capsule and OmpA. However, recent data has conclusively shown that NMEC can invade macrophage and monocytes in a manner which bypasses the normal process of phagocytosis and avoids killing of the bacteria (249). Once internalised, the bacteria can then replicate intracellularly and disseminate throughout the host without being exposed to the immune response. Invasion of macrophage is thought to involve at least two bacterial proteins, OmpA and TraJ (105, 249). In addition to promoting invasion of macrophage, OmpA also appears to inhibit apoptosis of infected macrophage by inducing \( \text{Bcl}_{\text{XL}} \), a potent inhibitor of apoptosis, which blocks the release of cytochrome \( c \) from mitochondria and hence
prevents activation of the apoptotic caspase cascade (248). OmpA⁺ NMEC were also shown to inhibit the production of pro-inflammatory cytokines in infected monocytes via the NF-κB pathway (232). Mutants defective in macrophage invasion were also significantly less virulent in the neonatal rat model, suggesting that the ability to survive intracellularly in the bloodstream is vital to the widespread dissemination of NMEC throughout the host (105). This suggestion is further supported by reports that a threshold level of bacteraemia in neonates is required for NMEC infection to progress to meningitis: of neonates with a bacteraemia of greater than 10⁵ CFU/ml of blood, 55% progressed to meningitis (56). It was even possible to induce meningitis in adult rats by increasing the inoculum size to produce a bacteraemia of greater than 10⁴ CFU/ml of blood (132, 245). This indicates that it is immune naivety that makes neonates susceptible to NMEC infection rather than some preference of the bacteria for neonatal epithelial or endothelial cells. It is therefore possible that NMEC infection of an immuno-compromised adult could result in sepsis and meningitis and even a healthy adult could theoretically be infected with as little as 10⁷ - 10⁸ CFU delivered via a puncture wound such as a needle-stick injury.

Following dissemination throughout the host NMEC have the ability to colonise and invade the brain microvascular endothelial cells (BMECs) which form the blood-brain barrier (BBB). This stage of NMEC pathogenesis has received the most attention and several models have been developed to study this process in vitro and in vivo. Several proteins involved in adhesion to and invasion of BMECs have been identified using primary and immortalised human and bovine BMECs and the neonatal rat model of NMEC sepsis and meningitis. Several different techniques were utilised to identify these bacterial factors, including TnphoA mutagenesis for AslA, IbeA, IbeB and YijP (106, 110, 113, 266), differential fluorescence induction for TraJ (11) and homology with known virulence factors for OmpA and CNFl (129, 204). Many of these same virulence factors were also identified in a study employing signature-tagged mutagenesis (12). In each case isogenic mutants were less invasive in BMECs in vitro and several also displayed virulence defects in the rat model. IbeA, originally Ibe10, was the first of these proteins to be described (113) and a receptor, Ibe10R, has been identified for this protein which localises to the surface of cultured BMECs (202). Strong evidence has been presented supporting a role for this protein in the invasion of
the cultured BMECs including complementation of the invasion defect and inhibition with purified protein (112). Following the identification of IbeA, the abundant outer membrane protein OmpA was also implicated in the invasion of BMECs (204). OmpA was first considered as it shares limited homology with the opacity protein Opa from *Neisseria meningitidis* which has been implicated in cellular invasion (262). *ompA* mutants were significantly less invasive *in vitro* and less virulent in the rat model (267). A BMEC receptor, gp96, has also been identified for OmpA (201). OmpA interacts with this protein via GlcNAc1,4-GlcNAc sugar epitopes (47, 203). Uptake of bacteria by BMECs involves large-scale rearrangements of cellular actin and these actin rearrangements are dependent on OmpA expression (247). Actin rearrangements occur downstream of PI-3 kinase activation and result in loosened tight-junctions (246), perhaps compromising the barrier function of the BMEC monolayer. CNFl was also found to contribute to invasion of BMECs in a manner similar to that described for UPEC invasion of bladder epithelial cells (section 1.1.3.1) via constitutive activation of Rho GTPases. Less work has been performed on the remaining protein factors involved in BMEC invasion. AslA is similar to an arylsulphatase from *K. pneumoniae*, which is involved in sulphate metabolism (106). Deletion of the *ibeB, yijP* or *traJ* genes reduces invasion of an *E. coli* K1 strain into BMECs and these defects can be complemented (11, 110, 266). However, the mechanisms by which these proteins contributes to invasion has not been elucidated. These proteins may be involved in receptor-ligand interactions or in regulatory pathways which in turn alter the expression of other virulence factors. For example, the *ibeB* gene is allelic to the *cusC* gene of *E. coli* K-12, which is induced at high copper concentrations and is thought to encode a copper efflux system (175). It is possible that a defect in copper efflux in an isogenic *ibeB* mutant could affect NMEC invasion through altered gene expression. Another important factor in successful NMEC penetration of the blood-brain barrier is the K1 capsule. It has been shown that although capsule deficient mutants of *E. coli* K1 strain RS218 can invade BMECs normally only bacteria expressing the capsule can cross the BMEC monolayer as viable bacteria (107). In the rat model animals infected with capsule deficient bacteria had sterile CSF samples but bacteria could be detected in the CSF following immunolabelling. It was later discovered that expression of the K1 capsule prevented lysosomal fusion with the *E. coli* containing vacuole in BMECs and hence enhanced intracellular survival (130).
1.2 Translocation, membrane insertion and structure of β-barrel proteins in *E. coli*

### 1.2.1 Mechanisms of inner membrane translocation in *E. coli*

All outer membrane proteins (OMPs) are initially synthesised as a preprotein with an amino-terminal signal sequence (also known as a signal peptide) which facilitates the translocation of the protein across the inner membrane and is subsequently removed upon translocation. *E. coli* utilise at least two distinct translocation pathways, Sec and TAT (twin-arginine translocation), which differ in their ability to translocate folded proteins. The Sec system relies on a cytoplasmic chaperone, SecB, to bind to unfolded preproteins and keep them in an unfolded translocation-competent state (206), whereas the TAT system appears to be capable of translocating at least partially folded and perhaps even oligomerised preproteins and even proteins with associated co-factors (219, 234). Proteins targeted to the Sec translocation pathway have a 16-23 residue signal or leader sequence which generally contains a basic residue at the N-terminal end, followed by a long stretch of (10-18) hydrophobic amino acids and finally a signal peptidase cleavage site with alanines in the -3 and -1 positions (42). Proteins translocated by the TAT pathway have a considerably longer signal peptide (30-50 residues) which contains a characteristic and essential twin-arginine motif (RR) (219).

### 1.2.2 The Sec translocation pathway

Most OMPs in *E. coli* appear to traverse the inner membrane via the Sec translocation pathway (Fig. 1.3) (reviewed in (45, 62, 64)). Upon translation, the nascent chain of a Sec-dependent preprotein is bound by the SecB chaperone which maintains the protein in an unfolded conformation and directs it to the membrane translocation machinery by direct interactions with the membrane-bound SecA translocase. The preprotein is then fed through a pore in the inner membrane consisting of the SecY, E and G proteins in an energy-dependent manner with SecA supplying the necessary ATP-hydrolysis function. The preprotein is then cleaved by the Lep leader peptidase, a serine peptidase, releasing the mature protein on the periplasmic side of the membrane. Several other
proteins play important although non-essential roles in this process, including: two additional translocase subunits, SecD and F, which are necessary for optimal secretion, and several additional chaperones, DnaK, DnaJ, GrpE, and the Ffh/4.5SRNA system.

1.2.3 Periplasmic transit and insertion into the outer membrane

The Sec pathway is involved in the translocation of both OMPs and periplasmic proteins and both classes of protein rely on similar signal peptides to direct them onto this pathway. However, once the membrane has been crossed additional signals are presumably required to target proteins to the outer membrane and these signals appear to reside within the mature sequence of at least some OMPs (24, 133, 134) although no specific targeting signal has yet been identified. It has been suggested that the overall conformation of the mature protein may itself be the signal which directs it to the outer membrane (250). The mechanism by which OMPs traverse the periplasm is also unclear with two competing theories suggesting that the proteins either reach the outer membrane after an intermediate periplasmic stage or directly cross from one membrane to the other at points of contact between the two known as Bayer’s junctions. Whichever of these two competing theories is true, the final stage of the process must involve the insertion of the protein into the membrane and the adoption of its final structural conformation. Outer membrane β-barrel proteins were thought to spontaneously fold and insert into the membrane (250) but recent evidence has suggested that specific chaperones are required for the insertion of putative β-barrel proteins into the outer membranes of both bacteria and mammalian mitochondria (214, 277).

1.2.4 β-barrel protein structure

The two membranes of E. coli differ in both structure and function. The inner membrane is a uniform lipid bilayer forming a highly selective barrier whereas the outer membrane is a highly asymmetric, relatively non-selective and porous barrier consisting of a lipid inner leaflet and an outer leaflet comprised entirely of LPS. The protein complement of the two membranes is also drastically different with transmembrane α-helices found exclusively in the inner membrane and the outer
**Figure 1.3** Export of proteins via the Sec pathway. Graphical representation of Sec-dependent protein export across the inner membrane in *E. coli*. The known components of the Sec pathway are shown including the SecB chaperone, SecA translocase, SecYEG pore with associated SecD and SecF and the leader peptidase, Lep. The heavier black line represents the preprotein leader peptide.
Figure 1.4 Three-dimensional structure of a β-barrel protein. Structure of the *E. coli* OmpX 8-stranded β-barrel as determined by X-ray diffraction analysis (264).
membrane proteins, with the exception of the lipoproteins, adopting a β-barrel confirmation (Fig. 1.4). The first outer membrane protein crystal-structure to be solved was the *Rhodobacter capsulatus* porin (272) providing the first evidence for the β-barrel structure of OMPs and to date all of the OMPs that have been crystallised adopt a β-barrel conformation. A β-barrel consists of between 8 and 22 antiparallel β-sheets (136) roled into a cylindrical shape in the membrane with the full hydrogen-bonding potential of the β-sheets fulfilled by inter-strand bonds with neighbouring β-sheets. The β-sheets are generally linked by short turns on the periplasmic side and longer loops on the external side of the membrane. In order to span a membrane, the β-sheets which form the body of the barrel must consist of a minimum of 6 amino acids, although the average is approximately 12 (136), and generally contain a hydrophobic residue in every second position along the strand. As successive residues in a β-sheet align with their side-groups pointing in opposite directions, these residues align with their hydrophobic side-chains exposed on the surface of the barrel forming a hydrophobic skin which is in contact with the core of the membrane while the side-chains of the alternating, usually hydrophilic, residues then point into the core of the protein and hence are shielded from the hydrophobic interior of the membrane (136, 205, 264). This property allows the membrane-spanning regions of β-barrel proteins to be generally less hydrophobic than transmembrane α-helices. One known exception to this is found in the *E. coli* phospholipase A β-barrel where a patch of hydrophilic residues are exposed on the outer surface of the barrel and this area has been implicated in the oligomerisation of the protein (240). These residues are involved in hydrophilic interactions with neighbouring protein molecules and thus are not exposed to the highly hydrophobic interior of the membrane. In a folded β-barrel the β-sheets are also commonly flanked by aromatic residues, forming a ring at the interface between the hydrophobic membrane core and the polar head groups (229). These aromatic residues serve to bridge the polar and non-polar environments that the outer surface of the protein is exposed to at this membrane interface.

β-barrel proteins are major components of the *E. coli* outer membrane, with some proteins such as OmpA being present in up to 100,000 copies per cell (136, 183), and therefore form a vital link between the bacterium and its environment whether by
sensing environmental signals and initiating signalling cascades or by mediating direct interactions between the bacterium and its host (47, 74).
1.3 Scope of this thesis

To date, little research has focussed on the initial stages of neonatal infection with NMEC. The purpose of the work presented in this thesis was to investigate the colonisation and invasion of the GI epithelium by *E. coli* K1 strain RS218. Putative novel virulence factors were identified by a bioinformatic examination of the unfinished *E. coli* RS218 genome sequence and the role of one novel putative virulence factor was assessed by standard techniques using an *in vitro* model of the human GI epithelium based on the T84 colonic carcinoma cell line. The role of known NMEC virulence factors in the invasion of the neonatal GI tract was also investigated.
Chapter 2  Materials and Methods
2.1 General Methods

2.1.1 Bacterial strains and culture conditions

2.1.1.1 Bacterial strains

All *E. coli* strains used in this study are described in Table 2.1. Stocks were maintained in LB broth supplemented with 8.7% (v/v) DMSO and stored at -70°C.

2.1.1.2 Bacterial growth media

All media were prepared using Millipore 18 MΩ cm⁻¹ grade water and chemicals obtained from Difco and Oxoid. Media were sterilised by autoclaving at 120°C for 20 min prior to use. Additives not suitable for autoclaving were sterilised by filtration through 0.2 µm Millex filters (Milipore). The quantities listed below are sufficient for 1 litre of medium.

Luria-Bertani (LB) broth and agar:

LB broth and agar were used throughout this study for the routine culturing of all bacterial strains except where otherwise stated.

- **LB broth:** 10 g Oxoid tryptone, 5 g yeast extract, 5 g NaCl
- **LB agar:** 10 g Oxoid tryptone, 5 g yeast extract, 5 g NaCl, 15 g Bacto agar

Columbia agar:

Columbia agar was used in the preparation of blood agar plates for recovery of bacteria following mammalian cell adhesion and invasion assays.

- **Columbia agar:** 39 g Oxoid columbia agar base

After autoclaving, 50 ml of defibrinated horse blood was added.

Top agar:
Top agar was used to prepare and titre bacteriophage lysates.

Top agar: 7 g Bacto agar

After autoclaving 10 ml of 1M MgSO₄ was added

SOC medium:
SOC medium was used following electroporation of *E. coli* strains to increase the recovery of viable bacteria following transformation.
SOC medium: 20 g Oxoid tryptone, 5 g yeast extract, 0.5 g NaCl

After autoclaving 0.95 g MgCl₂, 1.2 g MgSO₄ and 1.8 g glucose was added.

### 2.1.1.3 Bacterial culture conditions

Bacteria were routinely grown on LB agar plates and in shaken aerobic liquid cultures at 37°C, except where otherwise stated. Liquid cultures were inoculated by transferring single colonies from agar plate cultures into an appropriate volume of LB broth and grown overnight. Where mid-logarithmic cultures were required, overnight cultures were diluted 1:100 in fresh media and grown to the appropriate optical density at 600 nm.

### 2.1.1.4 Antibiotics and media additions

All stock antibiotics and media additives were filter sterilised through 0.2 μm Millex filters (Milipore) and stored at -20°C. Carbenicillin, kanamycin and spectinomycin were prepared as 50 mg/ml stocks in Millipore grade water and used at a working concentration of 50 μg/ml. Tetracycline was prepared as a 30 mg/ml stock in 70% (v/v) ethanol and used at a working concentration of 30 μg/ml. Chloramphenicol was prepared as a 10 mg/ml stock in 70% (v/v) ethanol and used at a concentration of 10 μg/ml. The chromogenic substrates for the beta-galactosidase and alkaline phosphatase enzymes, X-Gal and X-P respectively were prepared as 25 mg/ml stocks in DMF and used at working concentrations of 50 μg/ml. The lac operon inducer IPTG was prepared as a 100 mM stock and used at concentrations of 0.1–1 mM as appropriate.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Relevant details</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJD1484</td>
<td>VL386 <em>ihfA Δ82::Tn10</em> Tet' (pBR322)</td>
<td>(59)</td>
</tr>
<tr>
<td>CJD 1486</td>
<td>VL386 <em>ihfA Δ82::Tn10</em> Tet' (pBR322-<em>ihfA</em>)</td>
<td>(59)</td>
</tr>
<tr>
<td>DH5α</td>
<td>F' *endA1 hsdR17 (rK mK) glvV44 thi-1 recA1 gyrA (Nal') relA Δ(lacIZYA-argF)U169 deoR (Ph80dlacΔ(lacZ)M15)</td>
<td>Invitrogen Life Technologies</td>
</tr>
<tr>
<td>MC4100</td>
<td>F' araD139 Δ(argF-lac)U169 rpsL150 relA flb-5301 ptsF25 deoC1 thiA1</td>
<td>ATCC</td>
</tr>
<tr>
<td>O42</td>
<td>EAEC clinical isolate, serotype O44:H18</td>
<td>(99)</td>
</tr>
<tr>
<td>RS218</td>
<td>NMEC clinical isolate, serotype O18:K1:H7</td>
<td>(3)</td>
</tr>
<tr>
<td>RShek</td>
<td>RS218 <em>hek</em>:spc</td>
<td>This study</td>
</tr>
<tr>
<td>RSibeB</td>
<td>RS218 <em>ibeB</em>:spc</td>
<td>This study</td>
</tr>
<tr>
<td>RSihf</td>
<td>RS218 <em>ihfA Δ82::Tn10</em> Tet' (transduced from CJD1484)</td>
<td>This study</td>
</tr>
<tr>
<td>RSlac</td>
<td>RS218 <em>lac</em>:cat</td>
<td>This study</td>
</tr>
<tr>
<td>RSlac ihf</td>
<td>RS218 <em>lac</em>:cat <em>ihfA Δ82::Tn10</em> (transduced from CJD1484)</td>
<td>This study</td>
</tr>
<tr>
<td>RSompA</td>
<td>RS218 <em>ompA</em>:Tn7 transprimer Kan'</td>
<td>This study</td>
</tr>
<tr>
<td>XL-1 blue</td>
<td>*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lac''ZΔM15 Tn10 (Tet')]</td>
<td>Stratagene</td>
</tr>
<tr>
<td>VL386</td>
<td>*ara Δ(lac-pro) rpsL thi Φ(fimA-lacZ) fimE::ISI λpl(209)</td>
<td>(59)</td>
</tr>
</tbody>
</table>
2.1.2 Eukaryotic cell lines and growth conditions

2.1.2.1 Eukaryotic cell lines
All cell lines used were obtained from ATCC (Manassas, VA, U.S.A.). The mammalian cell lines used were T84 (ATCC CCL-248), CHO-K1 (ATCC CCL-61) and pgsA-745 (ATCC CRL-2242) cells. Stocks of all cell lines were maintained in cell freezing medium-DMSO (Sigma-Aldrich) under liquid nitrogen.

2.1.2.2 Cell line growth conditions
All cell lines were grown in a 1:1 mixture of DMEM and Ham's F12 medium supplemented with 10% (v/v) heat inactivated fetal bovine serum (Life Technologies). Cells were routinely grown in 25 or 75 cm² tissue culture flasks at 37 °C in 5% CO₂. Upon reaching confluence monolayers were disrupted by treatment with a trypsin/EDTA solution (Sigma-Aldrich) and diluted (1:4 for T84 cells and 1:20 for CHO-K1 and pgsA-745 cells) into fresh medium.

2.1.3 Plasmids, bacteriophage and oligonucleotides

2.1.3.1 Plasmids
All plasmids used in this study are listed in Table 2.2 along with relevant descriptions and sources. For plasmids constructed during this study the details of construction are described in the relevant chapters.

2.1.3.2 Oligonucleotides
The sequences of all oligonucleotides used in this study are listed in Table 2.3. All oligonucleotides were synthesised by MWG-Biotech, Ebersberg, Germany.
2.2 Nucleic acid methodologies

2.2.1 Transformation of *E. coli* strains

Two distinct methods were used for the introduction of foreign DNA into *E. coli* cells. Cells were made competent either by repeated washes with a cold calcium chloride solution or cold Millipore grade water and then transformed by heat-shock (155) or electroporation (60) respectively. Significantly greater transfection efficiencies were achieved using the electroporation method. The calcium chloride method was used for routine transformation of intact plasmids into *E. coli* K-12 strains. Due to the higher efficiencies achieved, the electroporation method was used with *E. coli* K-12 strains when higher efficiencies were required and routinely with *E. coli* K1 strains.

2.2.1.1 Transformation of *E. coli* K-12 strains using calcium chloride method

A 3 ml culture of the strain to be made competent was inoculated from a single colony into LB broth and incubated overnight with shaking. This culture was then diluted 1:100 in 100 ml of fresh LB broth and grown to an \( \text{OD}_{600nm} \) of between 0.4 and 0.6. The cultures were then incubated on ice for 20 min and the bacteria were harvested by centrifugation at 6,000 \( \times g \) for 8 min. The pellet was resuspended in 20 ml of cold \( \text{CaCl}_2 \) solution (60 mM \( \text{CaCl}_2 \), 15% (v/v) glycerol, 10 mM PIPES, pH 7) and harvested as before. This pellet was resuspended in 20 ml cold \( \text{CaCl}_2 \) solution and incubated on ice for 30 min before harvesting the bacteria as before. The pellet was then resuspended in a final 4 ml of cold \( \text{CaCl}_2 \) and incubated overnight on ice before aliquoting in 100 \( \mu l \) volumes and storage at \(-70^\circ C\).

Competent bacterial cells (100 \( \mu l \)) were mixed with between 10 ng and 1 \( \mu g \) of plasmid DNA (in a volume not exceeding 10 \( \mu l \)) and incubated on ice for 10 min. The cells were then transferred to a 42 \( ^\circ C \) water bath and incubated for 2 min, thus inducing the uptake of the DNA. Following a further 2 min incubation on ice the cells were mixed with 1 ml of pre-warmed LB or SOC broth and incubated at 37 \( ^\circ C \) for 1 h to allow expression of plasmid-borne antibiotic resistance genes. Aliquots of 1 \( \mu l \), 10 \( \mu l \), 100 \( \mu l \) and 1 ml of the transformation mixture were then spread on selective LB agar plates.
<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant details</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC184</td>
<td>pSC101 replicon Tet' Cm'</td>
<td>(31)</td>
</tr>
<tr>
<td>pAM3</td>
<td>pIbeB with Spc' omega cassette from pHp45Ω inserted into ibeB Spc' Ap'</td>
<td>(161)</td>
</tr>
<tr>
<td>pBhek7</td>
<td>hek ORF cloned downstream of the T7 promoter in pBSK II Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pBR322</td>
<td>pMB1 replicon Ap'</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>pBSK::hspc1</td>
<td>pBSKII containing a 4 kb XhoI/EcoRV fragment of E. coli RShek DNA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>including the hek::spc allele Spc' Ap'</td>
<td></td>
</tr>
<tr>
<td>pBSKII</td>
<td>ColE1 replicon Ap'</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pCGV1</td>
<td>pMB1 replicon Spc', source of phoA gene</td>
<td>(94)</td>
</tr>
<tr>
<td>pET125</td>
<td>Cosmid clone of the region of DNA from ETEC strain H10407 encompassing the tia gene Ap'</td>
<td>(74)</td>
</tr>
<tr>
<td>pHEK6</td>
<td>866 bp of RS218 DNA carrying the hek gene cloned into pBSKII Ap'</td>
<td>S. G. J. Smith, unpublished</td>
</tr>
<tr>
<td>pH+</td>
<td>pBSKII containing the hek gene with 1138 bp of upstream DNA Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pHnhe6</td>
<td>pHEK6 with a novel Nhel site inserted before the last codon of hek</td>
<td>This study</td>
</tr>
<tr>
<td>pHp45Ω</td>
<td>pMB1 replicon Spc'</td>
<td>(21)</td>
</tr>
<tr>
<td>pH-P</td>
<td>see Table 5.1</td>
<td>This study</td>
</tr>
<tr>
<td>pH+P</td>
<td>see Table 5.1</td>
<td>This study</td>
</tr>
<tr>
<td>pHSPC4</td>
<td>pHEK6 with Spc' omega cassette from pHp45Ω inserted into hek Spc' Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pIbeB</td>
<td>1871 bp of E. coli RS218 DNA containing ibeB cloned into pBSKII Ap'</td>
<td>S. G. J. Smith, unpublished</td>
</tr>
<tr>
<td>pKD3</td>
<td>Plasmid source of FRT-flanked cat gene for gene disruption Ap' Cm'</td>
<td>(46)</td>
</tr>
<tr>
<td>pKD46</td>
<td>Temperature sensitive plasmid carrying the red and gam genes of λ-phage under control of a P_bad promoter Ap'</td>
<td>(46)</td>
</tr>
<tr>
<td>pKOBEAGA</td>
<td>Temperature sensitive plasmid carrying the red and gam genes of λ-phage under control of a P_bad promoter Ap'</td>
<td>(33)</td>
</tr>
<tr>
<td>pL-P</td>
<td>see Table 5.1</td>
<td>This study</td>
</tr>
<tr>
<td>pL+</td>
<td>see Table 5.1</td>
<td>This study</td>
</tr>
<tr>
<td>pL+P</td>
<td>see Table 5.1</td>
<td>This study</td>
</tr>
<tr>
<td>pLacCat</td>
<td>lac::cat fragment cloned into pBSK II Cm' Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pLoop1</td>
<td>Derivative of pHEK6 with the DNA sequence corresponding to the first loop of the Hek protein replaced with CGCGCG Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pLoop2</td>
<td>Derivative of pHEK6 with the DNA sequence corresponding to the second loop of the Hek protein replaced with CGCGCG Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pLoop3</td>
<td>Derivative of pHEK6 with the DNA sequence corresponding to the third loop of the Hek protein replaced with CGCGCG Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>pLoop4</td>
<td>Derivative of pHEK6 with the DNA sequence corresponding to the forth loop of the Hek protein replaced with CGCGCG Ap'</td>
<td>This study</td>
</tr>
<tr>
<td>Vector/Plasmid</td>
<td>Description/Relevant Information</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>pM-Z, pM-ZS1, pM-ZS2</td>
<td>See Table 5.1 This study</td>
<td></td>
</tr>
<tr>
<td>pM+ZS1, pM+ZS2</td>
<td>See Table 5.1 This study</td>
<td></td>
</tr>
<tr>
<td>pMAL-c2</td>
<td>MBP protein-fusion expression vector (cytoplasmically expressed) Amp' New England Biolabs</td>
<td></td>
</tr>
<tr>
<td>pMhek1</td>
<td>DNA coding sequence of mature Hek cloned downstream of the malE gene on pMAL-c2 Amp' This study</td>
<td></td>
</tr>
<tr>
<td>pMIP</td>
<td>See Table 5.1 This study</td>
<td></td>
</tr>
<tr>
<td>pOmpA</td>
<td>pBSKII containing the E. coli RS218 ompA with a spontaneous deletion of the promoter region and the 5' portion of the ompA gene Amp' S. G. J. Smith, unpublished</td>
<td></td>
</tr>
<tr>
<td>pPD101</td>
<td>pSC101 replicon Cm' (54)</td>
<td></td>
</tr>
<tr>
<td>pPDhek22</td>
<td>866 bp of RS218 DNA carrying the hek gene cloned into pPD101 Cm' This study</td>
<td></td>
</tr>
<tr>
<td>pPDompA11</td>
<td>2,000 bp of RS218 DNA carrying the ompA gene with its native promoter cloned into pPD101 Cm' This study</td>
<td></td>
</tr>
<tr>
<td>pQF50</td>
<td>ColE1 replicon Amp' carrying a promoterless lacZ gene Amp' (72)</td>
<td></td>
</tr>
<tr>
<td>pRC1</td>
<td>pOmpA containing a Tn7-based transprimer in the ompA gene Amp' Km' (37)</td>
<td></td>
</tr>
<tr>
<td>pSag1</td>
<td>fimE promoter fragment cloned into pQF50 Amp' (84)</td>
<td></td>
</tr>
<tr>
<td>pThek6</td>
<td>hek ORF from pBhek7 subcloned downstream of the P-svg promoter in pTrc99a Amp' This study</td>
<td></td>
</tr>
<tr>
<td>pTia5</td>
<td>tia gene from pET125 subcloned into pBSK II Amp' This study</td>
<td></td>
</tr>
<tr>
<td>pTnhe4</td>
<td>pThek6 with a novel NheI site inserted before the last codon of hek This study</td>
<td></td>
</tr>
<tr>
<td>pTrc99a</td>
<td>Cloning vector with an IPTG-inducible P-svg promoter Amp' (4)</td>
<td></td>
</tr>
<tr>
<td>Primer name</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>hekfm</td>
<td>5' - AAG GAA GGT AAA AGC GG - 3'</td>
<td></td>
</tr>
<tr>
<td>hek-nhe1_f</td>
<td>5' - GTA TGA CTT ATA ACT TCG CTA GCT GAC GAC ACT GCT CC - 3'</td>
<td></td>
</tr>
<tr>
<td>hek-nhe1_r</td>
<td>5' - GGA GCA GTG TCG TCA GCT AGC GAA GTT ATA AGT CAT AC - 3'</td>
<td></td>
</tr>
<tr>
<td>hekrm</td>
<td>5' - CCG CTC TAG AAC TAG TG - 3'</td>
<td></td>
</tr>
<tr>
<td>hekrORF</td>
<td>5' - CAG AAG TTA TAA GTC ATA CC - 3'</td>
<td></td>
</tr>
<tr>
<td>iber</td>
<td>5' - TGG TCA TCT CCG GCC AGT TC - 3'</td>
<td></td>
</tr>
<tr>
<td>lac downstream f</td>
<td>5' - AAA CCG ACA TCG CAG GCT TCT GCT TCA ATC - 3'</td>
<td></td>
</tr>
<tr>
<td>lac downstream r</td>
<td>5' - TCC GTG ACG TCT CGT TGC TTG CAT AAA CC - 3'</td>
<td></td>
</tr>
<tr>
<td>lac upstream f</td>
<td>5' - GAA ATT GTA GGC CTG ATA AGC GGA GCG CAT C - 3'</td>
<td></td>
</tr>
<tr>
<td>lac upstream r</td>
<td>5' - AAA GAA ATG TTT GCC AGC GTA GGG GAA AAC GCC - 3'</td>
<td></td>
</tr>
<tr>
<td>loop1_F</td>
<td>5' - GCG GTA TTC AGT GGC GGT ATT G - 3'</td>
<td></td>
</tr>
<tr>
<td>loop1_R</td>
<td>5' - GCG GTA TAC CAT CAC AGA GCC ACC - 3'</td>
<td></td>
</tr>
<tr>
<td>newloop2 fp</td>
<td>5' - CGC GAG GTG TCA TGC AAC ACA CTA ATG - 3'</td>
<td></td>
</tr>
<tr>
<td>newloop2 rp</td>
<td>5' - CGC GTC CAG TTC TGT ACG AAC CGG - 3'</td>
<td></td>
</tr>
<tr>
<td>L3f</td>
<td>5' - CGC AAC TTC GCA TGG AGC CTT GG - 3'</td>
<td></td>
</tr>
<tr>
<td>L3r</td>
<td>5' - CGC GGC GTA GCC AAT CCC TGC - 3'</td>
<td></td>
</tr>
<tr>
<td>L4f</td>
<td>5' - CGC AAA AGT CAT GAC ATC ATG CTT GG - 3'</td>
<td></td>
</tr>
<tr>
<td>L4r</td>
<td>5' - CGC ATC AAG ATA GCG ATA GCT GAG - 3'</td>
<td></td>
</tr>
<tr>
<td>M13_(−43)</td>
<td>5' - AGG GTT TTC CCA GTC ACG ACG TTG - 3'</td>
<td></td>
</tr>
<tr>
<td>ompAf</td>
<td>5' - AGC CAA TGC TTT CAG TCG TG - 3'</td>
<td></td>
</tr>
<tr>
<td>ompAf1</td>
<td>5' - GCA CAC TGG TGC TAA ACT GG - 3'</td>
<td></td>
</tr>
<tr>
<td>ompAr</td>
<td>5' - ATG CAG AAC ACC AGC ATA CG - 3'</td>
<td></td>
</tr>
<tr>
<td>ompAr1</td>
<td>5' - GGA ATT TAT CGT CTG GCA AC - 3'</td>
<td></td>
</tr>
<tr>
<td>pQF50f</td>
<td>5' - CGC AAG GAA TGG TGC ATG CC - 3'</td>
<td></td>
</tr>
<tr>
<td>pQF50r</td>
<td>5' - GAC GGC CAG TGA ATT CCC TTT C - 3'</td>
<td></td>
</tr>
<tr>
<td>RI1</td>
<td>5' - CGC CAT GGA GAT AAT TGA AAT G - 3'</td>
<td></td>
</tr>
<tr>
<td>RSlac/catf</td>
<td>5' - GTT TTC CCC TAC CGT GGC AAA CAT TTT TTT AAT CAG TGT GTA GGC TGG AGC TGC TTT G - 3'</td>
<td></td>
</tr>
<tr>
<td>RSlac/catr</td>
<td>5' - GAT TGA AGC AGC CTG CGA TGT CGG TTT CCG CGA GGT GCC ATA TGA ATA TCC TCC TTA G - 3'</td>
<td></td>
</tr>
<tr>
<td>T7_promoter</td>
<td>5' - TAA TAC GAC TCA CTA TAG GGA GAC CG - 3'</td>
<td></td>
</tr>
</tbody>
</table>
and incubated overnight at 37 °C. Transformants were single colony purified on fresh LB agar plates containing appropriate antibiotics.

2.2.1.2 Transformation of E. coli strains by electroporation

A 3 ml culture of the strain to be made electro-competent was inoculated from a single colony into LB broth and incubated overnight with shaking. This culture was then diluted 1:100 in 100 ml of fresh LB broth and grown to an OD_{600nm} of between 0.4 and 0.6. The cultures were then incubated on ice for 20 min, the bacteria were harvested by centrifugation at 6,000 x g for 8 min and the pellet was resuspended in 100 ml of sterile ice-cold water. The bacteria were then harvested as before, resuspended in 50 ml cold water and harvested again. The pellet was then resuspended in 4 ml sterile ice-cold 10% (v/v) glycerol, harvested and resuspended again in ice-cold 10% glycerol to a final volume of 400 µl. 40 µl aliquots were stored at -70°C.

Electro-competent bacteria (40 µl) were mixed with between 10 ng and 2 µg of DNA (in a volume not exceeding 4 µl) and immediately transferred into a cold electroporation cuvette (BioSmith, 0.2 cm electrode width). The electroporation was then carried out in a Bio-Rad gene pulser at 12.5 kVcm⁻¹, 25 µF and 200Ω. The bacteria were then mixed with 1 ml of pre-warmed LB or SOC broth and incubated at 37 °C for 1 h to allow expression of plasmid-borne antibiotic resistance markers. Aliquots of 1 µl, 10 µl, 100 µl and 1 ml quantities of the transformation mixture were then spread on selective LB agar plates and transformants were single colony purified onto fresh LB agar plates following growth overnight at 37°C.

2.2.2 Transduction with bacteriophage P1vir dam rev6

2.2.2.1 Preparation of a P1vir lysate

A modified P1vir bacteriophage, P1vir dam rev6 was used for generalised transduction in E. coli K1. An overnight culture of the donor strain was diluted 1:50 into LB broth supplemented with 5 mM CaCl2 and grown at 37 °C to an OD_{600nm} of between 0.4 and 0.6. 2 ml of this culture was then mixed with 200 µl of P1vir phage stock and
incubated for a further 20 min at 37 °C to allow absorption of the phage particles onto the bacteria. 6 ml of top agar was then added and the resulting mixture was poured onto the surface two LB agar plates and incubated overnight at 37 °C. The top agar containing released phage particles and surviving bacteria was scraped into a 50 ml falcon tube. The surfaces of the plates were washed with 4 ml LB broth and this was added into the 50 ml tube. 200 µl of chloroform was added and the mixture was vortexed vigorously for 2 min. Cellular debris and agar was removed by centrifugation at 1,500 g for 15 min, the supernatant was transferred to a clean tube and a further 100 µl of chloroform was added. The resulting lysates were stored at 4°C.

2.2.2.2 P1vir transduction

An overnight culture of the recipient strain was harvested by centrifugation and resuspended in an equal volume of MC solution (10 mM MgSO4, 5 mM CaCl2). 100 µl of these bacteria was mixed with between 1 µl and 100 µl of a transducing lysate and incubated at 30 °C for 20 min. 100 µl of 1 M sodium citrate and 700 µl of warm LB broth was added and the mixture was incubated with shaking at 37 °C for 1 hour. 1 µl, 10 µl, 100 µl and 1 ml quantities of the bacteria were then spread on selective LB agar plates and grown overnight at 37 °C. Putative transductants were single colony purified successively three times onto fresh selective LB agar plates and screened by PCR.

2.2.3 λ Red-mediated allele replacement

Where suitable donor strains were not available for P1vir transduction novel mutated alleles were constructed in vitro and transferred to the bacterial chromosome using an allele-replacement system based on the gene products of the λ phage red operon (46). This system is based on the ability of the gam, bet and exo gene products to inhibit the Exonuclease V activity of RecBCD, allowing the transformation of linear DNA fragments, and promote homologous recombination at regions of homology between the chromosome and the transformed linear DNA fragment. The λ-phage redγα operon was located on the pKOBEGA (33) plasmid under the control of the strong, arabinose-inducible P_{BAD} promoter. This plasmid has a temperature sensitive pSC101 replicon, which permits plasmid replication at 30 °C but not at higher temperatures.
The plasmid also confers resistance to the antibiotic carbenicillin. The λ Red allele replacement system was initially adapted and optimised for use in E. coli K1 strain RS218 for the construction of a lac mutant derivative strain (see Chapter 5, section 5.2.2.3 for details).

2.2.3.1 Generation of mutated alleles in vitro

In order to generate a mutated allele, the appropriate region of the bacterial chromosome was first cloned into a plasmid vector and then disrupted by insertion of an antibiotic resistance cassette. The details of each gene disruption are described later in detail in the appropriate section. Once the mutated allele had been constructed and confirmed by restriction endonuclease mapping, a linear DNA fragment containing the novel allele was produced either by PCR amplification of the desired region or restriction endonuclease digestion with appropriate enzymes. The linear fragment was isolated and purified from an agarose gel (section 2.2.5.4) and concentrated by ethanol-precipitation (section 2.2.5.5). In the case of PCR products the DNA was digested with the restriction endonuclease DpnI prior to purification to decrease the amount of intact parental plasmid in the PCR mixture.

2.2.3.2 λ Red-mediated allele replacement in E. coli K1

E. coli K1 bacteria harbouring the pKOBEGA plasmid were grown to mid-logarithmic phase at 30°C in the presence of 50 μg/ml carbenicillin and expression of the Redα, β and γ proteins was induced by addition of 0.2% L-arabinose. The induced bacteria were grown for a further 2 hours and made electrocompetent (section 2.2.1.2). These electrocompetent bacteria were then transformed with 1-2 μg of purified linear DNA containing the relevant disrupted gene as described previously and following the 37°C expression step were spread on agar plates containing the appropriate antibiotic. All subsequent steps were carried out at 37°C in the absence of carbenicillin to ensure rapid loss of the pKOBEGA plasmid. Putative mutants were single-colony purified twice and then screened for the presence of the novel mutant allele by PCR. Loss of the pKOBEGA plasmid was confirmed by testing for carbenicillin sensitivity on agar containing the antibiotic.
2.2.4 Purification of plasmid and chromosomal DNA

2.2.4.1 Small-scale purification of plasmid DNA

The Genelute Plasmid Miniprep kit (Sigma-Aldrich) was used for the routine purification of plasmid DNA from overnight bacterial cultures. Purification was carried out according to the manufacturers instructions using a modified alkaline lysis method. Briefly, bacteria from a 3 ml overnight culture were harvested and lysed in an alkaline solution containing SDS, to denature proteins, and RNase, to degrade RNA. The alkaline conditions also denatured both the chromosomal and plasmid DNA which was released upon cell lysis. The solution was then rapidly neutralized causing the plasmid DNA to re-anneal and the chromosomal DNA to precipitate. The lysate was then cleared by centrifugation to remove cellular debris and chromosomal DNA and the soluble plasmid DNA was purified using a mini cation-exchange column. The purified DNA was eluted from the column in TE buffer following a desalting and washing step.

2.2.4.2 Large-scale purification of plasmid DNA

The Qiagen Plasmid Midi kit was used to purify plasmid DNA from 100 ml overnight cultures of \textit{E. coli} according to the manufacturers instructions. Purification was carried out according to the manufacturers instructions using a modified alkaline lysis similar to that described above. Bacteria were lysed as before and the cleared lysate is passed through a cation-exchange column, which binds the renatured plasmid DNA. The column with bound DNA was washed repeatedly and the DNA is eluted is a high-salt buffer. The DNA is then further purified and desalted by precipitation with isopropanol and resuspended in ddH$_2$O.

2.2.4.3 Purification of total genomic DNA

Chromosomal DNA was routinely purified using the Puregene genomic DNA purification kit (Gentra Systems). Briefly, bacteria from a 1.5 ml overnight culture were harvested by centrifugation, resuspended in a buffer containing tris [hydroxymethyl] aminomethane, EDTA and SDS and incubated at 80 °C for 5 min to lyse the bacteria. Contaminating RNA was removed by treatment with RNase and...
cellular proteins were removed by protein precipitation using ammonium acetate. The remaining DNA was then precipitated with isopropanol and resuspended in a tris/EDTA buffer.

2.2.5 In vitro manipulations of DNA

2.2.5.1 Restriction endonuclease digestion of DNA

All restriction digests were carried out using enzymes supplied by New England Biolabs (NEB) according to the manufacturers instructions. Briefly, 0.1-2 μg of purified DNA was incubated with 10-20 U of restriction enzyme in the appropriate NEB buffer for 1 hour at the appropriate temperature. Digests with multiple enzymes were carried out in the recommended double digest buffer or in an appropriate buffer in which all enzymes had 100 % activity. Where no suitable buffer was available sequential digestions were performed.

2.2.5.2 Purification of DNA fragments

Following digestion with restriction endonucleases or PCR amplification linear DNA fragments were purified directly from solution or from an agarose gel slice using the GENECLEAN II kit (Q BIO gene). To purify DNA from solution 3 volumes of 6 M sodium iodide were added followed by 5 – 10 μl of Glassmilk, an aqueous suspension of silica particles, and the mixture incubated for 5 min at 56 °C. The DNA bound to the silica particles was harvested by centrifugation at 15,800 x g for 10 sec and resuspended in New Wash, an ethanol based wash solution. This wash step is performed twice more and following the final wash the pellet is air-dried and the bound DNA is eluted with Millipore water. Purification from an agarose gel slice uses a similar protocol with some additions. The gel slice was melted at 56 °C following addition of 4.5 volumes of sodium iodide and 0.5 volumes of TBE modifier, a concentrated salt solution. The soluble DNA was then purified as before.
2.2.5.3 Ligation of DNA fragments

The T4 DNA Ligase was used to catalyse the formation of a phosphodiester bond between the exposed 5' phosphate and 3' hydroxyl groups of linear DNA fragments in an ATP-dependent fashion (68). These ligations were carried out using the Quick Ligation kit (New England Biolabs) and were routinely used for the cloning of DNA fragments into appropriate plasmid vectors. 50 ng of digested vector DNA was mixed with a three-fold molar excess of digested insert in a total volume of 10 μl and mixed with 10 μl of 2 x Quick Ligation Reaction Buffer. 1 μl of Quick T4 DNA Ligase was added and the reaction was incubated at room temperature for 15 min. The ligated DNA molecules were then transformed into either calcium-chloride or electro-competent E. coli XL-1 blue cells. Where calcium-chloride competent cells were used a 10 μl sample of the ligation reaction was directly transformed. For electro-competent cells the ligated DNA was first purified as previously described (section 2.2.5.2) and concentrated by ethanol precipitation (section 2.2.5.5) prior to electroporation.

2.2.5.4 Agarose gel electrophoresis

DNA samples were visualized following separation on a 1 % agarose gel. Briefly, agarose (1 g) was added to 100 ml of 0.5 x TBE buffer (44.5 mM tris borate, pH 8.3, 1 mM EDTA) and heated to 100°C to dissolve the agarose. Ethidium bromide was added to a final concentration of 1 μg/ml and the molten gel was poured into a gel mould and allowed to set. DNA samples were prepared by adding an appropriate volume of 5 x sample loading buffer (25 mM tris pH 7.6, 30% (v/v) glycerol, 0.125% (w/v) bromophenol blue) and these samples were electrophoresised through the gel at 135 V for 60 min in 0.5 x TBE buffer. The separated DNA fragments were photographed while illuminated under UV light.

2.2.5.5 Purification of DNA samples from solution by ethanol precipitation

Ethanol precipitation of DNA was routinely used for the preparation of desiccated DNA samples for DNA sequencing and for the concentration of dilute DNA samples. Ethanol is added to the aqueous DNA sample to deplete the hydration shell surrounding the DNA molecules exposing the negatively charged phosphate groups. A cation is
then added to mask this charge and allow a precipitate to form. Sodium acetate was routinely used as the cation. Briefly, the DNA sample was mixed with 0.1 volumes of 3 M sodium acetate (pH 5.2) followed by 2.5 volumes of ethanol and incubated at \(-70 ^\circ C\) for 20 min to allow a precipitate to form. The precipitated DNA was then recovered by centrifugation at 21,000 x \(g\) for 30 min, washed with 500 \(\mu\)l of 70 % ethanol to remove contaminating salt and centrifuged again as before. The precipitated DNA was then air-dried. If an aqueous DNA solution was required, the precipitated DNA was dissolved in Millipore grade water at 56 \(^\circ C\) for 10 min.

### 2.2.6 Polymerase chain reaction

The polymerase chain reaction (PCR) was used for the amplification of specific DNA fragments for use in cloning reactions and for the confirmation of constructed plasmids and mutant alleles. PCR is based on the ability of certain thermostable DNA polymerases to synthesise a new DNA strand complementary to a provided single-stranded denatured DNA template when primed with specific complementary oligonucleotides (216). The procedure involves successive rounds of thermal denaturation of a double-stranded DNA template, hybridization of two complementary oligonucleotides (primers) and synthesis of the new DNA strand by the DNA polymerase. The primers are designed to be complementary to opposite strands at either end of the fragment to be amplified and orientated such that their 3' ends face each other. The new DNA strand is synthesised from provided dNTPs by the DNA polymerase in the presence of Mg\(^{2+}\). Each new strand acts as a template in further rounds of amplification and the procedure thus results in exponential amplification of the desired DNA fragment.

#### 2.2.6.1 Amplification of DNA by polymerase chain reaction

Two different DNA polymerases were used for the routine amplification of DNA fragments. Taq DNA polymerase (New England Biolabs), a recombinant purified thermostable polymerase from Thermus aquaticus YT-1, which is highly efficient but lacks the 3'-5' exonuclease activity necessary for error correction (proof-reading), was used in PCR reactions were the sequence accuracy of the amplified DNA was not
critical; for example, screening recombinant plasmids for inserts or screening for mutant alleles in purified chromosomal DNA. Where a high degree of sequence fidelity was required, for example to generate DNA fragments for cloning reactions, Platinum Pfx polymerase was used (Life Technologies). Pfx is also highly processive but retains the 3'-5' exonuclease activity necessary for proof-reading resulting in a significant decrease in the rate of nucleotide misincorporation.

Pfx PCR reactions were carried out according to the manufacturers instructions in an MJ Research PTC-200 peltier thermal cycler. Briefly, 5 µl of 10 x Pfx buffer was mixed with 0.3 mM of each dNTP, 0.3 µM of each primer, 1 mM MgSO₄, 100 pg - 10 ng of template DNA, 1 U of Platinum Pfx and ddH₂O to a final volume of 50 µl. The PCR reactions were transferred to the thermal cycler and incubated as follows:

1. Denaturation: 94 °C for 2 min
2. Denaturation: 94 °C for 30 sec
3. Oligonucleotide annealing: a temperature corresponding to the lowest melting temperature of the primer pair for 30 sec
4. Extension: 68 °C for 1 min per kilobase of expected DNA product
5. Repeat steps 2-4 for an additional 29 cycles

Taq PCR reactions were carried out as above with some exceptions. Taq buffer (10 mM tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, pH 8.3) was substituted for Pfx buffer in the reaction mixture, MgCl₂ (1-3 mM) was used in place of MgSO₄ and 1 U of Taq polymerase was added to each reaction. In the thermal cycler extension was carried out at 72 °C, the optimal temperature for Taq activity and an additional 72 °C for 10 min step was included at the end of the program to complete any unfinished products.

Purified plasmid, genomic DNA or cell lysates were used as templates in PCR reactions. Cell lysates were prepared by transferring a single bacterial colony to a PCR tube containing 20 µl Millipore grade water using a sterile wooden tooth-pick and incubating at 100 °C for 5 min. 1 µl of this crude lysate was used in each PCR reaction.
2.3 Analysis and Manipulation of Proteins

2.3.1 SDS-PAGE
Proteins were separated on discontinuous denaturing polyacrylamide gels by the method of Laemmli (140). Using this method, proteins are denatured in SDS and β-mercaptoethanol and separated on the basis of their size as they travel through a polyacrylamide gel towards the anode. SDS binds to most proteins in a constant weight ratio, masking their natural charge with its own negative charge and giving each protein a similar mass:charge ratio allowing separation on the basis of size rather than charge. The discontinuous gel is formed using buffers of differing composition and pH to firstly focus the separating proteins into narrow well-defined bands and then separate these focused proteins on the basis of their size.

2.3.1.1 Preparation of total cellular protein for SDS-PAGE analysis
The OD<sub>600nm</sub> of an overnight culture was measured and 1 ml of the culture was then harvested by centrifugation at 15,800 x g for 1 min. The pelleted bacteria were then resuspended in an appropriate volume of Laemmli buffer (75 mM tris-Hcl, pH 6.8, 20 % (v/v) glycerol, 2 % (w/v) SDS, 2 % (v/v) β-mercaptoethanol, 10 μg/ml bromophenol blue) such that the final concentration was 10 OD<sub>600nm</sub> units/ml. The samples were boiled at 100 °C for 5 min and stored at −20 °C. Prior to use, samples were thawed and boiled again to denature all proteins.

2.3.1.2 Sarkosyl enrichment of outer-membrane proteins
Bacterial lysates were enriched for outer-membrane proteins as previously described (53). This procedure is based on the ability of the detergent N-laurylsarcosine (sarkosyl) to disaggregate and solubilise protein and lipid components of the bacterial cytoplasmic membrane while, due to its similar charge density to LPS, leaving the LPS-containing outer-membranes intact and insoluble.
The OD$_{600nm}$ of an overnight culture was measured and 12 OD$_{600nm}$ units (corresponding to 4 ml of a culture with an OD$_{600nm}$ equal to 3) of bacteria were harvested by centrifugation at 6,000 x g for 10 min and resuspended in 600 μl of sonication buffer (10% sucrose, 50 mM trisCl (pH 7.5), 100 mM NaCl, 1mM EDTA) and lysed by sonication. Intact bacteria were removed by centrifugation at 9,300 x g for 5 min and the supernatants were incubated with 0.5% sarcosyl for 30 min with continuous mixing to solubilise the inner-membranes. The sarcosyl-insoluble fraction containing the outer-membranes was harvested by centrifugation at 21,000 x g for 30 min and resuspended in 100 μl of Laemmli buffer. As before samples were stored at -20°C and boiled for 5 min prior to use.

2.3.1.3 Electrophoresis of protein samples

Discontinuous gels were prepared using standard protocols (8) and 12 or 15 % polyacrylamide gels were routinely used. Typically 10 μl of each boiled Laemmli sample was loaded per well and broad range molecular weight markers (New England Biolabs) were included on all gels. Electrophoresis was carried out in a Bio-Rad Mini-Protean III gel tank at 200 V for 50 min. Gels were stained using coomassie brilliant blue R-250 (0.25 % coomassie brilliant blue R-250, 45 % (v/v) methanol, 10 % (v/v) acetic acid) and destained using coomassie destain solution (45 % (v/v) methanol, 10 % (v/v) acetic acid) or transferred to nitrocellulose or PVDF membranes for use in western immunoblots.

2.3.2 Production of an anti-Hek antiserum and purification of an MBP-Hek fusion protein

The pMAL protein fusion and purification system (New England Biolabs) was used to construct a plasmid where the hek gene was fused to the malE gene, encoding the maltose-binding protein MBP, and this fusion gene was transcribed under the control of an IPTG-inducible Pt$^{IE}$ promoter. This plasmid was used to produce large amounts of an MBP-Hek fusion protein allowing generation of an anti-Hek antiserum and large-scale purification of the fusion protein. The construction of this plasmid is described in detail in Chapter 3.
2.3.2.1 Hek antiserum

An anti-Hek antiserum was generated by immunising a rabbit with a crude insoluble protein preparation containing the MBP-Hek fusion protein. Briefly, a 100 ml culture of *E. coli* K-12 expressing the fusion protein was harvested by centrifugation at 6,000 x g for 10 min, resuspended in 5 ml of column buffer (20 mM tris-HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA) and lysed by sonication. Any insoluble material was then harvested by centrifugation at 21,000 x g for 30 min and when the soluble and insoluble fractions were analysed by SDS-PAGE the majority of the fusion protein was found to be present in the insoluble fraction. The proteins contained in the insoluble fraction were solubilised by resuspension in 8 M urea and the solution was desalted using a PD-10 column (Amersham Biosciences). The column eluate was examined by SDS-PAGE and the protein concentration estimated. This crude preparation was then used to immunise a rabbit. The initial immunization consisted of 200 μg protein in Freund’s complete adjuvant with two further boosts containing 200 μg of protein in Freund’s incomplete adjuvant administered on days 14 and 28. A high antibody titre was detected by western immunoblotting after 38 days and the rabbit was exsanguinated. The resultant antiserum was absorbed against a lysate of *E. coli* K-12 expressing MBP to remove cross-reacting antibodies. The lysate was prepared from a 500 ml culture which was harvested, resuspended in 5 ml column buffer and lysed by sonication. The lysate was divided in two and one half was boiled at 100 °C for 5 min. The boiled and untreated lysates were recombined, mixed with an equal volume of antiserum and incubated overnight at 4 °C. Insoluble materials including protein-antibody complexes were removed by centrifugation at 15,800 x g for 30 min. This absorption was repeated a further two times. The absorbed antiserum was purified using a protein A-agarose affinity column (Sigma-Aldrich) according to the manufacturers instructions and stored in aliquots at -20 °C.

2.3.2.2 Pilot purification of MBP-Hek

The fusion protein was over-expressed in *E. coli* and a pilot purification carried out according to the manufacturers instruction. An overnight culture of *E. coli* K-12 containing the fusion plasmid was diluted 1:100 in 80 ml fresh broth and grown to an OD<sub>600nm</sub> of between 0.4 and 0.6. A 1 ml sample of culture was taken and resuspended in
Laemmli buffer as previously described (section 2.3.1.1), the remainder of the culture was induced by the addition of IPTG to a final concentration of 0.3 mM and incubated for an additional 2 h. A second sample was taken and resuspended in laemmli buffer and the remaining bacteria were harvested by centrifugation at 6,000 x g for 10 min and resuspended in 5 ml column buffer (20 mM tris-HCl, pH 7.4, 200 mM NaCl, 1 mM EDTA). The bacteria were lysed by sonication and the insoluble material was removed by centrifugation at 15,800 x g for 5 min. Samples of the soluble supernatant and insoluble pellet were retained and resuspended in Laemmli as before. Amylose resin (50 μl) was washed twice with column buffer in a microcentrifuge tube and mixed with an equal volume of the soluble supernatant. The mixture was incubated on ice for 20 min, harvested, washed twice with column buffer and resuspended in Laemmli buffer. The uninduced, induced, insoluble, soluble and amylose-bound protein samples were analysed by SDS-PAGE.

2.3.2.3 Large-scale purification of MBP-Hek

A 1 l induced culture of E. coli K-12 harboring the malE-hek fusion plasmid was harvested, resuspended in 50 ml column buffer and a soluble lysate was prepared. An affinity column was prepared by pouring amylose resin (15 ml) into a 2.5 x 10 cm column and the column bed was washed with column buffer (120 ml). The soluble lysate was diluted 1:5 in column buffer and loaded onto the column at a rate of approximately 1 ml/min. The column was then washed with column buffer (180 ml), the bound protein was eluted with 6 ml column buffer containing 10 mM maltose and collected in 10 fractions. The 10 fractions were analysed by SDS-PAGE and the protein containing fractions were pooled and stored at 4 °C.

2.3.3 Western immunoblotting

2.3.3.1 Electro-transfer of separated proteins

Proteins were separated on SDS-polyacrylamide gels and transferred to nitrocellulose or PVDF membranes using a Biometra Fastblot semi-dry transfer apparatus. For transfer to nitrocellulose membranes the polyacrylamide gel and the membrane were
equilibrated in transfer buffer (25 mM tris, 192 mM glycine, 20 % (v/v) methanol) and blotting was carried out according to the manufacturers instructions at 5 mA/cm² for 20 min. Where higher sensitivity of detection was required Immobilon-P PVDF membrane (Millipore) was used and transfer was carried out using a three buffer semi-dry system according to the manufacturers instructions. Following transfer all membranes were rinsed briefly with Millipore grade water and stained with Ponceau S stain (0.5 % (w/v) Ponceau S, 1 % (v/v) acetic acid) to visualise the transferred proteins. Excess ponceau stain was removed by further washes with Millipore purified water before subsequent manipulation of the membranes. Nitrocellulose membranes were blocked by incubation in blocking buffer (5 % non-fat powdered milk in phosphate-buffered saline) at 4 °C overnight prior to use. Immobilon-P PVDF membranes were dried overnight at room temperature according to the manufacturers instructions prior to use.

2.3.3.2 Detection of bound proteins

Blocked nitrocellulose membranes or dried PVDF membranes were incubated with anti-Hek antiserum diluted 1:2,000 in appropriate blocking buffer (5 % non-fat powdered milk in PBS for nitrocellulose or 3 % non-fat powdered milk in PBS containing 0.01 % Tween-20 for PVDF membranes) for 1 h at room temperature. The membrane was then washed 3 x 5 min with PBS and incubated with a secondary HRP-linked anti-rabbit antibody diluted 1:20,000 in blocking buffer for either 1 h (nitrocellulose) or 30 min (PVDF). The blot was then washed as before and developed using the SuperSignal West Pico chemiluminescent HRP substrate (Pierce). Chemiluminescence was detected using BioMax Light Film (Kodak) and films were developed manually using standard Kodak developer and fixer solutions according to the manufacturers instructions. The length of exposure was varied to adapt for the signal strength. Typical exposure lengths were 1 - 5 min for nitrocellulose and 10 sec – 1 min for PVDF membranes.

2.3.4 Flow-cytometry

To examine the surface exposure of the Hek protein intact bacterial cells were labelled with anti-Hek polyclonal antiserum and a fluorescently labelled secondary antibody and
analysed by flow-cytometry. To remove any further cross-reacting antibodies present in the anti-Hek antiserum the antiserum was mixed an equal volume of an overnight culture of *E. coli* K-12 and incubated for 12 h at 4 °C. Bacteria with bound antibody were removed by centrifugation at 15,800 x g for 10 min. This absorbed antiserum was then used in flow-cytometry. Briefly, approximately 5 x 10⁷ bacteria from an overnight cultures were harvested at 18,000 x g for 10 min, washed 3 times with PBS (with centrifugation as before) and resuspended in PBS containing 2 % v/v formaldehyde. Excess formaldehyde was removed by three successive washes with PBS and the bacteria were then incubated overnight with a 1:10 dilution of absorbed anti-Hek antiserum at 4 °C. The labelled bacteria were then washed three times with PBS and incubated with a 1:100 dilution of secondary FITC-conjugated anti-Rabbit antibody for 1 h at room temperature. The labelled samples were then washed again as before and fluorescence was analysed on a Beckman Coulter Epics XL flow cytometer according to the manufacturers instructions.

2.3.5 2-D PAGE

2.3.5.1 Sample preparation

Overnight bacterial cultures were diluted 1:100 into fresh broth and grown to an OD₆₀₀nm of approximately 1.0. 100 ml of culture was harvested by centrifugation at 4,600 x g for 10 min at 4°C and the pellet was resuspended in 10 ml of a low-salt wash buffer (3.0 mM KCl, 1.5 mM KH₂PO₄, 68 mM NaCl, 9.0 mM NaH₂PO₄) (www.expasy.org). The suspension was centrifuged again as before and washed a further three times with low-salt wash buffer. After the final wash the bacterial pellet was resuspended in 120 µl of resuspension buffer (10 mM tris-HCl (pH 8.0), 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, Complete mini (protease inhibitor cocktail, Roche), 0.1% (w/v) SDS) and the bacterial samples were stored at -20°C in 10-20 µl aliquots.

2.3.5.2 IPG strip loading

Complete solubilisation of proteins in the bacterial sample was ensured by incubation in a solution containing a high concentration of urea, thiourea and the detergent CHAPS.
Prior to loading onto an IPG strip each bacterial sample was thawed and 1 μl was mixed with 125 μl of solubilisation buffer (7 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 65 mM DTT, 0.0005 % (w/v) bromophenol blue) and incubated at room temperature for at least 3 hours. 1.25 μl of 20 % Biolytes (BioRad) was added to each sample (final 0.2 %) and incubated for a further hour at room temperature. Each sample was centrifuged at 10,000 x g for 5 minutes to remove insoluble matter and the supernatant was transferred into a Biorad rehydration tray. A 7 cm Proteogel 4-7 IPG gel strip (Sigma-Aldrich) was gently placed gel-side down onto each sample and allowed to rehydrate overnight at room temperature. In order to prevent evaporation of samples during the overnight incubation a small volume of water (200-300 μl) was added to unused lanes in the rehydration tray and the lid was sealed in place with parafilm. This method allows IPG strips to rehydrate overnight without an overlay of mineral oil and with a minimum of sample evaporation.

2.3.5.3 First dimension isoelectric focussing

Paper wicks were placed onto the electrodes of a Biorad focussing tray and each was wet with 5 μl of ddH₂O. Rehydrated IPG strips were placed gel-side down onto the paper wicks and 1 ml of mineral oil was added over each strip to prevent evaporation. The focussing tray was placed into a Biorad IEF cell and isoelectric focussing was carried out with a maximum of 50 μA per strip and a constant temperature of 20°C under the following conditions: rapid-ramp to 150 V and hold for 300 volt-hours, followed by a rapid-ramp to 6,000 V and hold for a total of 40,000 volt-hours.

2.3.5.4 Equilibration and second dimension SDS-PAGE

Second dimension polyacrylamide gels (12 % resolving gels without a stacking gel) were prepared using standard methods (8) and allowed to polymerise fully during the first dimension IEF step. Immediately following first dimension isoelectric focussing each IPG strip was incubated in 2.5 ml of equilibration buffer (6 M urea, 50 mM tris-HCl (pH 8.0), 30 % (v/v) glycerol, 2 % (w/v) SDS, 0.0005 % (w/v) bromophenol blue) containing 2.5 % (w/v) DTT for 10 min at room temperature followed by 2.5 ml of equilibration buffer containing 2.5 % (w/v) iodoacetamide for a further 10 min. Each strip was then dipped briefly in 1xSDS-PAGE running buffer and placed onto a
second dimension polyacrylamide gel. To ensure even contact between the IPG strip and the acrylamide gel each strip sealed in place with agarose overlay solution (molten 1 % agarose cooled to 55°C). Electrophoresis was carried out in a Bio-Rad Mini-Protean III gel tank at 200 V for approximately 40 min and gels were stained as before (Section 2.3.1.3). Stained gels were imaged using a BioRad GS800 calibrated densitometer and spot identification and analysis was carried out using the PDQuest™ software package (BioRad).

2.3.6 Enzyme assays for beta-galactosidase and alkaline phosphatase

Plasmids were constructed where the $lacZ$ gene, encoding beta-galactosidase, or $phoA$, encoding alkaline phosphatase, were placed under the control of the putative $hek$ promoter. The beta-galactosidase activity was measured using a standard Miller assay (168) and alkaline phosphatase activity was measured using a modified Miller assay utilising an alternative chromogenic substrate. To assay beta-galactosidase activity an overnight culture was diluted 1:10 in Z buffer (60 mM Na$_2$HPO$_4$.7H$_2$O, 40 mM NaH$_2$PO$_4$.H$_2$O, 10 mM KCl, 1 mM MgSO$_4$, 50 mM β-mercaptoethanol) in a total volume of 1 ml. Chloroform (50 μl) and 0.1 % SDS (25 μl) were added, the mixture was vortexed thoroughly to permeabilise the bacterial cells and incubated at 28 °C for 5 min. The chromogenic substrate, ONPG, (200 μl, 4 mg/ml) was added to start the reaction and 1 M Na$_2$CO$_3$ (500 μl) was added to stop the reaction when sufficient colour had developed. The optical density of the reaction was measured at 420 nm to measure the development of colour from cleaved ONPG and at 550 nm to correct for light scatter caused by cell debris. The β-galactosidase activity was then calculated using the following equation:

$$\text{β-galactosidase activity (Miller Units)} = \frac{1000 \times (\text{OD}_{420\text{nm}} - 1.75 \times \text{OD}_{550\text{nm}})}{t \times v \times \text{OD}_{600\text{nm}}}$$

where $t$ is the duration of the reaction in min and $v$ is the volume of culture used in the reaction expressed in ml.

The assay for alkaline phosphatase activity was performed in the same manner with slight variations. Bacterial cultures (1 ml quantities) were harvested and resuspended in cold 1M tris-HCl (pH 8.0) and lysed as before. The chromogenic substrate, pNPP (100
μl, 3.7 mg/ml), was added to start the reaction and the reaction was stopped addition of 0.5 M K$_2$HPO$_4$ (200 μl, pH 8.0). Colour development was measured as before. Each assay was performed in duplicate and repeated at least twice to ensure reproducibility.
2.4 Phenotypic assays

2.4.1 Haemagglutination assay

The ability of bacterial strains to agglutinate erythrocytes was determined using a 1% suspension of human blood. Briefly, sterile human blood was diluted 1:10 with PBS, centrifuged at 6,000 x g for 10 sec to harvest intact erythrocytes and resuspended in the same volume of PBS. This wash step was repeated until the supernatant was clear. Immediately prior to use, this 10% suspension of erythrocytes was diluted 1:10 with PBS containing 100 mM mannose. Mannose was included to inhibit agglutination due to type 1 fimbriae. Overnight bacterial cultures were harvested by centrifugation at 15,800 x g for 1 min and resuspended in an equal volume of PBS. The OD₅₀₀nm was measured and the bacteria were diluted with PBS such that their OD₅₀₀nm was equal to 1.0. The bacteria were then serially 2-fold diluted with PBS in a final volume of 100 μl in a 96-well microtitre plate. To each well was added an equal volume of 1% blood suspension and the plate was incubated at room temperature for 2 hours or at 4 °C overnight to allow un-agglutinated erythrocytes to settle out of suspension.

2.4.2 Autoaggregation assays

2.4.2.1 Quantification of the rate of autoaggregation

To examine the characteristic autoaggregation phenotype associated with the expression of the Hek protein in E. coli K-12 strains an assay was developed to measure the rate at which bacterial cells aggregated and settled out of suspension by measuring the decrease in optical density, measured at 600 nm, at the surface of the cultures. Overnight cultures (10 ml) were harvested by centrifugation and resuspended in 5 ml PBS. The OD₅₀₀nm were measured and adjusted to approximately 4.0 by the addition of PBS. 5 ml of each culture was then transferred to a Kahn tube, a 50 μl sample was taken and the OD₅₀₀nm measured again to determine the starting OD for each culture (T=0). Further 50 μl samples were taken from the surface of the cultures every 20 min and the OD₅₀₀nm were measured as before. Assays were performed in duplicate and the
rate of autoaggregation was determined by the mean decrease in optical density over time.

2.4.2.2 Competitive autoaggregation assays
To determine if autoaggregation was due to a homologous Hek-Hek interaction or a heterologous interaction between Hek and some other surface component a competitive autoaggregation assay was developed. Two overnight cultures of *E. coli* K-12, one carrying a Hek expression plasmid and the other carrying the empty vector, were prepared as before and mixed together in 1:1 ratio. The mixed culture was transferred to a Kahn tube, a 100 µl sample was taken and dilutions were spread on differential LB agar plates containing X-gal and IPTG to determine the relative proportions of the two strains in the mixture, as bacteria containing the empty vector would appear blue on these plates. The rate of autoaggregation was then monitored as before and after 4 h a second 100 µl sample was taken from the surface of the mixture and dilutions were spread on differential plates. The change in the relative proportions of the two strains was used to determine whether all of the bacteria or only those expressing Hek were settling out of suspension.

2.4.3 Adhesion, cell association and invasion assays

2.4.3.1 Qualitative adhesion assays
In order to perform qualitative adhesion assays T84 cells were seeded into chamber slides (Nalge Nunc International, 0.8 cm² per well) at a density of approximately 1.0 x 10⁵ cells per chamber and grown for 7 days at 37°C in 5 % CO₂. Bacterial strains were grown overnight and 6 µl quantities were added to each well of cultured T84 cells. Infected monolayers were then incubated for 6 h at 37°C in 5 % CO₂ and then washed extensively with PBS to remove non-adherent bacteria. The chamber slides were then disassembled and the mammalian cells with adherent bacteria were fixed in methanol for 5 min. The slides were then stained in freshly prepared Giemsa stain (5 % v/v methanol, 2 % v/v Giemsa stock solution) for 1 h at room temperature and excess stain
was removed by briefly rinsing with ethanol. Slides were air-dried and examined microscopically.

2.4.3.2 Quantitative cell association and invasion assays

CHO-K1, pgsA-745 and T84 cultured cells were infected with *E. coli* bacteria and the total numbers of cell-associated and intracellular bacteria were enumerated. All cell lines were grown in 12 well trays in the appropriate tissue culture medium and fed once on the day before the experiment. T84 cells were split 7 days prior to infection at a density of $2.5 \times 10^5$ per well and CHO-K1 and pgsA-745 cells were split 3 days prior to infection at densities of $3.0 \times 10^4$ and $6.0 \times 10^4$ per well respectively. Bacterial strains were grown overnight, harvested by centrifugation and resuspended in PBS. The OD$_{600nm}$ were measured and bacteria were diluted 1:500 in warm tissue culture medium. Mammalian cell monolayers were washed once with warm PBS and bacteria-containing medium (1 ml) was added to each well. The infected cells were centrifuged at 600 g for 5 min to initiate contact between bacteria and the mammalian cells and incubated at 37 °C in 5 % CO$_2$ for 1 h to allow adhesion to and invasion of the cultured cells. Samples of the medium containing bacteria were also diluted and spread on Columbia-blood agar plates to determine the numbers of bacteria present in each inoculum. The infected cells were then washed twice with warm PBS to remove any non-adherent bacteria. To determine the total number of cell-associated bacteria the monolayer was disrupted by treatment with 0.1 % Triton x-100 and the released bacteria were enumerated by spreading dilutions on Columbia-blood agar. To determine the number of intracellular bacteria a standard gentamicin protection assay was performed (114). Following the 1 h infection the cells were incubated with medium containing gentamicin (100 µg/ml) for 90 min at 37 °C in 5 % CO$_2$ to kill any adherent extracellular bacteria, washed with PBS, disrupted with 0.1 % Triton x-100 and the released bacteria enumerated as before.

In experiments where exogenous glycosaminoglycans were tested as potential inhibitors of invasion the bacterial inocula were prepared essentially as before but in media containing the inhibitor to be tested and incubated for between 30 min and 1 h at 37 °C prior to infecting the cultured monolayers. The experiments were then carried out as previously described.
Chapter 3  Phenotypic analysis of the Hek protein of

*Escherichia coli* K1
3.1 Introduction

3.1.1 *E. coli* K1 pathogenesis

A prerequisite to invasive bacterial infection is the ability to interact with and colonise the host’s epithelium. For bacteria whose niche is the human gastrointestinal (GI) tract this process requires acid resistance during gastric transit, resistance to killing by bile salts and antimicrobial peptides, the ability to penetrate the mucosal layer and attach to the epithelial cells lining the gastrointestinal tract, all tasks in which *E. coli* excels. These abilities are common to both commensal and pathogenic strains alike and have made *E. coli* one of the principal members of the human intestinal flora. The pathogenic strains of *E. coli* described in Chapter 1 have evolved further abilities in their quest to exploit new niches; in particular, many pathogenic *E. coli* have developed the ability to invade and survive within human cells (123). Of interest here are those strains of *E. coli* capable of causing meningitis in neonates, collectively known as neonatal-meningitic *E. coli* or NMEC. In order to traverse from the gut to the cerebrospinal fluid surrounding the brain, NMEC must first successfully colonise and invade the epithelium, survive uptake by professional macrophage in order to disseminate throughout the body and finally colonise and invade the microvascular endothelial cells which form the blood-brain barrier (30, 111, 113, 157, 165). As outlined in Chapter 1 a great deal is known about survival in the blood stream and penetration of the blood-brain barrier but little study has focussed on the initial stages of infection. Invasion and transcytosis of cultured GI epithelial cells by *E. coli* K1 strains has been demonstrated (30) and a single study utilising signature-tagged mutagenesis and a neonatal rat model identified 16 genes necessary for colonisation (157) but to date no gene has been specifically implicated in the invasion of GI epithelial cells.

3.1.2 Bacterial adhesins

To facilitate the interaction between a bacterium and the host cell the bacterium must express surface components with the ability to specifically interact with the host cell surface directly or with other bridging molecules which can in turn interact with the cell
surface (63). Long-range interactions are often facilitated by fimbrial adhesins (34, 35), which can project far beyond any bacterial LPS or capsular coat which can serve to mask shorter non-fimbrial adhesins (228, 263). More intimate interactions are subsequently formed using non-fimbrial adhesins (128, 205). These are often integral outer membrane proteins of the β-barrel structural family in gram-negative organisms (136). The specificity of such interactions varies widely allowing for tropisms to specific tissue types or interactions with a broad range of mammalian cells. For example, P type fimbriae bind specifically to Galα(1,4)Gal moieties on glycolipids of urinary tract epithelial cells and erythrocytes (241), whereas type 1 fimbriae can mediate attachment to mannose-containing receptors on many different host cell types (159).

3.1.3 Bacterial invasion

Bacteria colonising a human or animal host are exposed to a wide variety of stresses from the environment and the innate and adaptive immune response. One strategy which has evolved to allow bacteria avoid these stresses is the invasion of normally non-phagocytic cells. Although the intracellular niche does allow immune evasion and a certain degree of protection from environmental stresses it does present additional challenges for a bacterium, including innate cellular defences (122) and intracellular nutrient scarcity. Many different bacterial pathogens, including several E. coli pathotypes (see Chapter 1), have evolved mechanisms to induce their uptake by non-phagocytic cells and these can be divided into two mechanistic categories, the Zipper and Trigger mechanisms (39). The Zipper mechanism, typified by Yersinia and Listeria spp., involves the clustering of receptors in the host membrane around an adherent bacterium. As the clustered receptors interact with bacterial ligands farther around the surface of the bacterium the host membrane is drawn around the attached bacterium, forming a phagocytic cup, eventually meeting and fusing behind the bacterium. In Yersinia the bacterial ligand is invasin (114), whereas Listeria express at least two distinct ligands, InlA (internalin) and InlB (38). The Trigger mechanism is best represented by Salmonella and Shigella spp., which use type III secretion systems (TTSS) to inject effector molecules into the host cytosol. It is the action of these effectors, usually through cytoskeletal rearrangements, which induces membrane
extensions and ultimately the phagocytosis of the bacteria. The *Salmonella* TTSS is encoded on SPI-1 (*Salmonella* pathogenicity island-1) (83) whereas *Shigella* use a plasmid-encoded TTSS (see Chapter 1, section 1.1.2.6) (160).
3.2 Results

3.2.1 Identification of the hek gene in *E. coli* K1 strain RS218

NMEC strains produce several well-characterised fimbrial adhesins, including the type 1 S- and P-fimbriae (3, 118, 138), which have been implicated in infection by the closely related uropathogenic *E. coli* (UPEC) strains (25, 159). In addition, NMEC strains also express a number of membrane proteins that are necessary for invasion of endothelial cells in certain model systems. These membrane proteins, including OmpA, IbeA and B, and AslA, are described in detail in Chapter 1. As a partial unfinished genome sequence is available for *E. coli* K1 strain RS218 (www.genome.wisc.edu) a search for novel virulence factors was carried out. The nucleotide sequences of virulence determinants from several different bacterial species, including *Yersinia* spp., *Neisseria* spp., *Shigella* spp. and other virulent *E. coli* pathotypes, were used to perform BLASTN searches (www.ncbi.nih.gov) of the RS218 incomplete genome database (Table 3.1). Several genes were identified with between 35 and 41 % identity to known virulence determinants and a single gene was identified with 62 % identity to the gene encoding the Tia invasin from ETEC strain H10407. Tia is a well-characterised adhesin and invasin which has been shown to be necessary for invasion by ETEC and sufficient to promote invasion by a non-invasive *E. coli* K-12 strain (74). This RS218 gene was also found to be 91 % identical to the heat-resistant agglutinin, *hra*, gene from porcine *E. coli* strain O9:H10:K99 (153) and was hence designated heat-resistant agglutinin from *E. coli* K1, hek (Fig. 3.1). This was the first *hra*-like gene identified in a human-pathogenic strain of *E. coli*, however the gene has since been found to be widely distributed in UPEC strains and has been statistically linked to virulence in those strains (243). The high degree of identity between Hek, Hra1 and Tia suggests that the protein may form an 8-stranded β-barrel in the bacterial outer membrane (154) and this structural prediction will be dealt with in more detail in Chapter 4.

3.2.2 Cloning and expression of hek in *E. coli* K-12
3.2.2.1 Expression of hek from its native promoter

The hek gene from E. coli K1 strain RS218 and 76 bp of upstream DNA was amplified by PCR and cloned into the pBluescript II SK + cloning vector (pBSKII) cut with EcoRV. The resulting plasmid was sequenced to confirm the insert and designated pHEK6 (Table 2.2 and Fig. 3.2, Smith, S. G. J., unpublished). SDS-PAGE analysis of total cellular protein from a strain of E. coli K-12 harbouring the pHEK6 plasmid revealed a unique protein band with an apparent molecular weight of approximately 26 kDa, corresponding to the predicted size of the Hek protein (Fig. 3.3 A). This unique 26 kDa band was present in the insoluble fraction of sarkosyl-extracted protein preparations suggesting that the protein is likely to be present in the outer membrane was also recognised by an anti-Hek antiserum (Fig. 3.3 B and C). The 76 bp of upstream DNA was examined using the MacTargSearch program and a putative α70 promoter sequence with an associated Shine-Delgarno sequence was identified (Fig. 3.2 B, Smith, S. G. J., unpublished) and the expression of the Hek protein in strains harbouring this plasmid suggest that the gene can be transcribed from this promoter.

3.2.2.2 Construction of a low-copy Hek-expression plasmid

Due to a variable expression phenotype from the pHEK6 plasmid when transformed into the E. coli K1 strain RS218 background, which will be discussed in detail in Chapter 5, and to allow expression of Hek at levels close to those observed in RS218, the hek-containing insert from pHEK6 was transferred into a low-copy vector, pPD101 (Table 2.1) (54). The entire insert and some flanking DNA from pHEK6 was excised by digestion with KpnI and SacI which flank the multiple-cloning site of the pBSKII backbone. This fragment was ligated with pPD101 cut with KpnI and SacI yielding plasmid pPDhek22 (Fig. 3.3 D). Expression of Hek protein in strains harbouring this plasmid was confirmed by western immunoblot analysis (Fig. 3.3 E).

3.2.2.3 Construction of the IPTG-inducible pThek6 plasmid

In order to examine the Hek protein independent of its native promoter the hek open reading frame (ORF) was transferred from pHEK6 into the pTrc99a expression vector where transcription of the gene is placed under the control of an IPTG-inducible P^rec promoter. The hek ORF from plasmid pHEK6 was amplified by PCR using primers
TABLE 3.1 BLASTN searches of incomplete *E. coli* RS218 genome sequence

<table>
<thead>
<tr>
<th>Invasin</th>
<th>Species of origin</th>
<th>Gene name in <em>E. coli</em> K1</th>
<th>E value</th>
<th>% Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tia</td>
<td>ETEC</td>
<td><em>hek</em></td>
<td>$4e^{-47}$</td>
<td>62%</td>
</tr>
<tr>
<td>Inv</td>
<td><em>Yersinia</em> enterocolitica</td>
<td><em>eaeH</em></td>
<td>$5e^{-92}$</td>
<td>37%</td>
</tr>
<tr>
<td>EaeA</td>
<td>EPEC</td>
<td><em>eaeH</em></td>
<td>$5e^{-86}$</td>
<td>35%</td>
</tr>
<tr>
<td>Ail</td>
<td><em>Yersinia</em> enterocolitica</td>
<td><em>ompX</em></td>
<td>$3e^{-31}$</td>
<td>41%</td>
</tr>
<tr>
<td>Rck</td>
<td><em>Salmonella</em> typhimurium</td>
<td><em>Ecs1807</em></td>
<td>$2e^{-25}$</td>
<td>41%</td>
</tr>
<tr>
<td>OpcA</td>
<td><em>Neisseria</em> gonorrhoea</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>IpaC</td>
<td><em>Shigella</em> flexneri</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Figure 3.1 Clustal alignment of the primary sequences of the Hek, Hra1 and Tia proteins from *E. coli* strains RS218, O9:H10:K99 and H10407 respectively. Residues with a black background are identical and grey signifies conservative substitutions. Gaps are indicated with hyphens.
Figure 3.2 Construction of the pHEK6 plasmid. 

**A:** Representative map of the plasmid pHEK6 showing the positions of the *hek* and *bla* (encoding beta-lactamase) genes and the important restriction endonuclease recognition sites.

**B:** DNA sequence of the pHEK6 insert. The region of *E. coli* strain RS218 genomic DNA which was cloned into pBSKII is shown. The positions of putative promoter −35 and −10 boxes are indicated in **blue** and **red** respectively, whilst the Shine-Delgarno sequence (RBS) is in **pink** and the *hek* gene is shown in **black**. Additional flanking DNA is in **green**.
Figure 3.3 Expression of the Hek protein in *E. coli* K-12 from high and low copy plasmids. SDS-PAGE analysis of whole-cell lysates (panel A) or sarcosyl-insoluble fractions (panel B) of *E. coli* K-12 strain XL-1 harbouring pHEK6 (Hek+) or the vector control pBSKII (Hek-). The position of the Hek protein is indicated with and this protein was also detected using an anti-Hek antiserum (panel C). A representative plasmid map showing the salient features of the pPDhek22 plasmid is shown in panel D and expression of Hek protein in *E. coli* K-12 harbouring this plasmid (Hek+) compared to the vector control pPD101 (Hek-) was confirmed by western immunoblot analysis of whole cell lysates (panel E). A broad-range protein molecular mass standard (MW) was included on all gels.
RF1 and hekrm (Table 2.3). The forward primer, RF1, incorporated an additional CGCC at the 5' end of the ORF to introduce a new Ncol restriction site and the reverse primer, hekrm, was complementary to the pBSKII plasmid multiple-cloning site (MCS) 3' of the pHEK6 insert. The resulting PCR product was digested with BamHI and ligated into pBSKII cut with BamHI and HincII yielding plasmid pBhek7 (Fig. 3.4 A). The hek ORF with the new 5' Ncol site was excised from pBhek7 by digestion with Ncol and BamHI and the hek-containing fragment was cloned into the vector pTrc99a cut with Ncol and BamHI yielding plasmid pThek6 (Table 2.2 and Fig. 3.4 B). To confirm inducible expression of the Hek protein, a strain of E. coli K-12 harbouring the pThek6 plasmid was grown to mid-logarithmic phase, induced with IPTG and total cellular and sarkosyl-extracted protein samples were analysed by SDS-PAGE and western immunoblot (Fig. 3.4 C and D). A 26 kDa protein corresponding to Hek was visible in samples of uninduced cultures and this protein was highly over-expressed upon induction with IPTG. The identity of this protein was confirmed by western immunoblot analysis using the anti-Hek antiserum (Fig. 3.4 E). A second slower migrating species was also detectable after induction for 2 hours and this second form was exclusively found in the sarkosyl-insoluble fraction. This additional protein species may represent the membrane-inserted portion of the protein which is perhaps resistant to denaturation as a consequence of its stable tertiary structure. The relatively low amount of this protein species detected may be a result of the considerable time required to correctly fold a β-barrel protein (20 - 30 min at 37°C for OmpA) (250). Hek protein was also detectable in uninduced cultures and this agrees with the reports that the P<sub>lac</sub> promoter displays a low level of (leaky) transcription even in the absence IPTG (see pTrc99a description in the Amersham Biosciences catalogue 1999 no. 275007-01).

3.2.2.4 Construction of a malE-hek fusion expression plasmid and purification of the MBP-Hek fusion protein

Outer membrane proteins are often insoluble when over-expressed and thus can be difficult to purify. Fusions of the maltose binding protein, MBP, to amino-termini of target proteins has been found to enhance solubility of over-expressed proteins (124) and allow straightforward purification of fusion proteins on an amylose resin affinity column (127). The pMAL-c2 plasmid was used to construct an inducible malE-hek fusion allowing purification of the resulting approximately 67 kDa MBP-Hek fusion
protein. The hek DNA sequence corresponding to the mature Hek protein, lacking the 22 amino acid signal sequence, was amplified from plasmid pHEK6 with primers hekfm and hekrm (Table 2.3). The forward primer, hekfm, is identical to the first 17 nucleotides of the mature protein DNA sequence and the reverse primer is complementary to the pBSKII MCS as described above (section 3.2.2.3). The PCR product was cut with BamHI, cloned into pMAL-c2 digested with XmnI and BamHI resulting in an in-frame fusion to the malE gene and the resulting plasmid was designated pMhek1 (Table 2.2 and Fig. 3.5 A). SDS-PAGE analysis of E. coli K-12 harbouring the pMhek1 plasmid induced with IPTG revealed three inducible protein bands with apparent molecular weights of between approximately 55 and 66 kDa (Fig. 3.5 B). The upper band corresponded to the predicted size of the MBP-Hek fusion protein. As described in Materials and Methods, approximately 70% of the produced MBP-Hek was found to be insoluble and this insoluble protein was used to immunise a rabbit and generate anti-Hek antibody. The remaining 30% of soluble protein was purified on an amylose resin column and analysed by SDS-PAGE. The intermediate protein species (approximately 60 kDa) was present in the soluble fraction but not the insoluble fraction. However, this intermediate form was not present in the amylose column eluate, implying either that it did not bind the amylose resin or did not elute from the column. Aliquots of purified protein were found to migrate as two distinct species on an SDS polyacrylamide gel with apparent molecular weights of 55 and 66 kDa (Fig. 3.5 C) and both species were recognised by the anti-Hek antiserum (Fig. 3.5 D). In a given sample the relative quantities of the two species were found to vary depending on the temperature of denaturation and a third, slower migrating, form was also detected in samples treated with urea (Fig. 3.3 E) and the identity of all three forms was confirmed by western immunoblotting (not shown). This suggests that these apparently different protein species actually represent the MBP-Hek protein in different denatured forms. This possibility was supported when a sample of the purified MBP-Hek fusion protein, which resolves as two distinct protein species on an SDS-polyacrylamide gel, was analysed by MALDI-TOF mass spectrometry revealing a single protein species with a monoisotopic mass of 67,479 (S. G. J. Smith, personal correspondence). Multiple protein species were also detected by western immunoblot analysis of E. coli K-12 harbouring both the pHEK6 and pThek6 plasmids (Figs. 3.3 C and 3.4 E) indicating that the Hek protein itself displays a high degree of resistance to
Figure 3.4 Expression of the Hek protein from an inducible $P_{trc}$ promoter. Panels A and B: Plasmid maps of pBhek7 and pThek6. Panels C and D: SDS-PAGE analysis of whole-cell lysates (C) and sarcosyl-insoluble fractions (D) of uninduced and IPTG-induced cultures of E. coli K-12 strain XL-1 harbouring pThek6 with the Hek protein marked (<). Panel E: Western immunoblot analysis of E. coli K-12 harbouring pThek6, uninduced (1), induced for 1 h (2), induced for 2 h (3) and induced for 2 h sarcosyl-insoluble fraction (4) or the induced vector control pTrc99a (5).
Figure 3.5 Structure of pMhek1 and expression of an MBP-Hek fusion protein. Panel A: Plasmid map of the malE-hek fusion plasmid pMhek1 encoding the MBP-Hek fusion protein. Panel B: SDS-PAGE analysis of uninduced and induced cultures as well as soluble and insoluble fractions of the induced culture, induced protein bands are highlighted with <. Panel C: SDS-PAGE analysis of uninduced and induced cultures and two purified fractions containing MBP-Hek. Panel D: Western immunoblot analysis of an uninduced culture and soluble and insoluble fractions of the induced culture. Panel E: SDS-PAGE analysis of an induced culture insoluble fraction subjected to denaturation at 70 or 100°C or in 8 M urea prior to electrophoresis.
denaturation. In particular, the slower migrating species is exclusively observed in the sarkosyl insoluble fraction of bacteria harbouring the pThek6 plasmid (Fig. 3.4 E) as discussed in the previous section. This suggests that it may be the highly stable nature of the folded Hek protein which results in the apparent resistance to denaturation.

3.2.3 *E. coli* K-12 expressing Hek interact with human erythrocytes

The *E. coli* heat-resistant agglutinin, Hra1, has been shown to promote agglutination of human erythrocytes *in vitro* and this ability was found to be resistant to heat treatment of the bacteria prior to incubation with the erythrocytes. The primary sequence of the Hek protein from *E. coli* K1 strain RS218 is 91 % identical to that of Hra1 (153), suggesting that Hek may promote a similar haemagglutination activity. In order to test this, laboratory strains of *E. coli* K-12 harbouring the pHEK6 plasmid or the empty vector pBSKII were incubated with human erythrocytes as described in Materials and Methods. The relative abilities to promote haemagglutination were indicated by the highest bacterial dilution to give positive agglutination (Fig. 3.6 A). The Hek protein was found to confer a strong haemagglutination phenotype similar to that described for the Hra1 protein (153) and, as with Hra1, this phenotype was found to be resistant to heat treatment (Fig. 3.6 B). Following incubation at 70 °C for 20 min bacteria expressing the Hek protein were still capable of agglutinating human erythrocytes to the same extent as unheated bacteria. This heat resistant phenotype is undoubtedly related to the Hek protein’s strong resistance to denaturation as observed by SDS-PAGE analysis and discussed in the previous section. It is also possible to inhibit this Hek-associated haemagglutination phenotype by pre-incubating the human erythrocytes with purified MBP-Hek fusion protein (Fig. 3.6 C).

3.2.4 Expression of Hek in *E. coli* K-12 promotes autoaggregation

When high-density liquid cultures of *E. coli* K-12 harbouring the pHEK6 plasmid were allowed to stand it was observed that the cultures appeared to clear as the bacteria rapidly settled to the bottom (Fig 3.7 A). This phenomenon has been termed auto-aggregation and is due to the formation of bacterial aggregates following interactions between the bacteria in culture. This phenotype has been described previously for several *E. coli* proteins including Ag43, FimH and Cah and has been implicated in the
formation of biofilms (44, 227, 253). As that bacteria clump and settle to the bottom the density at the surface decreases. This allows the rate of aggregation to be observed by taking samples from the surface and measuring the decrease optical density (OD) at 600 nm over time. In cultures expressing Hek the surface OD decreases by 50 % in less than 60 min compared with only 10 – 15 % in 3 hours with the control strain (Fig. 3.7 B). The rate of aggregation is proportional to the density of the culture so as the culture density decreases so too does the rate of aggregation. This was confirmed by performing parallel autoaggregation assays with cultures normalised to a range of starting optical densities (2.0, 4.0 and 8.0) (Fig. 3.7 C). Auto-aggregation involves direct interactions between surface components of the bacteria. To investigate whether the aggregation phenotype associated with expression of the Hek protein was due to interactions between homologous interactions between Hek molecules on neighbouring cells or heterologous interactions between Hek and some other surface component, competitive aggregation assays were performed. Mixed cultures, containing equal amounts of Hek+ and Hek- E. coli K-12, were prepared at optical densities of 4.0 and 8.0 and the rate of aggregation was compared with pure cultures of each at optical densities of 2.0, 4.0 and 8.0. If Hek promotes aggregation by homologous interactions a mixed culture would aggregate more like a pure culture with a lower optical density. However, if Hek is involved in heterologous interactions with other bacterial surface components a mixed culture would aggregate more like a pure culture of the same density. Mixed cultures were found to aggregate at similar rates to Hek+ pure cultures of the same starting density (Fig. 3.8 A and B) suggesting that Hek is capable of interacting with additional E. coli K-12 surface components.

In addition to measuring the changing optical densities of cultures, the pure and mixed cultures were also examined by spreading diluted samples on agar plates containing the chromogenic beta-galactosidase substrate X-gal and the lac inducer IPTG. This allows the different bacterial populations to be differentiated based on the colour of their colonies. Bacteria harbouring the pBSKII plasmid will give rise to blue colonies whereas those harbouring pHEK6 will give white colonies. In this way the decreases in optical density in mixed cultures can be attributed to one or both of bacterial populations. Pure and 1:1 mixed cultures of E. coli K-12 harbouring pHEK6 (Hek+) or pBSKII (Hek-) were prepared at initial optical densities of 4.0 and samples were taken
**Figure 3.6** Haemagglutination of *E. coli* K-12 expressing Hek. Cultures of *E. coli* K-12 strain XL-1 harbouring plasmids pHEK6 (Hek+) or pBSKII (Hek-) were normalised to an optical density of 1.0 at 600 nm (titre=1) and serially two-fold diluted. An equal volume of 1 % human blood containing 100 mM mannose was added to each well and the assay was read after 2 h at room temperature or overnight at 4 °C. Assays were performed with fresh overnight cultures (panel A) or cultures were incubated at 70 °C for 20 min prior to the assay (panel B). Panel C: Protein inhibition was carried out using purified MBP-Hek fusion protein with titre 8 or 16 heat-treated bacteria used in panel B. A sample of purified MBP-Hek protein was serially two-fold diluted in PBS and added to the bacterial suspensions prior to addition of blood. Final protein concentration was 6 μg/ml in the first lane (titre 1). PBS alone was added to a lane as negative control (-).
Figure 3.7 Hek-promoted autoaggregation. Panel A: Settling out of cultures of *E. coli* K-12 strain XL-1 harbouring the pHEK6 (Hek+) or pBSKII (Hek-) plasmids. Panel B: The rate of settling of cultures at an optical density of 4.0 at 600 nm was quantified by plotting the decrease in surface optical density of cultures against time. Panel C: The effect of density on the rate of aggregation was examined by monitoring aggregation of cultures with a range of starting optical densities (2.0, 4.0 and 8.0).
**Figure 3.8** Analysis of Hek-promoted autoaggregation in mixed cultures. Pure cultures of *E. coli* K-12 harbouring pHEK6 or pBSKII or 1:1 mixed cultures of the two were normalised to optical densities of 4.0 (panel A) or 8.0 (panel B) at 600 nm and the surface density was monitored over 3 h.

**TABLE 3.2** Quantitative analysis of Hek-promoted autoaggregation in *E. coli* K-12

<table>
<thead>
<tr>
<th>Time sampled (min)</th>
<th>Pure cultures</th>
<th>1:1 mixed culture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hek&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Hek&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Hek&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Hek&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>Time = 0</td>
<td>1.59 (±0.08)</td>
<td>1.82 (±0.13)</td>
</tr>
<tr>
<td>Time = 180</td>
<td>0.75 (±0.01)</td>
<td>1.53 (±0.08)</td>
</tr>
<tr>
<td>% decrease</td>
<td>53.8 %</td>
<td>15.9 %</td>
</tr>
</tbody>
</table>

1, 2 - *E. coli* K-12 strain XL-1 harbouring pHEK6<sup>1</sup> or pBSKII<sup>2</sup>
to determine the total number of bacteria in each culture and the relative proportions of Hek+ and Hek- in the mixed cultures. The cultures were left undisturbed for 4 h to allow aggregation to occur and were again sampled. Pure cultures of *E. coli* K-12 harbouring pBSKII aggregated considerably slower than those harbouring pHEK6 (Table 3.2), 15.9% decrease in 4 h with pBSKII compared with 53.8% with pHEK6. Mixed cultures displayed a similar decrease to pure Hek+ cultures (53.7% decrease) and significantly the Hek- bacteria in mixed cultures aggregated as rapidly as those expressing Hek (53.4% and 53.7% respectively). This finding strongly suggests that the Hek protein promotes autoaggregation through interactions with other bacterial surface components.

3.2.5 Qualitative adhesion assays on Hek-expressing *E. coli* using T84 colonic epithelial cells

As bacteria expressing the Hek protein can agglutinate human erythrocytes it was likely that Hek could promote interactions with other mammalian cell types and this may play a role in the colonisation of human gastrointestinal epithelia. This hypothesis was supported by the high degree of similarity between Hek and other well-characterised adhesins such as Tia from ETEC. To investigate the role for the Hek protein in promoting adhesion to human epithelial cells, the T84 cell line, derived from a human colonic carcinoma, was used (55, 176). T84 cells retain an epithelial cell morphology and are widely used as a model of the human colonic epithelium (81, 147, 177, 275). T84 cells were grown at 37°C in 5% CO₂ on glass slides and incubated with *E. coli* K-12 strain XL-1 harbouring the pHEK6 plasmid or the vector control pBSKII. Non-adherent bacteria were washed away and the infected monolayers were stained with Giemsa to visualise the mammalian cells and any attached bacterial cells. The enteroaggregative *E. coli* strain 042 was included as a control as it displays strong adhesion to many mammalian cell types forming distinctive bacterial aggregates or micro-colonies on the mammalian cell surface (177). When cells were incubated with *E. coli* K-12 carrying the vector control a few isolated bacteria were observed adhering to the cultured monolayer (Fig 3.9 A). However, when cells were incubated with the Hek-expressing bacteria many more adherent bacteria were observed and often in clusters on the cell surface (Fig. 3.9 B). Interestingly, the bacteria expressing Hek
appear to cluster on particular cells (approximately 1 in every 100) suggesting that the Hek receptor may not be generally well expressed in this cell line. These clusters were very similar to the characteristic micro-colonies observed when cells were incubated with the EAEC strain (Fig. 3.9 C). These data suggested that the Hek protein increased the adherence of *E. coli* K-12 and perhaps facilitated the formation of micro-colonies through the cell-cell interactions as suggested by the auto-aggregation phenotype.

Despite indicating a possible role for Hek in adhesion to cultured epithelial cells the qualitative assay described above did not allow this adhesive ability to be accurately quantified and did not allow for a more detailed examination of the interaction between Hek expressing bacteria and the mammalian monolayer. To investigate the bacterial-cell interactions in greater detail, cell-association and gentamicin survival assays were employed as described in section 2.4.2.2.

3.2.6 Quantitative cell-association and invasion assays on Hek-expressing *E. coli*

3.2.6.1 Cloning of the Tia invasin from ETEC

The Tia adhesin and invasin from ETEC strain H10407 was used as a control in all cell-association and invasion assays. The pET125 plasmid bearing the *tia* gene from ETEC strain H10407 was obtained from J. M. Fleckenstein. *E. coli* K-12 bacteria carrying this plasmid were previously shown to express a 25 kDa protein corresponding to the predicted size of the Tia protein and were also highly invasive in several epithelial cell lines including T84 and CHO-K1 cells (74). Invasion of *E. coli* K-12 strain DH5α harbouring the pET125 plasmid into T84 cells was confirmed (DH5α/pET125 0.0113 % compared with DH5α/pBR322 0.0026 %) although the observed levels of invasion were approximately 300-fold lower than those previously reported in the literature (0.0113 % compared with approximately 3.4 %). In order to allow a direct comparison with the pHEK6 plasmid the *tia*-containing insert from pET125 was sub-cloned into the pBSKII vector. Briefly, The entire 2,000 bp *Clal-Sall* insert from pET125 was excised and ligated with pBSKII cut with *Clal* and *Sall*. The structure of the resulting plasmid was confirmed by restriction endonuclease mapping, sequenced to confirm the presence.
Figure 3.9 Qualitative analysis of Hek-promoted adherence. Photomicrographs of Giemsa stained T84 monolayers which were incubated with *E. coli* K-12 strain XL-1 harbouring pBSKII (panel A), pHEK6 (panel B) or EAEC strain O42 (panel C).
of the \textit{tia} gene and designated pTia5 (Table 2.2). \textit{E. coli} K-12 bacteria harbouring the pTia5 plasmid were also found to invade T84 cells with an efficiency considerably higher than the negative control. This plasmid was used as the positive control in all subsequent experiments involving \textit{E. coli} K-12.

### 3.2.6.2 Interaction of Hek- or Tia-expressing \textit{E. coli} with CHO-K1 epithelial cells

Chinese hamster ovary (CHO-K1) cells have an epithelial cell morphology and are readily invaded by strains of pathogenic \textit{E. coli} including ETEC (74). Although CHO-K1 cells are not human in origin and so are not a valid model for studying the specifics human disease, they are easily cultured and are often used in receptor identification studies as several derivative mutant cell lines are available (213) which are defective in the production of various cell-surface components.

To determine if Hek could also promote invasion of this cell line, cell-association and gentamicin survival assays were performed with confluent monolayers of CHO-K1 cells as described in Materials and Methods. In \textit{E. coli} K-12 strain XL-1 Hek was found to promote levels of cell association very similar to those observed with Tia (Fig. 3.10 A). However, Hek was found to promote considerably higher levels of invasion than Tia. This trend was observed in every assay, with \textit{E. coli} K-12 expressing Tia approximately 50 \% of cell-associated bacteria were found to be intracellular whereas in bacteria expressing Hek greater than 80 \% were found to be intracellular. These data were confirmed in \textit{E. coli} K-12 strain DH5\(\alpha\) and suggest that although both proteins can mediate attachment to cultured CHO-K1 cells, Hek is the more efficient at promoting invasion. The invasion efficiency of \textit{E. coli} K-12 harbouring the low-copy \textit{hek} plasmid pPDhek22 was also analysed and was found to be similar to that observed with pHEK6, however the vector control pPD101 was found to increase the level of background invasion by the \textit{E. coli} K-12 strain (Fig. 3.10 B).

### 3.2.6.3 Invasion of human T84 cells by \textit{E. coli} K-12 expressing Hek or Tia

The ability of Hek to promote cell-association and invasion of cultured T84 cells was tested as before and compared with the positive control Tia. \textit{E. coli} K-12 strain DH5\(\alpha\) harbouring plasmid pHEK6 displayed levels of cell-association 6.7-fold greater than the
strain with the vector control pBSKII but 4.5-fold lower than the Tia-expressing strain (Fig. 3.11 A). In invasion assays using the same strains, expression of Hek resulted in levels of invasion 4.3-fold higher than the background and 3-fold lower than the positive control Tia (Fig. 3.11 B). These results were also confirmed in a second E. coli K-12 strain, XL-1 (not shown).

3.2.6.4 Construction of an E. coli K1 hek mutant

In order to examine the effect of a mutation in the hek gene on the invasion of E. coli K1 into T84 cells an interrupted hek allele was constructed on a plasmid and transferred to the chromosome of strain RS218 using the λ Red allele replacement system as described in section 2.2.3.2. A 2.2 kb fragment encoding spectinomycin resistance was excised from plasmid pH45Ω (Table 2.1) by digestion with HindIII and the resulting 5' overhangs were filled in using the DNA polymerase I, large (Klenow) fragment enzyme. The resulting blunt-ended DNA fragment was ligated to pHEK6 cut with BsaBI (Fig. 3.1 A). The structure of the resulting plasmid was confirmed by restriction endonuclease mapping and designated pHSPC4 (Table 2.1). Bacteria harbouring pHSPC4 produced no detectable Hek protein and did not auto-aggregate or agglutinate human erythrocytes (Fig. 3.12 A). A linear DNA fragment containing the disrupted hek gene was generated by PCR amplification using primers hekfim and hekrORF (Table 2.3) using pHSPC4 as a template. Following amplification the template was digested with the restriction enzyme DpnI which is specific for methylated DNA and thus will not digest a PCR product. The linear PCR product was then purified and approximately 2 μg was transformed into electrocompetent arabinose induced RS218 harbouring the λ Red plasmid pKOBEGERA (Table 2.2). Putative mutants were confirmed by PCR and a single confirmed mutant was retained and designated E. coli strain RShek (Fig. 3.12 B, Table 2.1). The mutant strain produced no detectable Hek protein when analysed by western immunoblot (Fig. 3.12 C).

3.2.6.5 Invasion of human T84 cells by an E. coli K1 hek mutant

Invasion and transcytosis of T84 cells by E. coli K1 strain RS218 has been previously described (30). In order to determine if the Hek protein plays any role in the invasion of this cell line by E. coli RS218, wild-type and hek mutant bacteria, were compared in
Figure 3.10 Interaction of *E. coli* K-12 expressing Hek or Tia with CHO-K1 cells. Panel A: Cell association and invasion by *E. coli* K-12 strain XL-1 harbouring pHEK6 (Hek+), pTia5 (Tia+) or the vector control pBSKII. Panel B: Invasion efficiencies of *E. coli* K-12 strain XL-1 harbouring the low-copy number plasmid pDhek22 (Hek+) or its vector control pD101 compared with the strain harbouring the pBSKII vector.
Figure 3.11 Interaction bacteria expressing Hek or Tia with human colonic epithelial cells. Cell association (panel A) and invasion (panel B) of human colonic T84 cells by *E. coli* K-12 strain DH5α harbouring pHEK6 (Hek+), pTia5 (Tia+) or the vector control pBSKII.
Figure 3.12 Construction of an *E. coli* K1 RS218 *hek* mutant. Panel A: Haemagglutination by *E. coli* K-12 harbouring pHEK6, pHSPC4 (interrupted *hek* gene) or the vector control pBSKII. Panel B: PCR analysis of the *hek* gene in the wild type *E. coli* K1 strain RS218 (WT) and the isogenic *hek* mutant RShek (Δ*hek*). Panel C: western immunoblot analysis of Hek protein expression in the wild type *E. coli* RS218 and the *hek* mutant RShek.
gentamicin survival assays. The level of invasion from experiment to experiment was found to be highly variable as has been noted elsewhere (66, 74), allowing direct comparisons only between data from simultaneous experiments. However, the wild-type *E. coli* strain RS218 was found to be considerably more invasive than the isogenic *hek* mutant and this trend was confirmed in repeated experiments (RS218 0.122 %, RShek 0.026 %).

### 3.2.6.6 Complementation of the *hek* mutation in strain RShek

In order to confirm that the mutation in the *hek* gene was responsible for the apparent invasion defect displayed by isogenic *hek* mutant strain RShek it was necessary to complement this mutation by expressing Hek from a plasmid. In order to complement the invasion defect observed with T84 cells, *E. coli* strain RShek was transformed with the low-copy pPDhek22 plasmid and invasion of T84 cells by the complemented strain was compared with that of RShek harbouring the empty vector, pPD101. The presence of the pPDhek22 plasmid resulted in considerably higher levels of invasion than the empty vector (0.572 % compared with 0.252 %) and this trend was supported in repeated experiments using pPDhek22. It was also noted that the pPD101 vector appeared to increase the background level of invasion in the RShek strain in a manner similar to that observed with *E. coli* K-12 strain XL-1 (see section 3.2.6.2).
3.3 Discussion

The pathogenesis of *E. coli* K1 neonatal meningitis has been studied in great detail for a number of years and a great deal is now known about the establishment of septicaemia, survival and dissemination in the blood stream and infection of the blood-brain barrier. However, relatively little is known about the initial stages of infection including colonisation and invasion of the gastrointestinal epithelium. A number of bacterial proteins have been implicated in epithelial colonisation using a library of signature-tagged transposon-mutants in a neonatal rat model (157) and transcytosis of cultured human epithelial cells by the clinical *E. coli* K1 isolate RS218 has been demonstrated *in vitro* (30) but to date no bacterial factor has been specifically implicated in the invasion of human intestinal epithelial cells. In general, *E. coli* adhesion to host cells is initially mediated through long-range interactions by fimbrial adhesins (34, 228), which can easily extend beyond the physical and electrostatic barrier presented by surface lipopolysaccharide and capsular polysaccharide that can mask shorter adhesins (152, 230, 263)(M. Lambert, personal correspondence). More intimate interactions can then be formed using non-fimbrial adhesins, often integral outer membrane proteins (74, 77). Many of these adhesins are structurally incapable of extending beyond the LPS and capsular coat and yet are still capable of interacting with mammalian cell surface components (47). In some cases the existence of bridging molecules such as mammalian heparin binding proteins (MHBPs) has been hypothesised (73) and it is also conceivable that the mammalian receptor may itself be large enough to extend across this physical barrier. A third possibility is that the bacteria have some mechanism to modulate this barrier by altering the expression or composition of the LPS and capsule either across the entire cell surface or locally to create focal points for adhesion. One example of this was recently reported in UPEC where the initial long-range interactions between type 1 fimbriae and D-mannose receptors was shown to down-regulate transcription of the *kps* region 1 capsular assembly operon resulting in a decrease in the surface expression of the K1 capsule (230). Several non-fimbrial adhesins have already been implicated in the invasion of brain microvascular endothelial cells by *E. coli* K1, including OmpA and the Ibe proteins (110, 113, 204). The aim of this study was to identify novel outer membrane proteins with a role in the invasion of human
gastrointestinal epithelial cells and to examine what role, if any, is played by the other well-characterised adhesins expressed by \textit{E. coli} K1.

A gene with a high degree of identity to the \textit{tia} invasion locus of enterotoxigenic \textit{E. coli} (ETEC) strain H10407 was identified in the unfinished genome sequence of \textit{E. coli} K1 strain RS218 (www.genome.wisc.edu). This gene also shared 91 % identity with the \textit{hral} gene from uropathogenic \textit{E. coli} (UPEC) encoding the heat-resistant agglutinin and was hence designated heat-resistant agglutinin from \textit{E. coli} \textit{K1}, \textit{hek}. Expression of \textit{Hral} confers a strong haemagglutination phenotype that is highly resistant to heat treatment at 70 °C for 20 min prior to the assay. The \textit{tia} gene, which shares 62 % identity with \textit{hek}, has been implicated in adhesion to and invasion of cultured mammalian epithelial cells and has been shown to interact specifically with mammalian cell surface heparan sulphate proteoglycans (73). All three proteins are predicted to form 8-stranded β-barrels in the outer membrane (See Chapter 4). In order to study the function of the Hek protein a number of expression vectors were constructed in which the \textit{hek} gene was transcribed either from its own promoter or from a foreign inducible promoter and also as a fusion with the \textit{malE} gene, encoding the \textit{E. coli} maltose-binding protein. With these tools it was possible to examine the effect of expressing Hek in a non-pathogenic lab strain of \textit{E. coli} K-12 and also to raise a polyclonal antibody against the MBP-Hek fusion protein.

In \textit{E. coli} K-12 the Hek protein was found to confer a strong mannose-resistant haemagglutination phenotype with human erythrocytes and this phenotype was also found to be highly resistant to heat-treatment as with the very similar \textit{Hral} protein. Haemagglutination is not thought to be relevant to pathogenicity \textit{in vivo} but it does indicate an ability to specifically interact with human cells and as such is a useful initial assay for a novel adhesin. Further experiments using cultured mammalian cells were necessary to investigate these interactions in greater detail. Initial qualitative adhesion assays were carried out using cultured T84 human colonic epithelial cells and expression of the Hek protein appeared to enhance the adhesion of bacteria to this cell line. Furthermore, bacteria expressing Hek were observed to form clumps or microcolonies on the surface of cultured T84 cells similar to those formed by enteroaggregative \textit{E. coli}. These Hek-expressing bacteria were also observed to rapidly
aggregate and settle out of static liquid suspensions, indeed bacterial aggregates were even evident in shaken cultures. This phenomenon, known as autoaggregation, has been described for several other *E. coli* proteins including Antigen 43 (Ag43), a calcium-binding protein Cah, the type 1 fimbrial adhesin FimH and the diarrhoeagenic *E. coli* adhesin AIDA and has been implicated in the formation of biofilms (44, 227, 235, 253). The formation of bacterial biofilms requires the initial attachment of bacteria to a surface and subsequent interactions between bacteria to allow a three dimensional community to form (209). Taken together these data suggest a dual role for Hek in colonising cell surfaces and promoting the formation of biofilm-like microcolonies that may assist in other aspects of the infectious process such as invasion and resistance to environmental stress. The rate of Hek-promoted autoaggregation was also found to be proportional to the density of the culture. This allowed a closer examination of the mechanism of aggregation through the use of mixed-culture assays. Mixed cultures containing approximately equal numbers of Hek+ and Hek- *E. coli* K-12 were found to aggregate at a rate similar to that of a pure Hek+ culture with the same starting density. This suggested that the Hek protein induced aggregation through heterologous interactions with another bacterial surface component rather than with Hek molecules on neighbouring cells. To confirm this finding, the proportions of Hek+ and Hek- bacteria in mixed cultures were directly quantified before and after aggregation. Hek- bacteria in mixed cultures were found to aggregate as rapidly as those expressing Hek and considerably faster than Hek- bacteria in pure cultures. This clearly demonstrated that Hek is capable of interacting with some other bacterial surface component and that it is this interaction which is responsible for the Hek-promoted autoaggregation phenotype.

The ability of Hek to promote haemagglutination was unsurprising in light of the high degree of identity between Hek and Hra1. The significant similarity between Tia and Hek further suggested that Hek may also play a role in adhesion to and perhaps invasion of epithelial cells. To examine the Hek-promoted interactions with cultured cells in greater detail cell-association and gentamicin survival (invasion) assays were carried out using the human T84 cell line and the hamster-derived CHO-K1 cell line. Both cell types are epithelial in morphology and T84 cells are regularly used as a model of the human colonic epithelium (81, 177). The expression of either Hek or the positive
control Tia in *E. coli* K-12 dramatically increased the invasion efficiency in the T84 cell line (4.3- and 12.8-fold respectively) although expression of Tia resulted in considerably higher levels of invasion than Hek. With either protein the numbers of total cell-associated bacteria were significantly higher than the numbers of internalised bacteria with less than 1 in 20 of the adherent bacteria successfully invading a mammalian cell indicating that invasion may be saturable. Furthermore, these results show that of the approximately 5 x 10^5 mammalian cells per well only approximately 1 in 3,000 has an internalised bacterium. This suggests a relative scarcity of the receptors for Hek and Tia on the surface of T84 cells. This possibility was supported by the findings of the earlier qualitative adhesion assays using T84 cells which showed clusters of Hek-expressing bacteria clustering on certain cells in a monolayer but not others. This suggests that not all cells in the population are expressing the Hek-receptor or that the receptor is not accessible on all cells due to polarisation effects in the monolayer. Alternatively the relatively low numbers of internalised bacteria may be due to the defence mechanisms of the cultured epithelial cells, including the production of antimicrobial peptides (269). In contrast, expression of Hek or Tia in *E. coli* K-12 resulted in highly efficient invasion of cultured CHO-K1 cells compared with bacteria harbouring the empty vector (66- and 43-fold respectively). When the cell association data is compared with the results of the gentamicin survival assays it appears that Hek is the more efficient invasin in CHO-K1 cells with on average greater than 80% of cell-associated bacteria being recovered following gentamicin treatment compared with approximately 50% with Tia. As expression of Hek results in considerably higher levels of invasion than Tia in CHO-K1 cells but less in T84 cells this suggests that Hek and Tia recognise different receptors and that the differences in invasion between the two mammalian cell lines is due to differences in the relative expression of these receptors.

A mutation in the *hek* gene of *E. coli* K1 strain RS218 resulted in a dramatic decrease in the invasion efficiency in T84 cells and this invasion defect could be complemented using the low-copy number pPDhek22 plasmid which resulted in considerably higher levels of invasion than the vector control. However, the empty vector pPD101 appeared to increase the invasion frequency of the *hek* mutant strain and this
phenomenon was also observed in the *E. coli* K-12 strain XL-1. Without further data it is impossible to speculate on the ability of this plasmid to increase invasivity.
Chapter 4  Identification of the Hek-receptor and structure-function analysis of the Hek protein of *Escherichia coli* K1
4.1 Introduction

4.1.1 Proteoglycans

Proteoglycans are ubiquitous molecules on the surface of mammalian cells, consisting of a protein core covalently linked to glycosaminoglycan (GAG) chains, which contain repeats of alternating amino sugars and uronic acids. Based on the chemical nature of these repeats GAGs are defined as dermatan sulphate, chondroitin sulphate or heparan sulphate. Of interest here are the two major GAGs found on intestinal epithelial cells: heparan sulphate (HS), based on the disaccharide repeat (GlcUAβ1-4GlcNAcα1-4), and chondroitin sulphate (CS), based on the repeat (GlcUAβ1-3GalNAcβ1-4). Despite this seemingly simple chemical structure GAGs display an enormous amount of structural heterogeneity with large variations in chain length, glycosylation, sulfation and epimerisation, resulting in many thousands of distinct chemical species.

Proteoglycans are described on the basis of their protein core and the proteoglycans which are generally found on the surface of mammalian epithelial cells are divided into two families, the syndecans and glypicans. The syndecans (syndecan 1-4) are transmembrane proteins with an extended extracellular domain, which varies greatly among the members of the family, and are encoded by four distinct genes that appear to have diverged from a single ancestral gene following gene duplication. The glypicans (glypican 1-6) are anchored to the membrane by a covalent linkage to phosphatidyl inositol and are encoded by at least six genes in mammals. Many microbial pathogens have been shown to utilise interactions with mammalian cell proteoglycans to aid in the colonisation of their host (for a review see (213)). In particular, previous studies with the Neisseria gonorrhoeae invasion-associated opacity (Opa) protein and the ETEC Tia protein concluded that these proteins interact with heparan sulphate proteoglycans (HSPGs) on cell surfaces. In the syndecans the sites of attachment for HS chains are located at the N-terminal end of the protein, which extends a considerable distance from the membrane, and some may also have a CS attachment site closer to the membrane. In contrast to this the glypicans only contain binding sites for HS chains and these are located close to the membrane surface, the remainder of the protein is predicted to form a large globular
structure farther from the membrane, stabilised by numerous disulphide linkages (48). Due to their extended extracellular structure the syndecans appear to be the HSPG of choice for attaching microorganisms (78, 85, 101, 260). The CHO-K1 cell line has become the most useful tool in the study of bacterial interactions with proteoglycans (213). A number of derivative cell lines have been developed with characterised mutations in various enzymes involved in the synthesis of GAGs and the assembly of proteoglycans. One such CHO-K1-derivative, pgsA-745, carries a mutation in the pgsA gene encoding a xylosyltransferase enzyme and is completely deficient in the production of glycosaminoglycans (70). These cells still produce the proteoglycan core proteins but have no linked GAG chains. As such, this cell line represents a relatively blunt tool for the initial characterisation of a GAG binding protein. Further characterisation is then possible using other cell lines and even purified GAG molecules in inhibition studies.
4.2 Results

4.2.1 Identification of the Hek receptor

4.2.1.1 Invasion of proteoglycan deficient CHO-K1 cells by E. coli K-12 expressing Hek or Tia

The Tia protein from ETEC strain H10407 has been shown to interact with heparan sulphate proteoglycans (73) and the Tia and Hek proteins are quite similar in their primary sequences (62 %, see Chapter 3, Fig. 3.1). However, the surface-exposed loops of Hek and Tia are only 48 % identical and this falls to 38 % if only the first three loops are considered. Considering these significant differences in the primary sequences of the surface-exposed loops it was intriguing to examine if Hek also mediates attachment to mammalian cells via interactions with cell surface proteoglycans. Gentamicin protection assays were performed with CHO-K1 cells and a glycosaminoglycan deficient CHO-K1-derivative cell line, pgsA-745. The level of invasion of E. coli K-12 expressing Hek or Tia was dramatically reduced in the pgsA-745 cell line compared with invasion of CHO-K1 cells (Fig. 4.1 A), 23-fold for Hek-expressing bacteria and 13-fold for Tia. However the level of invasion by bacteria harbouring the vector control pBSKII was also considerably reduced in pgsA-745 cells suggesting that this E. coli K-12 strain, XL-1, may have some inherent proteoglycan binding ability. Furthermore, bacteria expressing Hek or Tia were still approximately 20-fold more invasive than the negative control in pgsA-745 cells suggesting that both Hek and Tia may also interact with some other cell surface component. The invasion efficiency of the clinical E. coli strain RS218 into CKO-K1 and pgsA-745 cells was also assayed and a large decrease in invasion was revealed in the pgsA-745 cell-line compared with CHO-K1 cells (pgsA-745 0.032 % compared with CHO-K1 0.135 %).

4.2.1.2 Inhibition of invasion of CHO-K1 cells with purified glycosaminoglycans

In order to examine the potential interactions between the Hek protein and the GAG chains on mammalian proteoglycans a series of gentamicin survival assays were carried
Figure 4.1 The contribution of GAGs to the invasion of CHO-K1 cells by *E. coli* K-12 expressing Hek or Tia. Panel A: Invasion of CHO-K1 and pgsA-745 cells by *E. coli* K-12 strain XL-1 harbouring pHEK6 (Hek+), pTia5 (Tia+) or pBSKII. Panel B: Inhibition of Hek-promoted invasion by the addition of exogenous GAGs. Heparin, heparan sulphate, chondroitin sulphate or the negative control dextran sulphate were added to bacteria 30 min prior to the gentamicin protection assay to determine the ability of each to inhibit invasion. The invasion efficiencies of each treated bacterial sample are expressed as a percentage of the untreated Hek-expressing control.
Figure 4.2 Inhibition of Hek-promoted invasion with exogenous glycosaminoglycans.

Panel A: Heparin (0.1, 1, 10 or 100 μg/ml) was added to cultures of *E. coli* K-12 harbouring pHEK6 45 min prior the invasion assay and these treated cultures were compared with untreated *E. coli* XL-1 harbouring pHEK6 or pBSKII as positive and negative controls. Panel B: Heparin, heparan sulphate, chondroitin sulphate or the negative control dextran sulphate (10 μg/ml of each) were added to cultures of *E. coli* K-12 harbouring pHEK6 45 min prior the invasion assay and these treated cultures were compared with untreated *E. coli* XL-1 harbouring pHEK6 or pBSKII. The invasion efficiencies in each experiment were expressed as a percentage of the untreated positive control.
and a number of purified glycosaminoglycans were tested for an inhibitory effect on the invasion of \textit{E. coli} K-12 expressing Hek. Prior to the assay bacteria were incubated in standard tissue-culture medium with 100 \(\mu\text{g/ml}\) of heparin, heparan sulphate (a less highly sulphated form of heparin) or chondroitin sulphate. The polysaccharide dextran sulphate was included as a negative control as it has a similar mass-charge ratio as heparin (197). Both heparin and heparan sulphate were found to strongly inhibit Hek-promoted invasion, heparin by 5.4-fold and heparan sulphate by 15.8-fold. In contrast, chondroitin sulphate and dextran sulphate resulted in reductions of only 2.4- and 2.6-fold respectively (Fig. 4.1 B). However, heparan sulphate was also found to strongly reduce the level of invasion of the control strain, \textit{E. coli} K-12 harbouring pBSKII, with little or no effect from the other three inhibitors. As a result the inhibition by heparin appears to be the more significant. To examine this in greater detail the inhibitory effect of heparin at a number of different concentrations was determined. The inhibitory effect was found to be dose-dependent and very sensitive, with considerable inhibition (2.5-fold) still apparent with as little as 100 ng/ml of heparin (Fig. 4.2 A). The experiment with all three GAGs was then repeated with 10 \(\mu\text{g/ml}\) of each and with dextran sulphate once again included as a negative control. Heparin was the only inhibitor to show a significant inhibitory effect with a 7.8-fold decrease in the level of invasion (Fig. 4.2 B). Heparan sulphate, chondroitin sulphate and dextran sulphate resulted in decreases of only 1.6-, 1.4- and 1.6-fold respectively. Interestingly, when GAG inhibition assays were performed with the \textit{E. coli} K1 strain RS218 and CHO-K1 cells, heparin was found to slightly reduce cell-association (3.94 ±0.045 %, RS218 alone; 2.8 ±0.229 %, RS218 with 100 \(\mu\text{g/ml}\) heparin) but had no effect on invasion. Heparan sulphate, chondroitin sulphate and dextran sulphate had no effect on either cell-association or invasion.

\section*{4.2.2 Prediction of the structure of the Hek protein}

\subsection*{4.2.2.1 Analysis of the amino-terminal sequence of the mature Hek protein}

Analysis of the predicted primary sequence of the Hek protein revealed the presence of a putative amino-terminal signal sequence similar to those described for proteins translocated across the inner membrane via the Sec system (see Chapter 1, section 1.2).
This prediction was supported by sequence similarity with other *E. coli* proteins including the Tia invasin from enterotoxigenic *E. coli* (74) and Hra1 from porcine *E. coli* strain O9:H10:K99 (153). The Hek amino-terminus displays the characteristic hallmarks of a Sec leader sequence; a 22-residue sequence with a basic amino-terminus followed by a stretch of 16 mostly hydrophobic amino acids and alanine residues in the –3 and –1 (residues 20 and 22 of the preprotein) positions. To experimentally confirm the cleavage of the Hek preprotein between residues 22 and 23, a sarcosyl-insoluble fraction of *E. coli* K-12 harbouring the inducible pThek6 plasmid (Table 2.1 and Fig. 3.2 B) was separated on an SDS-polyacrylamide gel and transferred to a nitrocellulose membrane. A 26 kDa protein unique to bacteria expressing Hek and assumed to be the mature Hek protein was excised and subjected to peptide-mass fingerprinting and to N-terminal amino acid sequencing by Edman degradation (Alta Biosciences). The protein’s identity was confirmed as Hek with a high probability (Expectation: 2.14551 x 10^{-17}, Fig. 4.3 A and B) and the sequence analysis results confirmed the predicted cleavage point of the signal sequence and identified the lysine at position 23 of the preprotein as the start of the mature Hek protein (Fig. 4.3 C). Interestingly, the amino-terminal section of Hek displays the greatest degree of sequence divergence from the closely related Hra1 protein (only 14/27 residues identical) while the highly dissimilar Tia and Hra1 proteins are almost identical over this same region (25/27).

### 4.2.2.2 Prediction of the secondary structure of the Hek protein

Several online computational structural modelling tools were used to derive a predicted secondary structure for the 224 amino acid mature Hek protein, including the PredictProtein program (www.embl-heidelberg.de/predictprotein/predictprotein.html), a porin-recognition utility, PORES (http://garlic.mefos.hr/pores/) and SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/). The protein was predicted to contain 8 β-sheets separated by alternating short (3-6) or long (28-33) stretches of residues (Fig. 4.4 A). A series of alternating hydrophobic amino acids occupying every second position along the strand can be found in each of the predicted β-sheets of the Hek protein in a manner characteristic of β-barrel proteins (Fig. 4.4 A) (136). The only exceptions to this are a stretch of three polar residues including two charged amino acids (RTE) in the centre of the third strand and a second polar patch on the neighbouring fourth strand, although this second polar area has two adjacent alternating valine residues which may
Figure 4.3 MALDI-TOF analysis and amino-terminal sequencing of the mature Hek protein. Panel A: Peptide mass fingerprint spectrum of the mature Hek protein. Panel B: Hek protein sequence coverage. The peptides identified by MALDI-TOF analysis of the trypsinised mature Hek protein are aligned with the predicted in silico Hek sequence. Panel C: Alignment of the amino-terminal portions of the Hek, Hral and Tia proteins showing the signal peptide cleavage site of the Hek protein as determined by N-terminal sequencing of the mature protein. A black background indicates identical residues and conservative substitutions marked in grey.
Figure 4.4 Panel A: Topological model of the mature Hek protein showing its important features: the predicted β-sheets are indicated by residues in diamonds with the presumed unstructured loops in circles. The A and R residues in orange hexagons show the extent of each loop deletion. The pattern of alternating hydrophobic amino acids is highlighted in yellow and the aromatic residues which may ring the barrel at the membrane interface are in green. The RTE motif is shown in pink and surface-exposed basic and acidic residues are also highlighted. Panel B: Diagrammatic representation of the technique of inverse PCR used to generate the loop deletion plasmids pLoop1-4. The grey half-arrows represent the divergent primers used in the PCR reaction with the CGC overhangs in red.
compensate for the interruption in the pattern. This suggests that the Hek β-barrel may have a polar patch exposed on the outside of the barrel which could potentially be involved in homomeric or heteromeric intermolecular interactions within the membrane (136, 240). In support of the β-barrel prediction several of the β-sheets are also flanked by aromatic residues in a manner also characteristic of other β-barrel proteins (205, 229). Based on these observations the Hek protein is likely to adopt a β-barrel conformation in the outer membrane consisting of 8 anti-parallel β-sheets connected by 4 long external loops and 3 short periplasmic turns. The predicted external loops were also found to contain a very high number of charged residues compared to similar β-barrel proteins (135, 205, 264). A topological model of the Hek protein is shown in Figure 4.4 A with the salient points highlighted, including the aromatic rings, alternating hydrophobic residues in the trans-membrane β-sheets and the surface-exposed charged residues.

4.2.3 Structure-function analysis of the Hek protein

4.2.3.1 Construction of hek loop-deletion mutants

According to the predicted structure of the Hek protein the only portions of the protein exposed to the external milieu are the four external loops and thus it can be assumed that the biological function of the protein resides with one or more of these loops. In order to elucidate which of the loops play roles in the various Hek-promoted phenotypes a series of hek loop-deletion mutant genes were constructed based on the pHEK6 plasmid each lacking the coding sequence for a single loop. Inverse PCR, utilising divergent primers flanking the region to be deleted (Table 2.3), was used to generate a linear DNA fragment lacking the coding sequence for a single loop. Ligation of this linear fragment then produced an intact plasmid with a defined deletion (Fig. 4.4 B). The aim was to remove as much of the external structure of the loop without affecting the overall folding of the protein. In order to promote correct folding of the β-barrel each primer was designed with a CGC 5' extension thus introducing two new residues, arginine and alanine, in place of the deleted loop. Arginine was chosen as a strong turn-promoting amino acid (215) and the coding sequence for alanine and arginine used, GCG CGC, also introduces a novel BssHIII restriction site, allowing for
easy screening of putative mutants. Both of these codons are decoded with maximal efficiency in *E. coli* (97) and should have no adverse affect on the translation efficiency of the protein. In this manner, four mutant plasmids were constructed, each lacking the coding sequence for a single loop. The structures of the plasmids were confirmed by PCR and nucleotide sequence analysis and the confirmed mutant plasmids were designated according to their shortened loop: pLoop1, pLoop2, pLoop3 and pLoop4. The predictive topological model of the Hek protein in Figure 4.4 A shows the surface-exposed loops and the extent of the deletions in each mutant.

4.2.3.2 *Expression of Hek loop-deletion mutants*

To confirm the expression of all four loop-deletion mutants whole-cell lysates of *E. coli* K-12 strains harbouring each of the four mutant plasmids and the parental construct pHEK6 were separated on denaturing SDS-polyacrylamide gels, transferred to PVDF membranes and analysed by western immunoblotting using a Hek-specific polyclonal antiserum (Fig. 4.5 A). Proteins of the correct predicted sizes were detected for all four mutants, however, the loop 3 mutant appeared to produce considerably less protein than the wild-type (approximately 5- to 10-fold on average by densitometric analysis). A small amount of degraded protein was also detected with the loop 4 mutant. To determine if the mutant proteins were being correctly translocated and inserted into the outer membrane bacteria expressing the mutant or wild-type proteins grown overnight in liquid cultures were fixed with formaldehyde and the surface-exposed Hek or mutant-Hek protein was detected using a Hek-specific polyclonal antibody by flow-cytometry (Fig. 4.5 B). The loop 1 and 4 mutants displayed approximately 40 - 50 % of wild-type Hek protein levels indicating that these mutant proteins are less efficiently targeted to and assembled in the outer membrane. The loop 2 mutant produced approximately 80 % of the wild-type levels and the loop 3 deletion produced approximately 10 % in line with the results of the western-immunoblot analysis.

4.2.3.3 *Haemagglutination and autoaggregation phenotypes of Hek loop-deletion mutants*

*E. coli* K-12 strains expressing each of the four Hek loop-deletion variants or the wild-type protein were assessed by standard haemagglutination assays with human blood to
Figure 4.5 Expression of Hek loop-deletion mutants. Panel A: Western immunoblot analysis of whole-cell lysates of *E. coli* K-12 strain XL-1 harbouring pBSKII (Hek-), pHEK6 (Hek+), pLoop1 (Δ1), pLoop2 (Δ2), pLoop3 (Δ3) or pLoop4 (Δ4). Panel B: Analysis of the surface exposure of Hek loop-deletion mutants by flow-cytometry. Surface exposed protein was detected using a rabbit anti-Hek antibody and a secondary anti-rabbit FITC-conjugated antibody. The mean fluorescence of 10,000 unique events was detected by flow-cytometry and each value was expressed as a percentage of the wild-type levels.
Figure 4.6 Haemagglutination and autoaggregation by bacteria expressing Hek or loop-deletion derivatives. Haemagglutination of 1% human blood (Panel A) or autoaggregation (Panels B) with *E. coli* K-12 strain XL-1 harbouring pHEK6 (Hek+), pBSKII (Hek-), pLoop1 (Δ1), pLoop2 (Δ2), pLoop3 (Δ3) or pLoop4 (Δ4). Panel C: The rate of settling in all six cultures was quantified by plotting the decrease in surface optical density against time.
determine the effect of the deletion of each surface-exposed loop had on the agglutination phenotype (Fig. 4.6 A). Deletion of loop 1 had no discernable effect on haemagglutination, with a titre of 32 for both the loop 1 mutant and wild-type Hek proteins but deletion of loop 4 reduced the titre to 8. Deletion of either loop 2 or 3 appeared to abolish haemagglutination completely although the lowest titre well of the loop 3 mutant appeared slightly agglutinated. These data suggested that loop 2 and perhaps loop 3 play a role in haemagglutination, however, the reduced expression of the loop 3 deletion mutant could account for its reduced haemagglutination phenotype. These mutations also had a dramatic and visible affect on Hek-promoted autoaggregation in standing cultures (Fig. 4.6 B). No visible aggregation was observed with the strains expressing the Hek proteins lacking loops 2 and 3 and a highly reduced phenotype was evident with the loop 4 mutation. As with haemagglutination, the loop 1 mutant appeared to aggregate as efficiently as the wild-type. To quantify the effect of each mutation on aggregation, the rate of settling in each culture was monitored as before by measuring the change in the optical density at the surface of the culture over time (Fig. 4.6 C). Expression of the loop 1 deletion mutant resulted in an aggregation profile essentially identical to or even slightly more efficient than the wild-type protein. Mutations in loop 2 or 3 reduce aggregation close to the background levels displayed by the negative control with the loop 3 mutant aggregating only slightly more efficiently than the loop 2 mutant. In repeat experiments the loop 2 mutant was indistinguishable from the negative control and the loop 3 mutant displays a similar low level of aggregation. The mutation in loop 4 results in an unusual intermediate phenotype where aggregation at first occurs as rapidly as the wild-type but slows and levels out at a much higher optical density than the wild-type. This intermediate phenotype was also very similar to that observed in the haemagglutination assays with this mutant.

4.2.3.4 Interaction of Hek loop-deletion mutants with mammalian epithelial cells

As mutations in loops 2 and 3 of Hek have dramatic effects on the ability of the protein to promote haemagglutination and autoaggregation it was decided to test all four loop deletion mutants for the ability to promote invasion of epithelial cells that is characteristic of expression of Hek in a non-invasive strain. As with the earlier haemagglutination and autoaggregation assays the four mutant proteins or wild-type Hek were over-expressed in an E. coli K-12 background and subjected to gentamicin
survival assays with CHO-K1 epithelial cells. The CHO-K1 cell line was chosen over
the more clinically relevant T84 cell type because its much greater sensitivity to
infection by Hek-expressing bacteria would allow small differences in invasion
efficiency to be easily differentiated. The deletion of loop 2 was found to completely
abolish the Hek-promoted invasion phenotype (Fig. 4.7) but interestingly, the mutation
in loop 3 resulted in only an approximate 2-fold reduction in the level of invasion
compared with the wild-type. As with the haemagglutination and auto aggregation
phenotypes the loop 1 deletion had little effect on the invasion efficiency, resulting in
only a slight decrease (1.25-fold) and in contrast with the autoaggregation data the loop
4 mutation actually appeared to increase the level of invasion slightly (1.4-fold),
although this was found to be quite variable from experiment to experiment.
Figure 4.7 Interaction of Hek loop deletion mutants with CHO-K1 cells. Invasion of CHO-K1 cells by *E. coli* K-12 harbouring pHEK6 (Hek+), pBSKII (Hek-) or the loop deletion plasmids, pLoop1-4 (Δloop1-4).
4.3 Discussion

As the ETEC protein Tia has been shown to interact with the heparin glycosaminoglycan (GAG) moieties found on many proteoglycans (73) it was decided to test the invasion of Hek-expressing bacteria in pgsA-745 cells, a CHO-K1-derivative cell line deficient in the production of GAGs. *E. coli* K-12 expressing Hek or Tia were significantly less invasive in pgsA-745 cells compared with the parental CHO-K1 cells. However the background level of invasion of the *E. coli* K-12 strain harbouring the empty vector control was also reduced in pgsA-745 cells perhaps indicating that the *E. coli* strain used has inherent GAG-binding capabilities. Bacteria expressing Hek or Tia were still significantly more invasive than the negative control in pgsA-745 cells. This suggests that Hek and Tia may be capable of interacting with additional CHO-K1 cell surface components. It is possible that the initial interaction between the bacterial protein and a GAG moiety facilitates further interactions with an additional cell surface component, perhaps even with the core protein of the receptor proteoglycan, which would still be present on the surface of pgsA-745 cells. To examine the interaction of Hek with glycosaminoglycans in greater detail purified GAGs were used to inhibit the invasion of CHO-K1 cells by *E. coli* K-12 expressing Hek. At a high concentration (100 µg/ml) both heparin and heparan sulphate were found to significantly inhibit invasion compared with the other common epithelial cell GAG, chondroitin sulphate, and the negative control dextran sulphate. However, heparan sulphate also reduced the background level of invasion by the *E. coli* K-12 strain alone suggesting that the inhibition by heparin was the more significant and Hek-specific. The inhibition by heparin was found to be dose dependent and invasion was still efficiently inhibited with as little as 100 ng/ml of heparin. Indeed, when the earlier inhibitions were repeated with all four inhibitors at a lower concentration (10 µg/ml) only heparin resulted in significant levels of inhibition. These data suggest that Hek can interact specifically with the highly-sulphated heparin moieties of proteoglycans and perhaps with some other additional mammalian cell surface component. Heparin was also the only GAG to inhibit the interaction of *E. coli* K1 strain RS218 with CHO-K1 cells. This indicates that *E. coli* K1 does utilise some interactions with heparinated-proteoglycans in its association with CHO-K1 cells. It is intriguing to speculate that the Hek protein may
be responsible for these heparin-dependent interactions but this is impossible to confirm without further data.

The large differences in invasion efficiency of T84 and CHO-K1 cells by Hek-expressing bacteria suggest that a specific receptor rather than a non-specific electrostatic interaction is involved in Hek-promoted binding and that this receptor is differentially expressed by these cell types. This is unsurprising as proteoglycan expression is tissue and temporally specific (18) and although both T84 and CHO-K1 cells are epithelial in morphology they originate from different species (human and hamster) and different tissue types (colon and ovary). Furthermore T84 cells are likely to have an altered expression profile as this cell-line originated as a lung-metastasis of a colonic carcinoma (55, 176) and metastatic cells have been shown to have altered proteoglycan expression, in particular syndecan 1 and 4 are often down-regulated to allow cells to detach from their tissue site and syndecan 2 is up-regulated to promote migration (96). The adhesion profile observed on T84 cells (Chapter 3, section 3.2.5), with bacteria clustering on particular cells, can also be explained by the apparent variations of proteoglycan expression within a cell monolayer that have been observed with several epithelial cell lines including Caco-2 cells (101).

The Hek protein was predicted to contain an amino-terminal signal-sequence of the type recognised by the Sec membrane translocation machinery and was detected in the sarcosyl-insoluble fraction of a bacterial lysate suggesting that it is associated with the outer membrane. Mature Hek protein isolated from a sarcosyl-insoluble fraction was found to lack this 22 amino acid signal-sequence, as predicted, and the amino-terminal sequence of the mature protein was confirmed. Computational modelling methods were employed to generate a prediction of the secondary structural characteristics of the Hek protein and from this to elucidate a model of the tertiary structure of the protein. The protein is predicted to share many secondary structural characteristics with outer membrane β-barrel proteins including the presence of 8 anti-parallel amphipathic β-sheets. This similarity is further supported by the presence of a putative ring of aromatic amino acids flanking several of the β-sheets in the region of the protein which was predicted to lie at the interface between the hydrophobic core of the membrane and the polar head-groups, a characteristic of many previously described β-barrel proteins
Based on these structural similarities the protein is predicted to form an 8-stranded β-barrel in the outer membrane with 4 long loops extending beyond the outer surface of the cell membrane and 3 short turns joining the β-sheets on the periplasmic side. One surprising aspect of the structural prediction was the presence on two charged residues in a polar patch (RTE) on the predicted outer face of the membrane-spanning barrel. It is possible that this area of the protein is involved in hydrophilic intermolecular interactions within the membrane, such as the formation of homomeric or heteromeric protein complexes. Further work would be necessary in order to test this hypothesis. Cross-linking experiments on intact bacteria expressing the Hek protein could be used in conjunction with western immunoblot analysis to determine if such interactions were indeed occurring within the membrane. This could then be confirmed by immuno-precipitation coupled with peptide-mass fingerprinting to identify co-precipitated proteins. Site-directed mutagenesis could then be employed to identify amino acids (for example, the RTE motif) involved in such interactions.

Although no additional secondary structure was predicted in the surface-exposed loops, the β-sheets on either side of the third loop were predicted to extend a considerable distance beyond the membrane surface. Such an arrangement may serve to extend and support this loop farther from the membrane and provide a support structure for the other loops in a manner similar to that observed with OpcA in Neisseria meningitidis and OmpX in E. coli (229, 264) (205). One striking feature of the surface-exposed loops of Hek is the extreme abundance of charged amino acids present, 20 basic and 26 acidic residues. The presence of surface-exposed basic residues has been shown to be involved in binding to proteoglycans (148) and in the case of OpcA in particular, the combination of extended β-sheet structure and an abundance of basic residues results in the formation of a positively charged cleft which is theorised to be the site of binding to heparinated proteoglycans (205). Although the data gathered here does not allow any accurate prediction of the final structure of the surface-exposed loops of Hek they may play a similar role in proteoglycan binding.

In an effort to analyse the contribution of the surface-exposed loops of Hek to the various Hek-promoted phenotypes a series of mutant plasmids were constructed each encoding a mutant Hek protein lacking a single loop. All four mutant proteins could be
expressed in *E. coli* K-12 and were detected on the surface of bacterial cells by flow-cytometry although the mutant lacking loop 3 was found to be poorly expressed compared with the wild-type protein. There are several possible explanations for this poor expression: the deletion may somehow affect the translation of the protein or result in a reduction in the stability of the protein leading to subsequent degradation in the cell. Several methods were originally used to predict the structure of Hek and all agreed on the gross structural attributes of the protein as described earlier (section 4.2.2.2). Of particular interest the fifth β-sheet, to the N-terminal side of loop 3, was predicted to extend considerably beyond the lipid bilayer into the region deleted in the loop 3 mutant. Deletion of such a highly structured region could explain a significant reduction in the stability of the protein as a whole. Alternatively if loop 3 is an antigenically dominant epitope for the Hek-specific polyclonal antiserum the apparent poor expression may represent a failure to detect the protein rather than an actual decrease in expression. Further experimentation is necessary to clarify this situation. However, as this mutation does not have a dramatic effect on the Hek-promoted invasion phenotype it was decided not to pursue any further analysis. Deletion of either loop 2 or 3 was found to completely abolish Hek-promoted haemagglutination and autoaggregation although the defect with the loop 3 mutant may be attributable to its potentially poor expression. The loop 1 deletion had no apparent effect on haemagglutination or autoaggregation and interestingly, the loop 4 deletion resulted in an intermediate phenotype in both assays. The loop 1 deletion was also found to have no significant effect on the level of invasion by Hek-expressing *E. coli* K-12, the slight decrease in observed invasion efficiency correlates with the slight decrease in expression of the mutant compared with wild-type. Taken together these data imply that loop 1 plays no role in any of the observed Hek-promoted phenotypes. In contrast with the data from both haemagglutination and autoaggregation assays the mutation in loop 4 had no detrimental impact on invasion and indeed in some experiments actually appeared to enhance invasion, despite a reduction in expression similar to that observed with the loop 1 mutant. This suggests that the mechanisms by which Hek promotes agglutination of erythrocytes and bacterial aggregation are related but may differ from that which is required for successful invasion of epithelial cells. Further analysis is required to clarify this possibility. The loop 3 deletion, which may be poorly expressed, did result in lower invasion efficiency (approximately 2-fold) but the strain expressing
this mutant was still highly invasive despite the apparently poor expression. Only the loop 2 deletion was found to completely abolish Hek-promoted invasion. Detection of the loop 2 mutant protein by flow-cytometry on intact cells suggests that the protein is correctly folded and inserted into the outer membrane despite the extensive deletion. These data clearly indicate that loop 2 plays an essential role in all three phenotypes conferred by Hek.

It has also been reported in the literature that the second loop of the related Tia protein is involved in invasion (154) and binding of heparin (73). The data presented to support these conclusions is based solely on the ability of two short peptides (GYDFYQHYNVPVRTEVEFY or AVGYDFYQHYNVPVRTEVEC) to inhibit invasion HCT8 human epithelial cells by Tia-expressing bacteria or bind directly to heparin-albumin-biotin. These peptides are reported as corresponding to residues 76 – 94 or 78 – 96 of the Tia protein, supposedly corresponding to loop 2 of Tia. However, the Hek protein structural prediction presented here (see Chapter 4, Fig. 4.4) combined with the homology between the Hek and Tia proteins (see Chapter 3, Fig. 3.1) would predict that this region of Tia is in fact mostly embedded within the membrane, including part of the second β-sheet, the periplasmic turn and most of the third β-sheet. Indeed the reported peptide sequences contain members of the predicted aromatic rings (205, 229) and the characteristic patterns of alternating hydrophobic amino acids, indicative of membrane spanning β-sheets (underlined) (264). Furthermore, the primary sequence of Tia flanking this region contains a very large number of charged residues, which are only rarely found within membrane-spanning β-sheets (136). Based on these findings it is clear that these peptides could not correspond to a surface exposed portion of the protein. The numbering system reported corresponds to the Tia pre-protein, including an intact signal sequence. However, residues 76 – 96 of the mature Tia would form part of the second surface-exposed loop. It therefore appears likely that the use of these peptides may have resulted from confusion between numbering systems of pre- and mature forms of the Tia protein.

The surface loops of β-barrel proteins were once thought to be highly flexible and unstructured appendages but recent X-ray crystallography evidence of the OpcA protein for Neisseria meningitidis (205) suggests that complex electrostatic and hydrogen-
bonding interactions between loops result in highly organised surface structures. OpcA is also a heparan sulphate binding protein and its crystal structure suggests that the interactions between the surface exposed loops form a heparin binding crevice lined with basic residues from two loops. A similar situation may exist in the case of the Hek loops with interactions between the loops providing a binding site lined with positively charged residues that are provided by loop 2, with perhaps contributions from other loops. The apparent increase in invasion observed with the loop 4 deletion mutant may be due to a change in the conformation of the remaining loops which enhances binding to the CHO-K1 cell surface receptor, perhaps by releasing a steric hindrance to binding, but which interferes with the autoaggregation and haemagglutination phenotypes. The conflicting results with CHO-K1 cells and human erythrocytes suggest that the Hek protein may interact with different receptors on these cell types or perhaps the same receptor is found on both but the extreme heterogeneity displayed by the glycosaminoglycans may explain the differences.
Chapter 5  Regulation of *hek* expression in *Escherichia coli* K1
5.1 Introduction

5.1.1 Phase variation

Survival of bacteria in a host requires the ability to adapt to changing environmental conditions and niches and to avoid clearance by the host's immune response. This adaptation necessarily involves alterations in bacterial gene expression. One mechanism by which bacteria alter gene expression in response to their environment is phase variation. Phase variation is generally considered to be the ability to reversibly switch the expression of a gene from on to off but variation between discreet levels of expression has also been described (156). Bacterial surface components, such as adhesins and invasins, are often strongly immunogenic and as a result are targeted by the immune response. As a result, their expression is a compromise between functional necessity and the danger of immune surveillance. Phase variation provides a mechanism by which the expression of such proteins can be optimally varied throughout the infectious process. A number of mechanisms by which phase variation can occur have been identified: small alterations in DNA sequence by slipped-strand mispairing (SSM), inversion of DNA elements by site-specific recombination and epigenetic mechanisms involving DNA methylation. SSM involves the expansion or contraction of DNA nucleotide repeats during replication of the DNA strand. Misalignment between the parental and daughter strand results in an increase or decrease in the number of repeats (146). SSM has been identified in many bacterial species as a source of phase variation (49, 87, 95, 141). SSM events can affect gene expression in numerous ways including shifts in the reading frame, when it occurs within an ORF (144), altered promoter spacing (259), and even altered regulator binding sites (156). SSM has been described with repeat units of between 1 and 7 nt and such repeat sequences are found in a wide range of prokaryotic genomes (258). No confirmed example of SSM-mediated phase variation in *E. coli* has been described to date and the genome of *E. coli* K-12 was found to have few examples of repeat sequences (258). However, SSM has been demonstrated in *E. coli* using exogenous DNA sequences (255) and with the large amounts of novel DNA present in pathogenic *E. coli* lineages SSM may be relevant in *E. coli* pathogenesis. Phase variation by site-specific recombination is typified by the invertible element controlling the expression
of type 1 fimbriae in *E. coli* (2). The promoter for the *fim* operon, encoding the structural genes of the type 1 fimbriae, is located on a 314 bp DNA element, the *fim* switch, flanked by short inverted repeats. Recombination between the repeats results in inversion of the *fim* switch, switching the orientation of the promoter either towards or away from the *fim* structural operon. Recombination at the switch is mediated by two site-specific recombinases, FimB and FimE, whose differing affinities for the switch depending on its orientation result in complex regulation of inversion (238). Similar recombination systems have also been described in other bacterial species, including *Proteus mirabilis* (282) and *Salmonella* spp. (283). Epigenetic phase variation involves the differential methylation of DNA within regulatory regions that affects the binding of an essential regulator but does not involve DNA sequence changes. Variation by this mechanism was first described for the *pap* operon in *E. coli* that encodes the P type fimbriae (23). Dam methylation of GATC sites in the *pap* regulatory region alters the binding of the global regulator Lrp to two sites within the regulatory region. Lrp acts as a repressor when bound to one site proximal to the *pap* operon and as an activator when bound to the second distal site. Differential methylation of GATC sites within each of the Lrp binding sites results in either phase On or Off fimbrial expression (103). The expression of the *E. coli* Ag43 protein is also varied by DNA methylation (100). Dam methylation blocks binding of OxyR to the regulatory region of the *agn43* gene. When OxyR binds it acts as a repressor of *agn43* transcription, thus the competition between OxyR and Dam for the *agn43* sites results in phase-variable expression of Ag43. In both cases the methylation state changes when DNA replication occurs. The resulting daughter strands are only hemi-methylated allowing fresh competition between the regulatory protein and Dam methylase (103).

5.1.2 Integration host factor

Integration host factor (IHF) is a nucleoid-associated protein and global regulator in *E. coli* (6). IHF plays an important role in the compaction of the bacterial chromosome (120) by binding to specific DNA sequences and inducing bends of approximately 160° in the DNA molecule (251). IHF is functional as a heterodimer of α and β subunits encoded by the *ihfA* and *ihfB* genes. IHF binds with high affinity to consensus sites with the sequence WATCARXXXTTR (where W is A or T, R is A or G and X is A,
T, G or C) and has been implicated in the regulation of up to 120 genes in *E. coli* K-12, with examples of both positive and negative regulation described (6). In many cases IHF-induced DNA bending mediates gene regulation by facilitating interactions between RNA polymerase and regulators bound upstream, known as action at a distance (192). IHF can also act to repress promoter activity by steric hindrance, either by directly binding to a consensus site within the promoter region or inducing conformational changes in the local DNA that hinder the activity of the promoter (27). IHF-induced DNA constrictions may also act to inhibit the dispersion of superhelical energy along the DNA molecule. Depending on the location of IHF binding sites this superhelical energy can be focussed onto a specific promoter resulting in enhanced open complex formation and increased transcription as in the *ilvGMEDA* operon (191).
5.2 Results

5.2.1 Plasmid nomenclature

Due to the large number of structurally similar plasmids used in this chapter a simplified plasmid nomenclature has been devised to aid the reader (Table 5.1). Plasmid names will begin with “p” as per standard nomenclature. The next letter of the plasmid name signifies copy number; H for high copy number, M for moderate copy number and L for low copy number. The third character of the scheme indicates the promoter structure preceding the gene, + indicates that an extended (1138 bp) region is included, - indicates a smaller (76 bp) region and I indicates that the gene is preceded by an inducible promoter. The fourth character, if present, indicates if the Hek coding sequence is fused to \textit{phoA} (P) or \textit{lacZ} (Z). Finally, the insertion of promoter fragments into the first or second pQF50 \textit{Smal} site is indicated with SI or S2 (proximal to \textit{lacZ}). The standard names of plasmids used in previous chapters are also listed alongside the new nomenclature.

5.2.2 Analysis of Hek protein expression in \textit{E. coli} K-12 and K1

Expression of the Hek protein in \textit{E. coli} K-12 strain XL-1 from the high-copy number plasmid pH- (Table 5.1, formerly pHEK6) was found to confer a strong heat-resistant haemagglutination phenotype and to promote autoaggregation of static bacterial cultures. Furthermore, the presence of this plasmid was found to enhance adhesion to and invasion of cultured epithelial cells by \textit{E. coli} K-12 (Chapter 3). These various Hek-promoted phenotypes were associated with a high level of Hek protein expression as detected by SDS-PAGE and western immunoblot analysis. In order to examine these phenomena in the context of the pathogenesis of \textit{E. coli} K1 sepsis and meningitis the expression of the Hek protein in \textit{E. coli} K1 strain RS218 was examined.

5.2.2.1 Expression of Hek from its native promoter in high-copy number

The expression of the Hek protein in strains of \textit{E. coli} K-12 harbouring the pH- plasmid was easily detected by western immunoblot analysis using a Hek-specific polyclonal
antiserum (see Chapter 3, section 3.2.2.1 and Fig 5.1 A) and appeared consistent in all samples analysed. Initial attempts to detect chromosomally expressed Hek protein in E. coli K1 strain RS218 by standard western immunoblotting procedures failed, suggesting that the protein was expressed poorly or not at all in this strain. In order to confirm that the protein could be expressed in E. coli K1, the plasmid pH- was transformed into the E. coli K1 strain and the expression of Hek was analysed. In contrast to the expression of Hek in E. coli K-12 the expression in E. coli K1 appeared to be highly variable. In repeated experiments high-level expression of the Hek protein was detected in only a single culture of E. coli K1 strain RS218 harbouring pH- (Fig. 5.1 A). No obvious reason for this variation was observed but all stages in protein synthesis, from transcription and translation to export and folding, may be subject to regulation.

5.2.2.2 Expression of Hek from the inducible \( P_{\text{reg}} \) promoter

To examine the expression of the Hek protein independent of its native promoter the expression of the protein from a strong inducible \( P_{\text{reg}} \) promoter was analysed in both E. coli K-12 and K1. When E. coli K-12 harbouring the pMI plasmid (Table 5.1, formerly pThek6) was induced with IPTG high-level production of Hek protein was observed (Chapter 3, section 3.2.2.3). This plasmid was transformed into E. coli K1 strain RS218 and 7 independent cultures were grown, induced and analysed by western immunoblot (Fig. 5.1 B). None of the cultures produced as much Hek protein as the E. coli K-12 control strain and among the 7 cultures the level of expression was variable. Densitometric analysis identified two distinct expression levels, the first was between 54 and 74% of that observed with E. coli K-12 and the other was approximately 25%. This suggests that there may be a degree of post-transcriptional variation in the expression of the Hek protein but this effect is not sufficient to explain the observed variation with the pH- plasmid in E. coli K1 described above.

5.2.2.3 Expression of Hek from its native promoter in low-copy number

As the plasmid pH- is present in between 200 and 400 copies per cell the low-copy number pL- plasmid (Table 5.1, formerly pPDhek22) was assumed to provide a closer approximation of the expression from the chromosomal allele. When this plasmid was transformed into the E. coli K-12 strain XL-1 the amount of detected Hek protein was
Figure 5.1 Comparison of Hek protein expression in *E. coli* K-12 and K1. Panel A: Western immunoblot analysis of *E. coli* K-12 strain XL-1 or K1 strain RS218 harbouring plasmid pH- (Hek+) or pBSKII (Hek-). Panel B: Western immunoblot analysis of IPTG-induced (+) or uninduced (-) *E. coli* K-12 or K1 carrying the pMI plasmid.
<table>
<thead>
<tr>
<th>Plasmid name</th>
<th>Original name</th>
<th>Copy number</th>
<th>Length of upstream</th>
<th>Fusion gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH-</td>
<td>pHEK6</td>
<td>High</td>
<td>76 bp</td>
<td>None</td>
</tr>
<tr>
<td>pH+</td>
<td>High</td>
<td>1138 bp</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>pH-P</td>
<td>High</td>
<td>76 bp</td>
<td>phoA</td>
<td></td>
</tr>
<tr>
<td>pH+P</td>
<td>High</td>
<td>1138 bp</td>
<td>phoA</td>
<td></td>
</tr>
<tr>
<td>pL-</td>
<td>pPDhek22</td>
<td>Low</td>
<td>76 bp</td>
<td>None</td>
</tr>
<tr>
<td>pL+</td>
<td>Low</td>
<td>1138 bp</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>pL-P</td>
<td>Low</td>
<td>76 bp</td>
<td>phoA</td>
<td></td>
</tr>
<tr>
<td>pL+P</td>
<td>Low</td>
<td>1138 bp</td>
<td>phoA</td>
<td></td>
</tr>
<tr>
<td>pM-Z</td>
<td>Moderate</td>
<td>76 bp</td>
<td>lacZ</td>
<td></td>
</tr>
<tr>
<td>pM-ZS1</td>
<td>Moderate</td>
<td>76 bp</td>
<td>lacZ</td>
<td></td>
</tr>
<tr>
<td>pM-ZS2</td>
<td>Moderate</td>
<td>76 bp</td>
<td>lacZ</td>
<td></td>
</tr>
<tr>
<td>pM+ZS1</td>
<td>Moderate</td>
<td>1138 bp</td>
<td>lacZ</td>
<td></td>
</tr>
<tr>
<td>pM+ZS2</td>
<td>Moderate</td>
<td>1138 bp</td>
<td>lacZ</td>
<td></td>
</tr>
<tr>
<td>pMI</td>
<td>pThek6</td>
<td>Moderate</td>
<td>P_{phoA} promoter</td>
<td>None</td>
</tr>
<tr>
<td>pMIP</td>
<td>Moderate</td>
<td>P_{phoA} promoter</td>
<td>phoA</td>
<td></td>
</tr>
</tbody>
</table>

1 - High copy number plasmids contain inserts in plasmid pBSKII

- Moderate copy number plasmids contain inserts in pQF50 (lacZ) or pTrc99a

- Low copy number plasmids contain inserts in pPD101
approximately the same as that observed with bacteria harbouring pH- (Fig. 5.2 A). This finding suggested that the production of Hek protein in this strain is saturable, perhaps due the titration of a necessary activator of hek transcription at high plasmid copy number or poor toleration of the over-expressed protein. When this plasmid was transformed into *E. coli* K1 strain RS218 the Hek protein was well expressed (Fig. 5.2 B) and furthermore, the expression did not appear to display the variability observed with plasmid pH- in this same strain. This finding suggests that the extreme variability of expression from plasmid pH- in *E. coli* K1 may have been an artefact of the artificially high-copy number. However, the level of Hek protein produced by bacteria harbouring this 2 - 4 copies/cell plasmid suggested that despite the earlier failure the Hek protein expressed from the chromosomally located *hek* gene in *E. coli* K1 strain RS218 should be easily detectable by western immunoblotting. A more sensitive western immunoblotting procedure was employed and adapted for use with the anti-Hek antiserum (see Chapter 2, section 2.3.3). This enhanced sensitivity allowed the detection of Hek protein in *E. coli* K1 strain RS218, however, the level of expression observed was considerably greater than 4-fold less than that observed in strains harbouring the pL- plasmid (Fig. 5.2 C). Densitometric analysis revealed at least an 8-fold difference in the Hek protein expression levels between the wild-type *E. coli* K1 strain RS218 and the isogenic *hek* mutant RShek harbouring the pL- plasmid and this is likely to be an underestimation due to the limitations in accurately assessing the density of over-exposed protein bands. This observation implied that the expression of Hek from the chromosomal locus was under some form of control that was absent from the plasmid-encoded gene. This could conceivably involve the titration of an extremely low abundance regulator, resulting in the de-repression of the additional copies of the *hek* gene, or perhaps indicate the absence of some regulatory sequence upstream of the *hek* gene in pL-. As the *hek*-containing fragment of RS218 DNA cloned in the pL- plasmid includes relatively little DNA from upstream of the gene (see Chapter 3, section 3.2.2.1) it is a distinct possibility that some control sequence is missing from the plasmid.

### 5.2.2.4 Cloning of the *hek* regulatory region

In order to determine the necessity or importance of regulatory sequences upstream of the *hek* gene on the *E. coli* K1 strain RS218 chromosome, a large fragment of upstream
DNA was cloned into the pBSKII plasmid and subsequently sub-cloned upstream of the *hek* gene in plasmid pH-. A marker-rescue approach using *E. coli* strain RSHek (Table 2.1) was adopted to clone a large fragment of the bacterial chromosome containing the mutated *hek::spc* allele. This avoided a potentially error-causing PCR step and negated the need to subsequently sequence the entire cloned DNA fragment. The unfinished RS218 genomic DNA sequence was examined to identify restriction endonuclease sites flanking the *hek* gene and including at least 1000 bp of upstream DNA. An *XhoI* site was identified 1138 bp upstream of the *hek* ORF and an *EcoRV* site 140 bp before the end of the *hek* gene. Furthermore, this *EcoRV* site lies downstream (367 bp) of the *BsaBI* site used in the construction of the mutated *hek* allele in *E. coli* K1 strain RSHek (Table 2.1, see Chapter 3, section 3.2.6.4). No *XhoI* and *EcoRV* restriction sites are found within the specinomycin-resistance cassette. Total genomic DNA was isolated from strain RSHek and digested with the *XhoI* and *EcoRV* restriction endonucleases. The digested genomic DNA was ligated with pBSKII cut with *XhoI* and *EcoRV* and transformed into *E. coli* K-12. Spectinomycin resistant bacteria were screened by PCR using *hek*-specific primers (*hekfm* and *hekrORF*, Table 2.3). Plasmid DNA was isolated from positive clones and further screened by restriction endonuclease mapping. Plasmid pBSK::hspcl (Table 2.2, Fig. 5.3 A) was found to contain a 4 kb fragment of RS218 genomic DNA including the *hek::spc* allele and 1138 bp of upstream DNA. This extended upstream region was then subcloned into plasmid pH- to generate a plasmid with the extended upstream region and an intact *hek* gene. The pH- plasmid was cut with *XhoI* and *BstAPI* to remove the 5' portion of the *hek* gene with the 76 bp of upstream DNA and the linearised plasmid backbone was retained (Fig. 5.3 B). The corresponding portion of the *hek* gene with the extended upstream region was then excised from pBSK::hspcl using the same enzymes and ligated to the pH- backbone resulting in a plasmid pH+ (Table 5.1, Fig. 5.3 C). As plasmid copy-number was earlier found to have an effect on Hek protein expression in the *E. coli* K1 strain RS218, the entire insert from pH+ was sub-cloned into the low-copy plasmid pPD101. The fragment containing the entire *hek* gene and the extended upstream region was excised from pH+ by digestion with *KpnI* and *SacI* and cloned into pPD101 cut with the same enzymes resulting in plasmid pL+ (Table 5.1).
Figure 5.2 Western immunoblot analysis of Hek protein in *E. coli* K-12 and K1. Panel A: Comparison of Hek expression from low-copy, pL- (Hek+) and pPD101 (Hek-), or high-copy, pH- (Hek+) and pBSKII (Hek-), plasmids in *E. coli* K-12 strain XL-1. Panels B and C: Expression of Hek in *E. coli* K1 strain RS218, the isogenic hek mutant RShek and RShek harbouring plasmid pL-. X-ray film was exposed to the blot for 10 sec (panel B) or 2 min (panel C). Bands corresponding to the Hek protein are indicated with ➤.
Figure 5.3 Construction of the pH+ plasmid. Total genomic DNA was isolated from *E. coli* K1 strain RShek (carrying the hek::spc mutation), cut with *XhoI* and *EcoRV* and ligated to pBSKII cut with the same enzymes yielding pBSK::hspc1 (Panel A). Both pBSK::hspc1 and pH- were cut with *XhoI* and *BstAPI* and the extended *hek* upstream region from pBSK::hspc1 was substituted for the corresponding shorter region of pH- (Panel B) yielding plasmid pH+ (Panel C).
5.2.2.5 Analysis of Hek expression from plasmids containing the extended upstream region

The expression of Hek protein in *E. coli* K-12 and K1 bacteria harbouring the high- and low-copy number plasmids with short (76 bp) or long (1,138 bp) upstream regions were compared by western immunoblotting (Fig. 5.4). *E. coli* K-12 strain harbouring either the pH- or pH+ plasmid displayed very similar identical levels of Hek protein expression and this was in line with the observation that the expression from pH- was apparently saturated. As pH- was present in excess of 50 times more copies per cell than pL- but results in similar levels of Hek protein expression, the upstream region in pH+ would need to reduce expression by greater than 50-fold before any detectable difference would be observed. However, when strains harbouring the pL- or pL+ plasmids were compared, an 8.4-fold difference was observed in the levels of Hek protein.

Interestingly, the *E. coli* K1 strains harbouring pH+, pL- or pL+ all produced detectable Hek protein (Fig. 5.4 B), with no apparent variation, providing additional evidence for the theory that the observed variation was specific to the pH- plasmid. The lack of variation in *E. coli* K1 harbouring plasmid pH+ ruled out plasmid copy number as the sole reason for variation in the expression of Hek. These findings suggested an alternative explanation, high plasmid copy number combined with the high-level expression from its deregulated promoter could be responsible for the variable expression observed with pH-. The levels of Hek protein expression in *E. coli* K1 harbouring pH+ and pL- were very similar and comparable to those observed with *E. coli* K-12. The *E. coli* K1 strain harbouring pL+ also displayed reduced Hek protein expression (3.4-fold), although the decrease was less than that observed in *E. coli* K-12. This discrepancy may be explained by the presence of exogenously expressed Hek protein in the *E. coli* K1 bacteria. Overall, the expression of Hek in bacteria harbouring the pL+ plasmid correlates well with the observed level of expression from the chromosomal locus in *E. coli* K1 strain RS218. This suggests that the additional upstream region does indeed play a role in the regulation of Hek expression.
5.2.2.6 Correlation between the level of Hek expression and Hek biological activity

To examine the effect of different levels of Hek protein expression on the efficiency of invasion, E. coli K-12 strains harbouring the pH-, pH+, pL- or pL+ plasmids were used to perform invasion assays with CHO-K1 cells. As expected, the strains harbouring pH- and pH+ displayed very similar invasion efficiencies (Fig. 5.5 A) whereas bacteria harbouring pL- displayed levels of invasion 3.3-fold greater than those harbouring pL- (Fig. 5.5 B). These data indicate a clear correlation between the invasion efficiency and the amount of Hek protein produced by a given bacterial population. This association was also apparent in haemagglutination assays using the same strains. E. coli K-12 bacteria harbouring pH-, pH+ or pL- all gave an agglutination titre of 16 but bacteria harbouring pL+ were significantly less able to promote agglutination with a titre of only 1 (Fig. 5.5 C). This finding also highlights the relative insensitivity of the haemagglutination assay.

5.2.3 Analysis of Hek translation and export using a PhoA reporter fusion

A series of fusions between the hek gene and phoA, encoding alkaline phosphatase, were constructed. The fusion plasmids were constructed such that the DNA sequence encoding the mature PhoA protein (without its own signal sequence) replaced the last codon of the hek gene. As the alkaline phosphatase enzyme is only active following export to the periplasm these fusions provide an indication of the combined effects of transcription, translation and export of the fusion protein to the periplasm via the Sec translocation pathway.

5.2.3.1 Construction of high- and low-copy hek-phoA fusion plasmids

Site-directed mutagenesis was employed to modify the hek gene in plasmid pH- by introducing a novel NheI site (GCTAGC) immediately before the last codon of the gene, yielding plasmid pHnhe6 (Fig. 5.6 A, Table 2.2). The modified hek gene still supported a strong haemagglutination phenotype in E. coli K-12 (not shown). The phoA gene (lacking all but 21 bp of the signal sequence coding sequence) was excised from plasmid pCGV1 with NheI and BamHI and ligated to pHnhe6 cut with the same
**Figure 5.4** Western immunoblot analysis of Hek expression in *E. coli* K-12 strainXL-1 and K1 strain RS218. Expression of Hek from: high copy plasmids pH- (76 bp upstream), pH+ (1138 bp upstream) or the vector control pBSKII (-ve); and low copy plasmids pL- (76 bp upstream), pL+ (1138 bp upstream) or the vector control pPD101 (-ve). Densitometric analysis of band intensity relative to the first lane is shown below each gel. Molecular weight standards are indicated with MW.
Figure 5.5 Phenotypic analysis of *E. coli* K-12 strains expressing different amounts of Hek protein. Panel A: Invasion of CHO-K1 cells by *E. coli* K-12 strain XL-1 harbouring pH-, pH+ or the vector control pBSKII. Panel B: Invasion of CHO-K1 cells by *E. coli* K-12 strain XL-1 harbouring pL-, pL+ or the vector control pPD101. Panel C: Haemagglutination with the same six *E. coli* K-12 strains as panels A and B using 1% human blood.
Figure 5.6 Construction of *hek-phoA* fusion reporter plasmids. Site-directed mutagenesis was performed to modify the *hek* gene on plasmid pH-, introducing a novel *NheI* site before the last codon and generating plasmid pHnhe6 (A). The *phoA* gene from plasmid pCGV1 was than excised using *NheI* and *BamHI* and ligated to pHnhe6 cut with the same enzymes (B). The resulting plasmid was designated pH-P (C). A low copy number version of this fusion was generated by subcloning the *hek-phoA KpnI/SacI* fragment from pH-P into pPD101 (not shown).
enzymes (Fig. 5.6 B). Bacteria were screened for a PhoA+ phenotype on agar containing X-P, a chromogenic substrate for alkaline phosphatase, and a single positive clone was confirmed by restriction endonuclease mapping and designated pH-P (Fig. 5.6 C, Table 5.1). The *hek-phoA* fusion gene was subsequently sub-cloned into the low-copy number pPD101 plasmid. The entire insert from plasmid pH-P (including the *hek* promoter) was excised with *KpnI* and *SacI* and ligated to pPD101 cut with the same enzymes yielding pL-P (Table 5.1).

### 5.2.3.2 Expression of the Hek-PhoA fusion protein

The presence of the pH-P plasmid in either *E. coli* K-12 strain XL-1 or K1 strain RS218 resulted in a highly variable PhoA phenotype on agar containing X-P (Fig. 5.7 A). A number of different colony morphologies were apparent. Two populations of blue colonies were distinguishable, pale blue and segmented dark blue colonies, and occasionally a third group of almost white colonies was also observed. Also, dark blue colonies were occasionally observed with white segments. To assess if the high plasmid copy number contributed to this variable phenotype (as suggested by earlier data using the pH- plasmid, section 5.2.2) the expression of the Hek-PhoA fusion in bacteria harbouring the low copy construct pL-P was examined. *E. coli* K1 strain RS218 harbouring pL-P gave a mostly uniform colony colour with occasional (approximately 1 in 500) darker blue colonies (Fig. 5.7 B). However, when one of these darker colonies was sub-cultured a highly mixed population resulted (Fig. 5.7 C) consisting of approximately 50:50 dark:pale blue colonies. These two colony phenotypes probably correspond to the two distinct populations of pale and dark blue observed with pH-P discussed above. The third population of almost white colonies observed with pH-P were unique to that plasmid. Taken together these data suggest that high plasmid copy number combined with de-repressed expression does indeed contribute to variation of Hek expression but that some underlying variability in the expression of the protein does occur. The occurrence of distinct colony types further suggests that this phenomenon is heritable. Earlier findings suggested that Hek variation in *E. coli* K1 was independent of the *hek* promoter, as expression from an inducible promoter also appeared variable between two discreet levels (section 5.2.2.2). To examine this possibility in further detail, a *hek-phoA* fusion that was transcribed from the inducible \( P_{\text{on}} \) promoter was constructed.
5.2.3.3 Construction of an inducible hek-phoA fusion plasmid and expression in E. coli K1

In order to construct an inducible hek-phoA plasmid the same approach was employed as described previously (section 5.2.3.1). Site-directed mutagenesis was performed to modify the hek gene in plasmid pMI and introduce a novel NheI site (GCTAGC) immediately before the last codon of the gene, yielding plasmid pTnhe4 (Table 2.2). The phoA sequence encoding the mature protein was excised from plasmid pCGV1 with Nhel and BamHI and ligated to pTnhe4 cut with the same enzymes. Bacteria were screened for a PhoA+ phenotype as before and a single positive clone was confirmed by restriction endonuclease mapping and designated pMIP (Table 5.1). Uninduced cultures of E. coli K1 strain RS218 harbouring pMIP displayed a visible PhoA+ phenotype, consistent with leaky expression from the P_Rec promoter (Fig. 5.8 A). The Hek-PhoA expression from this strain was also variable but the observed variability was much less extreme than that observed with bacteria harbouring the high-copy pH-P plasmid with only a slight difference in colour between the light and darker colony types. This observation perhaps reflects the relatively low copy number of the pMIP plasmid (20-25 copies/cell). However, when a single culture of bacteria harbouring pMIP was replica plated onto agar plates containing increasing levels of IPTG the variability became steadily more extreme with increasing IPTG concentration (Fig. 5.8 A - E). At moderate IPTG levels (0.01 mM) three distinct colony colours were apparent, dark blue, pale blue and white, and as the IPTG concentration was increased the proportion of white colonies also increased. This suggested that the pale and dark blue colonies were a result of an inherent variability in Hek expression as they were present at all IPTG concentrations. The white colonies, which appeared only at higher IPTG concentrations, seem to be an artefact of high-level expression of the fusion protein and may correspond to the similar white colonies described with plasmid pH-P above. This further supports the theory that high-level expression of Hek in pH- is responsible for the highly variable expression observed in E. coli K1 harbouring this plasmid. It is possible that high-level expression of the fusion, or even of Hek itself, actually results in a repression of its own expression.
Figure 5.7 Variable PhoA phenotypes from high and low copy *hek-phoA* fusion plasmids. Panel A: *E. coli* K1 strain RS218 harbouring the high copy number pH-P plasmid grown on L-agar containing the chromogenic PhoA substrate X-P. Examples of each different colony morphology are highlighted. Panel B: *E. coli* K1 strain RS218 harbouring the low copy number pL-P plasmid grown on L-agar containing X-P. The single highlighted dark blue colony was sub-cultured and the resulting mixed population is shown in panel C.
Figure 5.8 Variable PhoA phenotypes from an IPTG-inducible hek-phoA fusion with increasing IPTG concentration. A single culture of *E. coli* K1 strain RS218 harbouring pMIP was replica plated onto L-agar plates containing the indicated concentration of the inducer IPTG (A-E).
5.2.3.4 *Addition of regulatory region upstream of hek-phoA fusions*

In order to examine the effect of the upstream regulatory region on the expression of the Hek-PhoA fusion protein the *hek-phoA* fusion gene from pL-P was substituted for the wild-type *hek* gene in plasmid pH+ (Table 5.1). The 3' portion of the *hek* gene and the fused *phoA* gene were excised from pL-P with *BsaBI* and *SacI*. The pH+ plasmid was digested with the same enzymes to remove the 3' portion of the *hek* gene and the plasmid backbone was retained. The *hek-phoA* fragment was then ligated to the pH+ backbone yielding pH+P (Table 5.1). pH+P contains a complete *hek-phoA* fusion gene with the 1138 bp of upstream DNA in the high-copy pBSKII backbone. The insert from pH+P was subsequently subcloned into the low copy vector pPD101 following excision with *KpnI* and *SacI* and ligation to pPD101 cut with the same enzymes. The resulting low copy plasmid containing the *hek-phoA* gene with 1138 bp of upstream DNA was designated pL+P (Table 5.1). *E. coli* K1 strain RS218 harbouring either pH+P or pL+P had similar colony morphologies and colouring on X-P as strains harbouring the respective promoter-only constructs, pH-P and pL-P. Both plasmids also conferred the same variable PhoA phenotype as described above.

5.2.3.5 *Quantitative analysis of PhoA activity from low copy constructs*

The expression of the Hek-PhoA fusion protein in overnight cultures of *E. coli* RS218 harbouring pL-P or pL+P was directly quantified by assaying alkaline phosphatase activity using the chromogenic pNPP substrate. The plasmid containing the additional regulatory region did appear to reduce the expression of PhoA slightly: RS218 harbouring pL-P gave 299.4 arbitrary units (±2.1), or harbouring pL+P gave 247.1 arbitrary units (±22.8), compared with only 3.1 arbitrary units (±2.2) with RS218 alone. Although this difference is not as great as that observed for Hek protein expression from the pL- and pL+ plasmids (section 5.2.2.5) it does fit the same trend with the addition of upstream DNA reducing expression. The high error shown above for *E. coli* RS218 harbouring pL+P reflects an underlying variation in the level of PhoA activity with all of the examined strains. In some cases the PhoA activity in duplicate cultures varied by as much as 10 %. This variation perhaps highlights the underlying variation described for Hek protein expression. Additional experimentation would be required to clarify this phenomenon.
5.2.4 Analysis of *hek* transcription using a *lacZ* transcriptional reporter

Expression of the Hek protein in *E. coli* K1 appeared to be variable and this variation appeared to be, at least in part, independent of native promoter (section 5.2.2.2). In order to examine if any degree of variation occurred at the transcriptional level a series of transcriptional fusions between the putative *hek* gene promoter and the *lacZ* reporter gene (encoding the enzyme beta-galactosidase) were constructed.

5.2.4.1 Construction of *lacZ* transcriptional fusion to the putative *hek* promoter

The 76 bp immediately upstream of the *hek* ORF in plasmid pH- (containing the putative *hek* promoter) and the first 233 bp of the *hek* gene were cloned upstream of the promoterless *lacZ* gene in plasmid pQF50 (Table 2.2). The *hek* promoter fragment was excised from plasmid pH- by digestion with *SalI* and *BsaBI* and ligated into pQF50 cut with *SalI* and *SmaI*. Bacteria harbouring correct *hek* promoter-*lacZ* constructs were found to produce pale blue colonies when grown on L agar containing the beta-galactosidase chromogenic substrate X-Gal compared with the white colonies formed by bacteria harbouring pQF50. The structure of a single correct plasmid clone was confirmed by restriction endonuclease mapping and the plasmid was designated pM-Z (Table 5.1).

5.2.4.2 Transcriptional analysis of the putative *hek* promoter in *E. coli* K-12

The *hek* promoter-*lacZ* reporter plasmid pM-Z and the parental pQF50 were transformed into the *E. coli* K-12 strain MC4100 and the beta-galactosidase activity was measured using the standard Miller assay (168). Bacteria harbouring the pM-Z plasmid produced 50.5 Miller units of beta-galactosidase compared with 0.5 Miller units from bacteria harbouring the parental plasmid control pQF50. This suggests that the 309 bp of *E. coli* K1 strain RS218 DNA cloned upstream of the *lacZ* gene in pM-Z contained a functional but relatively weak promoter. To assess if the growth media used could affect the level of transcription from this promoter *E. coli* K-12 MC4100 harbouring either pM-Z or pQF50 were grown overnight in LB broth, L0 (LB broth without salt), CFA broth, BHI broth or BHI supplemented with 50 % heat-inactivated
foetal bovine serum (FBS) and the beta-galactosidase activity was measured as before. The media used have all been shown to affect *E. coli* gene expression in other model systems: low osmolarity (L0) affects the expression of many genes through the OmpR/EnvZ two-component system (170), CFA is known to enhance fimbral expression in ETEC (239), BHI has been found to enhance invasion by *E. coli* K1 and this effect is even greater in the presence of 50% serum (10). The beta-galactosidase activity of the test samples (*E. coli* MC4100 harbouring pM-Z) was found to vary slightly depending on the growth media (Fig. 5.9 A). However, the different growth media also affected the level of activity detected in the negative control strain (*E. coli* MC4100 harbouring pQF50). In order to normalise this data and allow comparison of the different test samples, the values for each negative control were subtracted from the appropriate test value. Transcription was highest in cultures grown in CFA broth (100.7 miller units), followed by BHI/FBS (84.6 miller units), L broth (65.2 miller units), BHI broth (63.7 miller units) and L0 (41.5 miller units).

### 5.2.4.3 Construction of a Lac-* E. coli* K1 RS218 derivative strain

As *E. coli* K1 strain RS218 is naturally Lac- it was necessary to mutagenise the endogenous lac operon to allow analysis of transcription using the pM-Z plasmid. As a method of phage-transduction was not available for *E. coli* K1 at this time a disrupted lac allele was constructed in vitro and transferred to the *E. coli* RS218 chromosome using the λ Red allele replacement system (Fig. 5.10). Datsenko *et al* have reported that as little as 36 bp of DNA sequence homology is required to promote recombination in laboratory strains of *E. coli* using their version of the λ Red gene disruption system, pKD46 (Table 2.2) (46). A pair of primers were designed to amplify the cat gene from plasmid pKD3 (Table 2.2) with 5' extensions that were homologous to regions of the *E. coli* RS218 chromosome upstream and downstream of the lacZ gene respectively. Primer RSlac/catf had 36 bp homology to a region upstream of the *E. coli* RS218 lacZ gene and RSlac/catr had 41 bp homology to a region 3661 bp downstream of RSlac/catf (Table 2.3). PCR amplification using these primers and the pKD3 plasmid as a template generated an approximately 1 kb DNA fragment containing the cat gene (encoding chloramphenicol resistance) with homology to the *E. coli* RS218 lac operon at each end (Fig. 5.10 A). However, allele replacement in *E. coli* K1 strain RS218 harbouring the plasmid pKD46 using this DNA fragment and the conditions described
by Datsenko et al (2000) proved unsuccessful. Allele replacement in *Shigella* spp. by a similar approach requires 300 – 500 bp of homology either side of the selectable marker using the pKOBEGA $\lambda$ Red plasmid (Table 2.2) (C. Beloin, personal correspondance). Additional fragments of upstream (504 bp) and downstream (369 bp) *E. coli* RS218 lac DNA were generated by PCR using the primer pairs lac_downstream-f and -r and lac_upstreamf and -r (Table 2.3). The additional DNA fragments were designed such that the upstream fragment shared a 20 bp overlap with the 5' end of the lac-cat fragment and the downstream fragment shared a 20 bp overlap with the 3' end of the lac-cat fragment (Fig. 5.10 B). The three fragments were then fused together in a single PCR reaction by gene SOEing (108), resulting in an approximately 2 kb lac-cat DNA fragment containing the cat gene with long flanking regions of *E. coli* RS218 lac DNA (Fig. 5.10 C). As PCR amplification of the lac-cat fragment proved quite inefficient the DNA fragment was ligated into plasmid pBSKII cut with EcoRV to generate the plasmid pLacCat (Table 2.2). The lac-cat fragment was excised from plasmid pLacCat using KpnI and SacI and used for allele replacement in *E. coli* RS218 harbouring the pKOBEGA plasmid (33) as described in Chapter 2, section 2.2.3. *E. coli* RS218 lac mutants were selected on the basis of chloramphenicol resistance and loss of the Lac+ phenotype on agar containing X-Gal. A single mutant was confirmed by PCR analysis and designated RSlac (Table 2.1).

### 5.2.4.4 Transcriptional analysis of the putative hek promoter in *E. coli* K1 strain RSlac

The construction of the *E. coli* RSlac strain allowed analysis of transcription from the putative hek promoter using the lacZ reporter fusion plasmid pM-Z (section 5.2.4.1) in the native *E. coli* K1 background. As a control, a fimE promoter-lacZ fusion, pSag1 (Table 2.2), was included. The fimE gene in *E. coli* encodes one of the site-specific recombinases responsible for the inversion of the fim switch and phase variation of type 1 fimbriae (238) and is weakly transcribed (84). *E. coli* RSlac was transformed with pM-Z, pSag1 and the parental plasmid pQF50 and the beta-galactosidase activity was measured as before. In overnight cultures the beta-galactosidase levels with pM-Z and pQF50 were similar to those observed with *E. coli* K-12 (41 Miller units for *E. coli* RSlac harbouring pM-Z and 1.2 Miller units with the vector control pQF50). Strains harbouring the fimE fusion pSag1 gave 40.5 Miller units indicating that the hek
Figure 5.9 Analysis of transcription from the putative hek promoter. Panel A: *E. coli* K-12 strain MC4100 harbouring the pM-Z plasmid or the vector control pQF50 were grown in standard LB medium, L0 (LB without salt), CFA, BHI or BHI/FBS (1:1). After overnight growth the beta-galactosidase activity of each culture was determined. In each case the background activity from the pQF50 vector control was subtracted from the pM-Z value to eliminate variation due to the differing media. All standard deviations were less than 10%. Panel B: *E. coli* K1 strain RSlac harbouring pM-Z of pQF50 grown for 24 or 48 h in static LB cultures, overnight in shaken LB or in DMEM tissue culture medium. The beta-galactosidase activity of each culture was determined and background activity was subtracted as before.
**Figure 5.10** Construction of disrupted lac operon *in vitro*. Panel A: PCR amplification of the *cat* gene with primers RSlac/catf and RSlac/catr. The primers are represented with half arrows and the red portions represent homology to the *E. coli* RS218 lac operon. Panel B: PCR amplification of additional up and downstream lac operon DNA using primer pairs lac upstream-f and -r and lac downstream-f and -r. The regions in red are homologous to the 5' extensions of RSlac/catf and RSlac/catr. Panel C: SOEing PCR to combine the three separate PCR products into a single amplimer.
promoter is very similar in strength to that of fimE. E. coli RS lac harbouring either plasmid grown in DMEM/Ham's F-12 (with 10% heat-inactivated FBS) or in static L broth cultures were then compared with standard shaken L broth cultures to assess if invasion assay conditions or conditions known to enhance type 1 fimbriae expression respectively had any effect on transcription from the putative hek promoter. As the static cultures grew quite slowly they were sampled after 24 h and again after 48 h. The different growth conditions were found to have only a slight effect on the beta-galactosidase levels (Fig. 5.9 B). Expression was highest in shaken L broth (44.1 Miller units) followed by shaken DMEM/Ham's F-12 (36.5 Miller units), static L broth after 48 h (31.2 Miller units) and static L broth after 24 h (18.6 Miller units).

5.2.4.5 Cloning of the hek regulatory region upstream of a promoterless-lacZ in plasmid pQF50

As the presence of additional upstream DNA was found to affect the expression of Hek protein (section 5.2.2.5) a series of lacZ transcriptional reporter plasmids were constructed containing the putative hek promoter alone or with an additional 1 kb of upstream E. coli RS218 DNA. The 1138 bp upstream of the hek ORF in plasmid pH+ (Table 5.1) and the first 233 bp of the hek gene were excised with XhoI and BsaBI, ligated to pQF50 cut with SmaI and SalI and transformed into E. coli K-12 strain XL-1 (Fig. 5.11 A). Bacteria displaying a Lac+ phenotype were found to fall into two distinct populations, one being considerably bluer than the other on agar containing X-Gal. The multiple-cloning site (MCS) of plasmid pQF50 contains two adjacent SmaI sites (separated by 8 bp) (Fig. 5.11 B). The two distinct Lac+ populations were found to correlate with an insert utilising one or other SmaI site; the darker blue colonies contained inserts using the SmaI site proximal to the lacZ gene in pQF50 and the lighter blue colonies contained inserts in the upstream SmaI site. One clone of each population was retained and designated pM+ZS2 (dark blue) and pM+ZS1 (pale blue) (Fig. 5.11, Table 2.2). The lacZ gene in pQF50 has its own Shine-Delgarno sequence and there are translational stop codons in all three reading frames between the MCS and the lacZ gene. It is therefore surprising that this 8 bp difference in insertion point had such a dramatic influence on the level of transcription of lacZ.
To allow a direct comparison with the new pM+ZS1 and pM+ZS2 constructs the shorter promoter-only region from pH- was re-cloned into pQF50 (see Section 5.2.4.1). As before, pH- (Table 5.1) was cut with Sall and BsaBI and ligated into pQF50 cut with Sall and Smal (Fig. 5.11 A). Two different Lac+ populations were identified and were again found to correlate with one or other Smal insertion site (Fig. 5.11 B). A representative clone from each population was retained and designated pM-ZS2 (dark blue) and pM-ZS1 (pale blue) (Fig. 5.11, Table 5.1).

5.2.4.6 Transcriptional analysis of the hek promoter in E. coli K-12 and K1

Cultures of E. coli K-12 strain XL-1 (Table 2.1) harbouring each of the new hek promoter-lacZ fusions (pM-ZS1, pM-ZS2, pM+ZS1 and pM+ZS2) or the vector control pQF50 were grown overnight and the beta-galactosidase activity was measured as before (Fig. 5.12 A). The level of beta-galactosidase activity observed for each strain mirrored the observations of colour intensity of plate cultures, bacteria harbouring pM-ZS1 (51.5 Miller units) or pM+ZS1 (86.2 Miller units) displayed lower activity than those harbouring pM-ZS2 (91.9 Miller units) or pM+ZS2 (422.9 Miller units) respectively. Surprisingly the constructs with additional upstream DNA (pM+ZS1 and 7) resulted in higher levels of activity than the respective promoter-only constructs. This finding is in direct disagreement with the observations of Hek protein levels from low-copy number plasmids with the same upstream sequences (section 5.2.2.5). All four reporter plasmids and the vector control were then transformed into the E. coli K1 strain RS218 and the beta-galactosidase activity was again assayed (Fig. 5.12 B). The same pattern of activity as in E. coli XL-1 was observed: E. coli RS218 harbouring pM-ZS1 (23.9 Miller units), pM+ZS1 (77.5 Miller units), pM-ZS2 (36.1 Miller units) or pM+ZS2 (133.3 Miller units). Once again these findings contradicted the observed levels of Hek protein expression (section 5.2.2.5). Although the pattern of activity was similar in both E. coli K-12 and K1 strains there was a striking difference in the levels of expression from each plasmid in the two strains. This suggested that the hek promoter was subject to some form of negative regulation in the E. coli K1 strain that was absent in the E. coli K-12 background. The genome of E. coli K1 strain RS218 (www.genome.wisc.edu/) contains large regions of additional DNA compared to the sequenced E. coli K-12 strain MH1655 (15) allowing many possible candidate proteins for this apparent negative regulation. Alternatively, the putative regulator may be
Figure 5.11 Construction of *hek* promoter-*lacZ* transcriptional fusions. Panel A: Plasmids pH- and pH+ were digested with *SalI* and *BsaBI* or *XhoI* and *BsaBI* respectively and the resulting 332 bp or 1371 bp fragments were ligated to pQF50 cut with *SalI* and *SmaI*. Panel B: A representative map of pQF50 is shown highlighting the duplicated *SmaI* site in the MCS and showing the insertion points of plasmids pM-ZS1 and S2 and pM+ZS1 and S2. The colony morphology and colour of *E. coli* K1 strain XL-1 harbouring each of the four plasmids is also shown.
Figure 5.12 Transcriptional analysis of the putative hek promoter in *E. coli* K-12 and K1. Beta-galactosidase activity of *E. coli* K-12 strain XL-1 (panel A) or *E. coli* K1 strain RSlac (panel B) harbouring each of the four hek-lacZ transcriptional reporter plasmids or the vector control pQF50. Panel C: beta-galactosidase activity of *E. coli* K-12 co-transformed with pPD101 or pL- and each of the four transcriptional fusion plasmids or the vector control pQF50.
present in both \textit{E. coli} lineages but expressed at different levels. One potential candidate protein that is expressed by \textit{E. coli} K1 but not K-12 is the Hek protein itself and negative autoregulation of its own promoter could explain the observed differences in transcription levels. In order to examine this possibility, \textit{E. coli} K-12 strain XL-1 was co-transformed with all four of the \textit{hek} promoter reporter plasmids or the vector control pQF50 and the Hek-encoding plasmid pL- or its vector control pPD101 (Table 2.2 and 5.1). The beta-galactosidase activity of the ten resulting bacterial strains was compared to examine if the expression of Hek protein affected the level of transcription from the \textit{hek} promoter (Fig. 5.12 C). However, neither the pL- or pPD101 plasmids were found to have any significant affect on the level of beta-galactosidase activity.

5.2.4.7 Identification of putative regulators of \textit{hek} transcription

The addition of further \textit{E. coli} RS218 DNA upstream of the putative \textit{hek} promoter in plasmid expression constructs was found to affect both the transcription from the promoter and the levels of Hek protein expressed, albeit in an apparently contradictory manner. This data suggested that the DNA immediately upstream of the putative promoter contained regulatory sequences which impact on Hek expression. As this effect occurs in both \textit{E. coli} K-12 and K1 backgrounds the regulation is likely to involve a transcription factor that is conserved across \textit{E. coli} and not an \textit{E. coli} K1-specific regulator. Furthermore, the transcription from the \textit{hek} promoter was greater in \textit{E. coli} K-12 than K1 suggesting a further \textit{E. coli} K1-specific aspect to the regulation of \textit{hek}.

In order to identify potential regulators of \textit{hek} transcription the region upstream of the \textit{hek} ORF and the 5' portion of the \textit{hek} gene itself was searched for the consensus binding sequences of several \textit{E. coli} global transcriptional regulators, including IHF, Lrp, CRP and FNR. The search identified three possible IHF binding sites, each with a single mismatch from the consensus binding-site of WATCARXXXXXTTR (where W = A or T, R = A or G and X = A, T, C or G). Binding sites with as many as three mismatches have been previously identified as positive IHF binding sites (192). One binding site is located inside the \textit{hek} ORF immediately following the first ATG codon, the second lies mid-way between the predicted -10 box of the promoter and the Shine-Delgarno sequence and the third is located 460 bp upstream of the putative -35 box.
(Fig. 5.13 A). The next best match was an Lrp binding site 410 bp upstream of the −35 box with 3 mismatches.

5.2.4.8 Construction of an E. coli K1 double lac and ihfA mutant

In order to study the role of IHF in the regulation of transcription from the hek promoter a mutation in the ihfA gene, encoding the α-subunit of IHF, was transduced from E. coli K-12 strain CJD1484 (Table 2.1) into the E. coli K1 Lac- strain RSlac. Transduction was carried out using a modified P1 phage as described in Chapter 2. The ihfA mutation in E. coli CJD1484 consists of an 82 bp deletion linked to a Tn10 insertion and this mutation had been previously shown to completely abolish IHF function at the invertible DNA element which is responsible for phase-variation of type 1 fimbriae in E. coli (59). Successful transfer of the ihfA deletion into E. coli RSlac was confirmed by PCR and subsequent western immunoblot analysis using a Salmonella IhfA-specific antibody (Fig. 5.13 B). The Δlac ΔihfA E. coli K1 strain was designated RSlac ihf.

5.2.4.9 Analysis of hek transcription in E. coli RSlac ihf

The beta-galactosidase activities of E. coli RSlac ihf or RSlac harbouring pM-ZS1, pM+ZS1 or pQF50 were compared to assess if the ihfA mutation had any effect on transcription from the hek promoter (Fig. 5.13 C). The ihfA mutation had no effect on the beta-galactosidase activity of strains harbouring plasmid pM+ZS1 (RSlac 77.5 Miller units, RSlac ihf 86.2 Miller units). However the ihfA lesion did have a dramatic effect on the activity from plasmid pM-ZS1 (RSlac 23.9 Miller units, RSlac ihf 83.5 Miller units). This trend was also confirmed with the remaining pair of hek promoter-lacZ fusions, pM-ZS2 and pM+ZS2 (not shown). These data suggest that the lower level of transcription from plasmids lacking the upstream region (promoter-only constructs) is dependent on IHF. Furthermore, this effect must involve one or both of the IHF binding sites proximal to the hek promoter as the third upstream binding site is absent in these promoter-only constructs.
A:  

![Graphical representation of the hek gene](image)

**Panel A:** graphical representation of the first third of the hek gene and the 1100 bp immediately upstream. The extent of the insertions in the pM-ZSl and pM+ZSl plasmids and the locations of putative IHF binding sites are indicated.

B:  

![Western immunoblot analysis](image)

**Panel B:** Western immunoblot analysis of IhfA (IHF α-subunit) in *E. coli* RS218 (lane 1), RSlac ihf (lane 2), RSihf (lane 3), VL386 (lane 4) CH1484 (lane 5) and CH1484 complemented with ihfA on the low copy pBR322 plasmid (lane 6). The bands corresponding to IhfA are indicated.

C:  

![Beta-galactosidase activity](image)

**Panel C:** beta-galactosidase activity in *E. coli* K1 RSlac and RSlac ihf harbouring the pM-ZSl, pM+ZSl and pQF50 plasmids.

**Figure 5.13** IHF and transcription of the hek gene. Panel A: graphical representation of the first third of the hek gene and the 1100 bp immediately upstream. The extent of the insertions in the pM-ZSl and pM+ZSl plasmids and the locations of putative IHF binding sites are indicated. Panel B: Western immunoblot analysis of IhfA (IHF α-subunit) in *E. coli* RS218 (lane 1), RSlac ihf (lane 2), RSihf (lane 3), VL386 (lane 4) CH1484 (lane 5) and CH1484 complemented with ihfA on the low copy pBR322 plasmid (lane 6). The bands corresponding to IhfA are indicated. Panel C: beta-galactosidase activity in *E. coli* K1 RSlac and RSlac ihf harbouring the pM-ZSl, pM+ZSl and pQF50 plasmids.
5.2.4.10 Variation of hek transcription

When *E. coli* K1 strain RSlac harbouring any of the *hek* promoter-lacZ fusion plasmids were grown on L agar containing the beta-galactosidase indicator X-Gal occasional colonies were observed that were a darker blue than the rest. When sub-cultured these darker colonies gave rise to mixed populations of pale and darker blue colonies. This suggested that transcription from the *hek* promoter may be phase variable. However, no clear phase-Off colonies were observed suggesting that if phase variation were occurring it was between discreet levels of expression rather than On to Off. An example of such variation was recently reported for the *nadA* gene in *Neisseria meningitidis* where a relatively small change in the level of transcription resulted in phase-variation at the protein level (156). As variation occurred in *E. coli* K1 strains harbouring all four of the reporter plasmids the minimum sequence necessary to support variation was the short region of DNA cloned into the pM-ZS1 and 9 plasmids. No Dam methylation sites (GATC) were found within this region and no evidence suggested the presence of an invertible DNA element. In order to examine if the variable beta-galactosidase expression was associated with DNA sequence changes within this region, the promoter inserts from both pale and dark blue colonies were sequenced. A dark blue colony of *E. coli* strain RSlac harbouring plasmid pM-ZS1 was sub-cultured onto fresh agar containing X-Gal resulting in a mixed population of pale and dark blue colonies. Five representative pale and five dark colonies were selected. Each of the ten colonies were boiled and used as templates in PCR reactions with primers pQF50f and pQF50r (Table 2.3), designed to amplify the pQF50 insert. The ten amplified inserts were sequenced using the pQF50r primer and the sequences were compared with the available *hek* promoter sequence from the unfinished *E. coli* RS218 genome-sequencing project (www.genome.wisc.edu/). No consistent differences were observed in the sequences of either group of inserts that would suggest a slipped-strand mispairing mechanism of variation at the *hek* promoter. The observed variation may perhaps be an indirect effect, caused by variation in the expression of a transcriptional, or even post-transcriptional, regulator of *hek.*
Variability in the expression of surface exposed proteins is a common theme among many bacterial species (2, 23, 100, 141, 156, 259). The dangers posed by the host’s immune response have presumably imposed selective pressure favouring the development of complex regulatory mechanisms by which the expression of potentially antigenic surface components can be limited to functional necessity. Expression of the Hek protein from a high copy number plasmid in the *E. coli* K1 clinical isolate RS218 initially appeared to be phase-variable, with phase-off being the dominant state. Expression of the Hek protein from an inducible P_{inc} promoter in the same strain confirmed that the Hek protein could be over-expressed in *E. coli* K1 but small variations in the level of Hek protein expression among bacteria harbouring this plasmid suggested that the expression was indeed variable and that this variation was independent of the *hek* gene’s native promoter. However, when the *hek* gene with its native promoter was transferred into a low copy number plasmid vector the apparent variation was abolished. This suggested that the extreme on to off variation observed with high copy expression was an artefact of that high copy number plasmid. Furthermore, the difference in the level of Hek expression between the low copy number plasmid and the native chromosomal gene did not correlate with the difference in gene copy number. Bacteria harbouring the low copy plasmid produced greater than two-fold more protein than would be expected based on gene copy number alone. This indicated that the chromosomally encoded *hek* gene was subject to some form of negative regulation absent from the plasmid constructs. In an attempt to restore this regulation to the plasmid-borne *hek* gene a large fragment of RS218 chromosomal DNA from upstream of the *hek* gene was cloned upstream of the gene on the low copy number plasmid. The addition of approximately 1 kb of upstream DNA was found to restore *hek* regulation and resulted in levels of Hek protein expression that correlated well with gene copy number. This finding indicated that the *hek* gene is subject to regulation involving sequences within the 1 kb upstream of the gene. Using these various plasmid constructs in *E. coli* K-12 it was also demonstrated that the efficiency of Hek-promoted invasion of CHO-K1 cells was proportional to the level of Hek protein expression.
In order to support these data a series of plasmids were constructed carrying a \textit{hek-phoA} fusion gene under the control of either the native \textit{hek} promoter, in both high and low copy number, or the IPTG-inducible \textit{P}_{\text{te}} promoter. The \textit{phoA} gene, encoding the enzyme alkaline phosphatase, provided a readily assayable reporter of Hek expression and export, as the PhoA protein is only enzymatically active in the periplasm. \textit{E. coli} K1 harbouring either the low or high copy number native promoter construct displayed variable PhoA phenotypes. These PhoA phenotypes supported the conclusion that Hek expression is variable between two distinct levels of expression rather than on to off. Bacteria harbouring either plasmid gave rise to pale and dark blue colonies on agar containing X-P. These distinct phenotypic groups were also apparent with \textit{E. coli} K1 harbouring the inducible fusion gene and probably correlate with the two distinct levels of Hek protein expression observed with the inducible pMI in \textit{E. coli} K1 (section 5.2.2.2, Fig 5.1 B). Considered together these data suggest that expression of Hek in \textit{E. coli} K1 is variable between two distinct levels of expression and that this variation is independent of the native \textit{hek} promoter.

The high copy native promoter fusions also produced a third colony morphology, small colonies with a white halo and only a small blue core. Close examination of the dark blue colonies also reveals occasional colonies with white segments seemingly growing out from the original dark blue colony. Similar white colonies were also apparent with the inducible fusion when induced with high levels of IPTG. This suggests that the PhoA negative phenotype results from extreme high level expression of the Hek-PhoA fusion protein rather than variation of Hek expression \textit{per se}. This could also explain why Hek protein expression in \textit{E. coli} K1 appeared highly variable when harbouring the high copy number and de-repressed pH- plasmid but not with the normally regulated high copy pH+ or either of the low copy plasmids, pL- or pL+ (section 5.2.2.3 and 5.2.2.5). This phenomenon indicates that very high level expression of Hek is not tolerated well in \textit{E. coli} K1. The apparent repression of Hek expression may be the result of a regulatory mechanism inherent to \textit{E. coli} K1 strain RS218 or may be the result of escape mutations to relieve a toxic effect of Hek over-expression. The second explanation seems the more plausible as this effect was not apparent when Hek was expressed in low copy number or from the native promoter with upstream regulatory
elements. Whichever explanation is true this phenomenon does not appear to occur at levels of expression relevant to those observed in the original \textit{E. coli} K1 strain RS218.

Hek-PhoA fusion plasmids with and without the upstream regulatory region were also used to further examine the effect of this region on expression of the Hek protein. The expression of the Hek-PhoA fusion was lower in the plasmid containing the upstream region compared with a similar plasmid containing the putative \textit{hek} promoter alone. This trend supports the western immunoblot data showing lower levels of Hek expression when the upstream regulatory region was included. However the difference in expression detected with the PhoA fusions was considerably lower than that observed with the wild-type Hek protein by western immunoblotting. This poor correlation may be an artefact of the Hek-PhoA fusion protein itself. It is possible that the apparent poor toleration of high level Hek expression and the added bulk of the Hek-PhoA fusion protein results in lower expression and/or export of the fusion. Such an effect would have a greater impact on the promoter-only plasmid, which produces more of the Hek-PhoA fusion. This could explain the reduced relative difference in fusion expression between the promoter-only and regulatory region containing plasmids.

In order to examine transcription from the native \textit{hek} promoter a \textit{lacZ} transcriptional reporter was constructed using the pQF50 reporter plasmid. A construct containing approximately one third of the \textit{hek} gene and 76 bp of upstream DNA gave a relatively low level of beta-galactosidase activity indicating the presence of a weak promoter within this insert. The level of transcription from this promoter in a Lac- \textit{E. coli} K1 strain was similar to that of another weak \textit{E. coli} promoter, that of the \textit{fimE} gene, encoding one of the \textit{fim} switch site-specific recombinases responsible for type 1 fimbriae phase variation.

As the region upstream of the \textit{hek} gene on the \textit{E. coli} RS218 chromosome was suspected to contain regulatory sequences involved in the expression of Hek, additional \textit{lacZ} fusion plasmids were constructed containing the \textit{hek} promoter with and without this upstream region. The presence of these upstream sequences in a low-copy Hek-expression plasmid resulted in up to 8.4-fold lower protein expression in \textit{E. coli} K-12
and at least 3.4-fold lower in *E. coli* K1. However, when the *hek* promoter-*lacZ* fusions with and without the upstream region were compared the opposite trend was observed. The presence of the upstream region resulted in between 1.7- and 4.6-fold higher levels of transcription in *E. coli* K-12 and between 3.2- and 3.7-fold in *E. coli* K1. These findings directly contradict the western immunoblot data examining Hek protein expression. No obvious explanation presents itself to reconcile these contradictory findings and any possible explanation remains speculative without further data. The variation in expression of the Hek protein and the Hek-PhoA fusion from the inducible P_{rc} promoter suggests post-transcriptional regulation of Hek expression. It is possible that post-transcriptional regulation controls the production of Hek protein to such a degree as to disregard any transcriptional effects although this seems unlikely. A second possibility is that the transcription of *hek* in constructs containing the additional upstream DNA involves a different promoter (Fig. 5.14) but it is impossible to test this possibility without performing primer-extension analysis on the *hek* transcript. If this were true the different transcripts could have very different properties in terms of RNA stability, possible secondary structure and even efficiency of translation. Such an explanation would involve a second more efficient upstream *hek* promoter producing more transcript than the promoter proximal to the *hek* gene but producing a transcript that was less well translated. This would not affect the translation of *lacZ* in the reporter fusions as the *lacZ* gene used has its own Shine-Delgarno sequences and does not rely on translational signals from the gene fused upstream. The result would be increased beta-galactosidase activity in the *lacZ* reporter fusions with the additional upstream DNA but less Hek protein expression in the expression constructs with the same upstream DNA. The most likely explanations for poor translation of a longer *hek* transcript would be secondary structure within the mRNA which interfered with translation or the binding of a regulatory protein to the *hek* mRNA modulating its translation. A third possibility involving reduced stability of the longer *hek* transcript appears less likely as this would presumably also affect the *hek-lacZ* transcript and hence reduce beta-galactosidase activity.

An alternative explanation for the difference in the level of transcription from the putative *hek* promoter or the promoter with additional upstream DNA was the presence of regulator binding sites upstream of the *hek* promoter. The DNA sequence around
and upstream of the putative *hek* promoter was searched for the consensus binding sites of several characterised regulator proteins including IHF, CRP, Lrp and FNR. Several candidate binding sites were identified with the most promising being 3 potential IHF binding sites, one far upstream of the *hek* ORF, one between the putative -10 and Shine-Delgarna sequences and one within the 5' end of the *hek* ORF (Fig. 5.13 A). The transcription of *hek* was examined using all four of the lacZ transcriptional reporters in a *ihfA* mutant derivative of the Lac- *E. coli* K1 strain RSlac (*ihfA* encodes the α-subunit of IHF). The *ihfA* mutation was found to have no effect on transcription from constructs containing the extended upstream region but resulted in increased expression from the constructs containing only the putative promoter. This finding also fits the two-promoter model described above (Fig. 5.14). IHF binding to sites proximal to the putative *hek* promoter may repress transcription from this promoter without affecting transcription from a second upstream promoter.

In addition to the difference in transcriptional levels between *hek-lacZ* fusion plasmids with and without the additional upstream DNA there was also a small degree of variation observed between bacteria harbouring the same plasmid. This variation occurred with all of the *hek-promoter-lacZ* fusion plasmids and involved a change between pale and darker blue colony colour on agar containing X-Gal. To examine if this variation was due to changes in the DNA sequence in and around the putative *hek* promoter the insert from plasmid pM-ZS1 in pale and darker blue colonies was amplified by PCR and sequenced. No differences were observed in the DNA sequences of the five pale and five dark blue colonies examined. This demonstrates that no phase-variation event involving DNA recombination/inversion or slipped-stranded mispairing is occurring at the putative *hek* promoter. This region also contains no GATC sequences thus ruling out DNA methylation as a possible cause of variation. The observed variation may perhaps involve an additional regulatory factor interacting within the region surrounding the putative promoter or perhaps interacting with the *hek-lacZ* mRNA transcript.

In summary, the *hek* gene was found to have a weak promoter within the 76 bp immediately preceding the gene on the *E. coli* RS218 chromosome and this weak promoter is subject to repression by the nucleoid-associated protein IHF. The level of
**Figure 5.14** Two promoter model of *hek* transcription and translation. A possible explanation of the contradictory effect of the upstream regulatory region on transcription of the *hek* gene (as examined using *lacZ* fusions) compared with production of Hek protein (as examined by western immunoblotting). The poor translation of the putative longer transcript could be explained by secondary structure in the mRNA (perhaps sequestering the Shine-Delgarno region) or by an RNA-binding regulatory protein.
Hek protein expression from this promoter on a low copy number (2-4 copies/cell) plasmid did not correlate with the observed exogenous protein expression in *E. coli* RS218. Bacteria harbouring the low copy plasmid produced more Hek protein than could be accounted for by gene copy number alone. The replacement of the 76 bp promoter region with an extended 1138 bp of upstream *E. coli* RS218 DNA restored the apparent repression. However, when the level of *hek* transcription was measured using plasmids containing a *hek-lacZ* fusion with the short (76 bp) or long (1138 bp) upstream region the exact opposite trend was observed. The presence of the extended upstream region greatly increased the observed level of transcription. In order to reconcile these apparently contradictory findings a model of Hek expression was devised which proposed the existence of a second promoter in the extended upstream region capable of driving *hek* transcription. This second promoter would be stronger than the alternative promoter immediately upstream of the *hek* gene but would produce a transcript which was less well translated, either due to secondary structure in the mRNA or the action of an RNA-binding regulatory protein.

The expression of plasmid-encoded Hek was also subject to a small degree of variation and this variation also occurred with plasmids where *hek* transcription was driven from an inducible promoter. This variability was confirmed using plasmids expressing a Hek-PhoA fusion from the native or inducible promoter. *hek* transcription was also subject to variation and this variation was not linked DNA methylation or changes and the DNA sequence level. It is conceivable that this variation is an indirect effect caused by the variation of a regulatory protein that is involved in the expression of Hek. Additional work is required to assess if this variation also occurs in exogenous expression of Hek and is not a plasmid-specific phenomenon.
Chapter 6  Proteomic analyses of invasion deficient mutants of *E. coli* K1
6.1 Introduction

6.1.1 Proteomics

6.1.1.1 2-D PAGE

Several techniques are available to study the result of mutations or growth conditions on global gene expression in bacteria including DNA microarray analysis. DNA microarrays provide a powerful tool to directly measure the amount of transcription of virtually every gene in many species, including \textit{E. coli} K-12. However, arrays are not available for every organism (eg. \textit{E. coli} K1) and even where arrays do exist changes at the transcriptional level do not necessarily always reflect changes in the copy number of a protein species. In order to directly study the end-products of gene expression, the proteins themselves, a number of powerful techniques have been developed which form the basis of proteomics. The term proteomics refers to the study of the protein complement (or proteome) of a given organism. In order to study a proteome specialised techniques are required which facilitate the separation and visualisation of as many proteins as possible. Techniques such as one-dimensional SDS-PAGE are severely limited in their ability to separate large numbers of proteins, as proteins with similar molecular weights are difficult or impossible to resolve. The development of two-dimensional polacrylamide electrophoresis (2-D PAGE) has provided the ability to separate potentially hundreds of distinct protein species from a single sample (185) and directly compare and quantify their expression. The first dimension of a 2-D PAGE gel involves the separation of proteins on the basis of their isoelectric point (pl) on an acrylamide strip containing a defined pH gradient. The pl of a protein is the pH at which the protein is electrically neutral and is a function of the amino acid composition of the protein. Proteins at a pH lower than their pl hold a net positive charge and vice versa. When the proteins in a pH gradient are subjected to an electrical field they migrate along the strip until they reach a pH equal to their pl and become electrically neutral, a process known as isoelectric focussing. The pH gradient strip is then placed on top of a standard SDS-polyacrylamide gel and the proteins are separated on the basis of their molecular weight in a second dimension. The first dimension pH gradients
were originally created using solutions of carrier ampholytes with a range of different pI values which migrate in an electrical field to produce a pH gradient. However, the resulting gradients were quite unstable and difficult to reproduce precisely. More recently immobilised pH gradients have been adopted which consist of polyacrylamide strips containing covalently attached acidic and basic buffering molecules (IPG strips) (274). As the buffering moieties are covalently attached to the gel the pH gradient remains stable throughout sample separation and the IPG strips are mass-produced resulting in consistent gradients.

6.1.1.2 MALDI-TOF

The ability to separate and visualise hundreds of protein species is of very limited use without the further ability to identify the resulting separated proteins. However, advances in mass spectrometry (MS) technology and applications has made the identification of the majority of proteins a practical and financial possibility. The most common identification method used today is peptide-mass fingerprinting (PMF). PMF involves the tryptic digestion of a particular protein yielding a characteristic set of peptide fragments which are unique to that protein. The individual peptides are then ionised and their masses are determined using a mass spectrometer. The resulting pattern of fragments is then compared with a database of in silico generated-tryptic profiles of all known or predicted proteins to identify a match for the protein. Several different MS technologies are in common use but the most accessible and commonly used technology is matrix-assisted laser desorption ionisation - time of flight (MALDI-TOF) MS. MALDI uses laser energy to ionise the tryptic digest fragments which are then accelerated by a powerful electric field through a high vacuum to a detector which registers the impact of each fragment. Ionisation is achieved through the use of a matrix chemical (such as sinapinic acid or di-hydroxy benzoate), in which the peptides are embedded, that absorbs the laser energy and desorbs this energy resulting in the ionisation of neighbouring peptides. The time of flight from ionisation to striking the detector is determined solely by the mass of the fragment. In this way the mass of each fragment can be accurately measured and a tryptic profile for the protein created. Other MS technologies are also used but the underlying principle of accelerating ionised peptides through an electric field to determine their mass remains the same (190).
MS technologies are constantly improving, allowing greater sensitivity and resolution, and along with these improvements in hardware come novel applications for MS. For example, improvements in resolution have allowed the differentiation of individual amino acids and made direct protein sequencing possible.

6.1.1.3 OmpA and IbeB

Both OmpA and IbeB in *E. coli* K1 have been the subject of considerable research in recent years. Both proteins have been implicated in the invasion of the blood-brain barrier and in the case of OmpA a specific receptor has been identified on the surface of BMECs (110, 199, 204). OmpA has been further implicated in almost every stage of NMEC pathogenesis (80, 200, 248, 249). Both proteins also have the potential to impact on the expression of other proteins in *E. coli*. IbeB is allelic to the putative copper transporter CusC and may alter gene expression by modulating intracellular copper concentrations. OmpA is one of the dominant outer membrane protein species and the removal of OmpA from the membrane would be expected to have drastic knock-on effects on the protein composition of the outer membrane. In order to examine the possible indirect roles of IbeB and OmpA isogenic *E. coli* K1 mutants were constructed and the proteome of these mutants was examined by 2-D PAGE.
6.2 Results

6.2.1 Construction of *E. coli* K1 *ompA* and *ibeB* mutants

6.2.1.1 Construction of an *E. coli* RS218 *ibeB* mutant

The *ibeB* gene from *E. coli* K1 strain RS218 was cloned into the high copy pBSKII vector yielding pIbeB and disrupted by insertion of the spectinomycin resistance cassette from plasmid pH45Ω to produce plasmid pAM3 (161)(Table 2.2). The pIbeB plasmid contains a 1871 bp insert containing the 1383 bp *ibeB* gene with 254 bp and 234 bp upstream and downstream of the gene respectively. The pAM3 plasmid consists of pIbeB with the pH45Ω spectinomycin cassette (excised with EcoRI) inserted in an MfeI site 548 bp downstream of the start of the *ibeB* ORF.

The interrupted *ibeB* gene from plasmid pAM3 was amplified by PCR using primers iber and ibef (Table 2.3) and the resulting 1871 kb linear DNA fragment was transformed into *E. coli* RS218 harbouring the induced pKOBEGA plasmid. Following overnight growth spectinomycin resistant colonies were selected and screened for the presence of an interrupted *ibeB* gene using the iber and ibef primers. A single mutant was retained and designated RSibeB (Fig. 6.1 A).

6.2.1.2 Invasion of human T84 cells by *E. coli* K1 RSibeB

As mutations in *ibeB* were found to impair invasion of cultured BMECs by *E. coli* RS218 and reduce virulence in an animal model (110) the invasivity of *E. coli* RSibeB into human T84 cells was determined. The isogenic *ibeB* mutant strain was found to be between 6.5- and 7.6-fold less invasive than the wild type *E. coli* K1 strain RS218 in repeated experiments. To determine if this defect could be complemented the *E. coli* RSibeB strain was transformed with pIbeB or the parental plasmid pBSKII and again compared with the wild type *E. coli* strain RS218. RSibeB harbouring pBSKII gave considerably lower cell association and invasion values than the wild type strain, RS218, and the complemented strain, RSibeB harbouring pIbeB (Fig. 6.1 B and C).
Figure 6.1 Mutation of the *E. coli* K1 *ibeB* gene and the interaction between RSibeB and T84 cells. Panel A: PCR analysis of the *ibeB* gene in *E. coli* K1 strain RS218 (wt), RSibeB (*ibeB* mutant) and pAM3 (mutated *ibeB* on pBSKII plasmid). Panels B and C: Cell association (B) and invasion (C) of cultured T84 cells by *E. coli* K1 strain RS218, RSibeB harbouring pBSKII or RSibeB harbouring plbeB.
This finding indicates that, in addition to its role at the blood-brain barrier, IbeB is also important in the invasion of intestinal epithelial cells, at least in vitro.

### 6.2.1.3 Construction of an *E. coli* RS218 ompA mutant

The *ompA* gene from *E. coli* K1 strain RS218 was cloned into plasmid pBSKII yielding pOmpA, this *ompA* gene was then disrupted by the insertion of a Tn7-based transprimer, conferring kanamycin resistance, and the mutated allele was transferred to the *E. coli* RS218 chromosome using the λ Red allele replacement system (37)(Table 2.2). The *E. coli* RS218 *ompA* mutant was designated RSompA (Table 2.1). Plasmid pOmpA was constructed by PCR amplifying the *E. coli* RS218 *ompA* gene using primers ompAf and ompAr (Table 2.3) and the resulting amplimer was ligated to pBSKII cut with EcoRV. The pOmpA plasmid insert was found to contain a spontaneous deletion removing the native promoter and a small portion of the 5' end of the gene, probably as a result of the toxic effects of over-expressing OmpA from such a high copy number plasmid. The truncated *ompA* gene was then disrupted using a Tn7-based transprimer GPC-linker scanning system generating plasmid pRC1 (Table 2.2). The disrupted allele was amplified by PCR using primers ompAfl and ompArl and transformed into *E. coli* RS218 harbouring the induced pKOBEKA plasmid (37)(Table 2.2). Kanamycin-resistant bacteria were screened by PCR for a mutated *ompA* gene and one such mutant was designated RSompA (Fig. 6.2 A).

### 6.2.1.4 Construction of a low copy OmpA-expressing plasmid

As the pOmpA plasmid has a spontaneous deletion that prevents expression of OmpA, a low copy plasmid was required to allow complementation of the *ompA* mutation in *E. coli* RSompA. The *ompA* gene complete with its own promoter was amplified from purified *E. coli* RS218 genomic DNA using primers ompAf and ompAr and ligated to the low copy plasmid pPD101 cut with EcoRV. Putative clones were screened by PCR using primers T7_promoter and M13_(-43) (Table 2.3), specific to the pPD101 backbone and amplifying across the plasmid’s multiple-cloning site. A single correct plasmid was retained and designated pPDompA11 (Table 2.3). The *E. coli* K1 *ompA* mutant RSompA produced no detectable OmpA and expression was restored in the mutant harbouring the pPDompA11 plasmid (Fig. 6.2 B and C).
6.2.1.5 Invasion of human T84 cells by E. coli K1 RSompA

OmpA has been implicated in the non-phagocytic invasion of macrophage and in the invasion of BMECs (204, 249). An ompA mutant of E. coli RS218 was also less virulent in an animal model of neonatal disease (267). In order to examine if OmpA played any role in the colonisation and invasion of the intestinal epithelium the ability of E. coli RSompA to adhere to invade cultured human T84 cells was compared with that of the wild type strain. Cell-association with E. coli RSompA was found to be 5.3 fold less than the wild type RS218 in T84 cells (RS218 1.06 % ± 0.0299, RSompA 0.202 % ± 0.0387) and the ompA mutant was also found to be 5.8-fold less invasive than the wild type (RS218 0.123 % ± 0.0075, RSompA 0.021 % ± 0.0035). The ompA mutant strain RSompA was then transformed with pPD101 or pPDompA to examine if the invasion defect could be complemented (Fig. 6.3). E. coli RSompA harbouring the pPDompA plasmid was almost 2-fold more invasive than the wild type RS218. However, the ompA mutant harbouring the pPD101 was found to be almost as invasive as the wild type (67 % of wild type invasion). The pPD101 plasmid was previously found to increase the recovery of E. coli K-12 strain XL-1 and K1 strain RShek bacteria in invasion assays using T84 and CHO-K1 cells (Chapter 3, Fig. 3.10 B, Chapter 5, Fig. 5.5 A and B). The mechanism of this apparent increase in invasivity of strains harbouring pPD101 is not known.

6.2.2 Proteomic analysis of E. coli K1

A 2-D PAGE sample preparation protocol was developed (described in Chapter 2) using standard E. coli harvesting, washing and lysis methods (www.expasy.org) combined with a novel protein denaturation and solubilisation procedure. Maximum protein denaturation and solubilisation was achieved using a solution containing the reducing agent DTT, a high concentration of the zwiterionic detergent CHAPS and a mixture of the chaotropes urea and thiourea. CHAPS was included to increase protein solubility and prevent aggregation formation, particularly when loading the sample onto a dehydrated IPG strip (194). Urea is commonly used to denature proteins and the addition of thiourea was recently found to enhance denaturation, particularly of membrane proteins (171). DTT is also a standard addition in most 2-D PAGE protocols.
Figure 6.2 Mutation of the *E. coli K1 ompA* gene. Panel A: PCR analysis of the *ompA* gene in *E. coli K1* strain RS218 (WT), RSompA (Δ*ompA*) and pRC1 (mutated *ompA* in pBSKII). Panels B and C: SDS-PAGE (B) and western immunoblot (C) analysis of OmpA expression in *E. coli K1* strain RS218 (WT), RSompA harbouring pPD101 or RSompA harbouring pPDom pAll. Bands corresponding to the OmpA protein are marked with arrows and protein size standards are indicated with MW.
Figure 6.3 Invasion of human T84 cells by an *E. coli* K1 ompA mutant. Invasion of T84 cells by *E. coli* K1 strain RS218, an isogenic *ompA* mutant RSompA harbouring plasmid pPD101 (OmpA-) or RSompA harbouring pPDompA11 (OmpA+). The level of invasion of each is expressed as a percentage of the wild type levels.
as a reducing agent to break down disulphide bonds and maintain proteins in a fully reduced state. Other reductants such as tributyl phosphine (TBP) and dithioerythritol (DTE) with greater reducing power have become popular recently but the reported benefits do not appear to justify the additional expense and problems with stability and gel artefacts, particularly in the case of TBP (102). Following extensive washes to remove excess salts and contaminating medium components, bacteria were lysed using SDS and stored frozen. Prior to loading the IPG strip the lysed bacterial sample was diluted 1:100 in the solubilisation solution. The charged detergent SDS is normally excluded from the 2-D PAGE sample preparation as its charged nature would interfere with IEF. However, the SDS concentration following dilution in solubilisation solution was low enough so as to not interfere to any noticeable extent. Extended incubation at room temperature was found to increase protein solubility and improve gel resolution. Finally, before IEF, carrier ampholytes were added to samples. Carrier ampholytes are no longer necessary to form the pH gradient in IEF but do help to maintain protein solubility by reducing charge interactions between proteins and substitute for salts as a charge carrier during IEF.

6.2.2.1 Proteomic analysis of E. coli K1 RSihf

As the IHF proteins is involved in the regulation of at least 120 genes in E. coli K-12 and E. coli K1 ihfA (encoding IHF α-subunit) mutant was constructed as a control throughout the development of the E. coli K1 2-D PAGE protocol described above. A ihfA mutation from E. coli strain CJD1484 (Table 2.1) was transduced into E. coli RS218 using a modified Pl phage as described previously (see section 5.2.3.8) resulting in strain RSihf (Table 2.1). Successful mutation of ihfA was confirmed by western immunoblot analysis using a Salmonella IhfA-specific antibody (Fig. 5.13 B). E. coli RSihf and the parental strain RS218 were grown overnight in LB broth and whole cell protein samples were prepared as described. Solubilised protein samples were loaded directly onto dehydrated pH 4-7 IPG strips (Sigma-Aldrich) and isoelectric focussing was carried out as described in chapter 2. The second dimension was carried out using standard 12 % SDS-polyacrylamide gels. Following staining, the gels were scanned using a calibrated densitometer and analysed using the PDQuest™ software package (BioRad) (Fig. 6.4). 23 distinct protein spots were identified whose intensity varied by 4-fold or greater between the two gels. Of these, 13 were upregulated in the ihfA
mutant with the remaining 10 being downregulated. Identification of these proteins by MALDI-TOF will be attempted at a later date.

6.2.2.2 Proteomic analysis of E. coli K1 ompA and ibeB mutants

Both IbeB and OmpA have been implicated in the invasion of the blood-brain barrier by E. coli K1 strain RS218 (110, 204). However, IbeB has been identified as a putative copper transporter in E. coli K-12 (175) and as a result may play a role in the regulation of gene expression by controlling intracellular copper concentration. The reduced invasivity and virulence of E. coli RS218 ibeB mutants could therefore be due to an indirect effect of the mutation on the expression of other virulence genes rather than a direct result of the loss of IbeB expression. Mutation of ompA may also have subsequent effects of gene expression. In addition to its role in the invasion of BMECs a series of additional functions have also been attributed to OmpA including resistance to complement and antibacterial peptides and invasion of and survival within macrophage and neutrophils (80, 200, 248, 249). It is becoming apparent that the outer surface of E. coli is a highly adaptable structure utilising numerous cross-regulation pathways to rapidly alter its composition and function. For example, the binding of type 1 fimbriae to its D-Mannose receptor results in down-regulation of the K1 capsule assembly operon (230) and type 1-mediated adherence to abiotic surfaces results in drastic changes in the protein composition of the outer membrane (187). There has also been regulatory cross-talk reported in the expression of type 1 and P fimbriae (278). As the OmpA protein is present in as many as 100,000 molecules per cell (136, 183) it is conceivable that the removal of such an abundant protein could have subsequent effects on gene expression in general and on the composition of the outer membrane in particular. It is possible that such changes in membrane composition are ultimately responsible for at least some of the phenotypes associated with ompA mutants. In order to examine if mutations in the ibeB and ompA genes had any knock-on effects on gene expression in E. coli K1, a proteomic examination of each of the E. coli K1 mutants and the parental K1 strain RS218 was undertaken.

Bacterial growth conditions were chosen based on those used in the original BMEC invasion assays (110, 204). E. coli RSompA and the parental strain RS218 were grown overnight in BHI broth and E. coli RSibeB and RS218 were grown in LB broth.
Figure 6.4 Proteomic analysis of E. coli RSihf. Comparative 2-D PAGE of mid-logarithmic LB broth cultures of E. coli RS218 and the isogenic \textit{ihfA} mutant, RSihf. Samples were prepared as described and separated on pH 4-7 IPG strips followed by 12% polyacrylamide gels. Protein spots which vary by more than 4-fold are highlighted and their relative abundance is shown below. In each graph the left bar represents the spot intensity from the E. coli RS218 gel.
Figure 6.5 Proteomic analysis of *E. coli* RSibeB. Comparative 2-D PAGE of mid-logarithmic LB broth cultures of *E. coli* RS218 and the isogenic *ibeB* mutant, RSibeB. Samples were prepared as described and separated on pH 4-7 IPG strips followed by 12% polyacrylamide gels. Protein spots which vary by more than 4-fold are highlighted and their relative abundance is shown below. In each graph the left bar represents the spot intensity from the *E. coli* RS218 gel.
Proteomic analysis of *E. coli* RSompA. Comparative 2-D PAGE of mid-logarithmic BHI broth cultures of *E. coli* RS218 and the isogenic *ompA* mutant, RSompA. Samples were analysed as before. Protein spots which vary by more than 4-fold are highlighted and their relative abundance is shown below. In each graph the left bar represents the spot intensity from the *E. coli* RS218 gel.
Solubilised whole-cell protein samples, prepared as described, were separated on pH 4-7 gradient strips (Sigma-Aldrich) in the first dimension and on 12 % SDS-polyacrylamide gels in the second dimension. Each pair of samples was prepared and examined under identical conditions and in parallel. 18 proteins were found to be differentially expressed by greater than 4-fold between \textit{E. coli} RSibeB and the parental strain RS218 (Fig. 6.5). Of these, only 5 were expressed at a higher level in the \textit{ibeB} mutant. However the greatest effect was observed with the \textit{E. coli} K1 \textit{ompA} mutant: 30 proteins were differentially expressed between \textit{E. coli} RSompA and RS218 (Fig. 6.6), of which 17 were upregulated in the \textit{ompA} mutant. These data indicate that mutation of either \textit{ibeB} or \textit{ompA} has dramatic knock-on effects on protein expression in \textit{E. coli} K1. However, without MALDI-TOF identification of these proteins it is impossible to speculate what role, if any, these differentially expressed proteins play in the virulence of \textit{E. coli} K1.
6.3 Discussion

NMEC virulence is a complex, multi-factorial process involving the products of at least 16 identified genes (131, 157). Among these putative virulence factors the IbeB and OmpA proteins have been the subject of considerable research in recent years. Both have been implicated in the invasion of BMECs, with a specific receptor identified in the case of OmpA (110, 199, 204). OmpA has been further implicated in almost every stage of NMEC pathogenesis (80, 200, 248, 249). The data presented here also suggests roles for both proteins in the invasion of human colonic epithelial cells. Mutation of the \textit{ibeB} gene in \textit{E. coli} K1 RS218 considerably reduced the efficiency of invasion into cultured human T84 cells and this defect could be complemented by expressing the \textit{ibeB} gene from a high copy number plasmid. Mutation of \textit{ompA} also reduced invasion by \textit{E. coli} RS218. Complementation of this mutation was complicated by the apparent toxicity of OmpA when expressed from a high copy number plasmid. A low copy number vector system was adopted but resulted in a higher level of background invasion. This effect was also observed in earlier attempts to complement a \textit{hek} gene mutation in \textit{E. coli} RS218 (see chapter 3) although the exact mechanism is unknown. Despite this failure to clearly complement the epithelial cell invasion defect in the \textit{ompA} mutant it is apparent that expression of OmpA is important in the invasion of these cells by \textit{E. coli} K1.

In order to allow proteomic examination of \textit{E. coli} K1 strains a 2-D PAGE protocol was developed using the \textit{E. coli} K1 strain RS218 and an isogenic \textit{ihfA} mutant, RSihf, as controls. The \textit{ihfA} gene, encoding the IHF \(\alpha\)-subunit, was chosen because removal of functional IHF was expected to have a broad range of effects on the expression of a large number of proteins in \textit{E. coli} K1. By comparison with \textit{E. coli} K-12 IHF is assumed to regulate in excess of 100 genes in \textit{E. coli} K1 (6). This assumption was partially confirmed by the discovery of 23 proteins differentially expressed by greater than 4-fold between \textit{E. coli} RS218 and RSihf. The apparent discrepancy between the assumed number of regulated genes and the number of differentially expressed proteins observed can be explained by: the examination of only a single growth condition (mid-log phase cultures in LB broth), a narrow pH range in the first dimension (pH 4-7) and
a relatively insensitive gel staining method (coomassie). Also the proteomic methodology used here is limited to the separation of only 500–1000 proteins. A more thorough investigation of the *E. coli* RSihf proteome would require comparisons of bacteria grown in different media and sampled at different stages of the growth cycle. Samples would then be separated on IPG strips covering the entire pH range; although the majority of *E. coli* proteins do have pIs between pH 4 and 7 a significant number of proteins are missed if only this range is considered. A more sensitive staining method, such as silver staining, could also be employed to increase the detection of low-abundance proteins. Despite these limitations, the findings with *E. coli* RS218 and RSihf do act as a proof of principle for the application of the described 2-D PAGE method to the proteomic analysis of *E. coli* K1.

Interestingly both IbeB and OmpA are expressed by non-virulent *E. coli* K-12. *ibeB* is allelic to the *E. coli* K-12 *cusC* gene, a copper-induced gene encoding a putative copper ion antiporter (175), and the *ompA* gene in *E. coli* K1 is identical to that found in *E. coli* K-12. This suggests that although necessary for virulence neither protein is sufficient to promote invasion of epithelial cells and indeed both proteins may play indirect roles in virulence. It is possible that IbeB impacts on the regulation of other virulence-associated genes through the control of intracellular copper concentration. OmpA has not been directly implicated in gene regulation but the protein does have a large periplasmic domain in addition to its β-barrel membrane-spanning domain (210) and this two-domain structure could conceivably serve as a membrane sensor and signal transduction apparatus. However, even in the absence of a direct regulatory role for OmpA the removal of such an abundant outer membrane protein could have dramatic effects on gene expression in general and the composition of the outer membrane in particular. Regulatory cross-talk has been demonstrated between many components of the outer surface of *E. coli* (230, 278) and the deletion of outer membrane proteins has also been shown to have an impact on the relative abundance of outer membrane components (187). To examine the effect of mutations in IbeB or OmpA the whole-cell proteomes of *E. coli* K1 isogenic *ibeB* or *ompA* mutants were compared with the wild type strain RS218 by 2-D PAGE analysis. Bacteria were grown in media and conditions as closely as possible mimicking the growth conditions described in the literature from the original BMEC invasion assays (110, 204) and 2-D PAGE samples
were prepared according to the protocol developed for the examination of the *E. coli* RSihf proteome. 18 proteins were found to be differentially expressed by more than 4-fold between *E. coli* RS218 and RSibeB of which only 5 were expressed as a greater level in the *ibeB* mutant. This suggests that the *ibeB* protein may indeed play a role in gene regulation and this effect appears to be preferentially through gene activation rather than repression. Mutation of *ompA* was also found to have a dramatic effect on protein expression in *E. coli* RS218 with 30 proteins differentially expressed by greater than 4-fold between *E. coli* RS218 and the isogenic *ompA* mutant RSompA. These 30 proteins were almost evenly divided between up- and down-regulation with 17 being expressed at a higher level in the *ompA* mutant. The identity of these differentially expressed proteins in *E. coli* RSibeB and RSompA is currently unknown but it is hoped that future MALDI-TOF analysis of these proteins will shed some light on their function and possible role in the virulence of NMEC. However, even without positive identifications of these proteins it is clear that mutations in *ibeB* and *ompA* have considerable knock-on effects on the expression of a large number of additional proteins. These findings have significant implications for the study of virulence where gene knockouts and complementation form the cornerstone of research into pathogenicity without regard for the possible knock-on effects of such mutations. It is clear that closer examination of putative virulence factors by proteomic and even genomic methods is necessary before conclusive functions can be accurately assigned. The data presented here comprises a starting point for the proteomic investigation of the roles of the *ibeB* and *ompA* proteins. A considerable volume of additional data will be required to allow a complete understanding of their roles in *E. coli* K1 virulence. As with the *E. coli* RSihf strain different growth conditions must be examined in combination with extended IEF pH ranges, increased staining sensitivity and MS identification of differentially expressed proteins. In the case of *ibeB*, in particular, it is critically important that the effect of copper concentration on the proteome be examined. A closer examination of fractionated protein samples (such as membrane-associated and soluble proteins) would also prove useful.
Chapter 7  General Discussion
The β-barrel conformation is a common structural theme among outer membrane proteins of gram-negative species (136) and many such proteins have been identified as virulence factors (74, 204, 205). A gene with 62 % identity to the enterotoxigenic *E. coli* tia gene was identified in the unfinished genome sequence of *E. coli* K1 strain RS218, a CSF isolate from a neonate with meningitis. The identified gene was found to be homologous to the *hral* gene from porcine *E. coli* and was designated *hek* (for *hral* from *E. coli* K1). The Tia protein, which is predicted to adopt a β-barrel conformation in the outer membrane, has been implicated in the invasion of human epithelial cells by ETEC strain H10407 through interactions with heparinated-proteoglycans (73, 74). Hral is a mannose-insensitive heat-resistant agglutinin initially identified in porcine *E. coli* strain O9:H10:K99 (153) and has been since found to show significant association with virulence in UPEC (243). Tia, Hral, and Hek are all synthesised as a preprotein with an amino-terminal signal sequence, which is cleaved upon translocation across the inner membrane of *E. coli*, yielding the mature protein. Interestingly Hek and Hral share only 17 identical residues in the first 31 amino acids of the preprotein and have only 6 differences across the remainder of the two proteins. In contrast, over the same 31 amino acid stretch Hra1 and Tia have only 2 differences but have 81 differences over all. As the 5' portion of a gene would be expected to have the greatest impact on its expression this finding may indicate that the expression of Hek is controlled differently than both Hral or Tia. Amino-terminal sequencing of the mature Hek protein indicated that the Hek preprotein is cleaved after residue 22, confirming the predicted cleavage point. Computational modelling using the Hek protein primary sequence predicted that the protein would adopt a transmembrane β-barrel conformation, consistent with predictions of the structure of the similar Tia protein (74). The Hek β-barrel was predicted to consist of 8 anti-parallel amphipathic β-sheets linked by 3 short turns on the inner face of the outer membrane and 4 longer (28-33 amino acid) loops on the outer surface.

When the *hek* gene with only 76 bp of upstream DNA was cloned into the high copy number pBSKII plasmid or the low copy pHDI01 plasmid a novel protein of approximately 26 kDa was observed by SDS-PAGE and western immunoblot analysis. Further analysis of this novel 26 kDa protein by peptide-mass fingerprinting positively identified the protein as Hek with an expectation value of 2.14551 x 10^{-17}. The Hek
protein was found to migrate on SDS-polyacrylamide gels as two distinct molecular weight species that were both recognised by an anti-Hek polyclonal antiserum. A similar phenomenon was observed with an MBP-Hek fusion protein and (albeit that this protein resolved in three distinct forms), in the case of the fusion protein, the relative proportions of the different protein species was found to vary under different protein denaturation conditions. This suggested that the distinct species of differing mobilities were in fact Hek in different denatured forms.

When over-expressed in \textit{E. coli} K-12, the Hek protein was found to promote agglutination of human erythrocytes and, as with Hral, this haemagglutination phenotype was resistant to heat treatment. Furthermore, Hek-promoted haemagglutination could be inhibited by the addition of a purified MBP-Hek fusion protein. Hek was also found to promote autoaggregation of bacterial cultures, a phenomenon linked with the formation of biofilms for some proteins (44, 235, 253). Interestingly, Hek-promoted autoaggregation appears to involve interactions between Hek and another bacterial surface component rather than homologous interactions between Hek molecules on neighbouring cells. Autoaggregation has not been reported with either of the Tia or Hral proteins.

As the Hek protein is quite similar to Tia, a known adhesin and invasin, it was suspected that Hek could also play a role in colonisation and perhaps even invasion of human epithelia. To investigate this possibility the T84 cell line was chosen as a model of the human colonic epithelium and qualitative adhesion assays were performed using \textit{E. coli} K-12 expressing either Hek or Tia. Expression of Hek was found to dramatically increase the adhesion of \textit{E. coli} K-12 to T84 cells and promoted the formation of micro-colonies on the surface of cultured cells. Such a phenomenon has been described for EPEC adhesion to epithelia (226) and is perhaps also related to the aggregative adhesion pattern of EAEC to epithelial cells (179). It is likely that Hek-associated autoaggregation plays a role in the formation of these micro-colonies and perhaps represents a bacterial strategy to enhance colonisation of the host by promoting adherence directly to host cells and simultaneously to other bacteria. To further characterise the interaction between Hek-expressing bacteria and epithelial cells, cell-association and gentamicin-protection (invasion) assays were performed using Chinese
hamster ovary (CHO-K1) cells and then repeated in the clinically relevant human colonic T84 cell line. Invasion of CHO-K1 cells by E. coli K-12 was greatly enhanced when either Hek or Tia were expressed. However, when cell-association and invasion assays were performed together it was found that on average greater than 80% of cell-associated Hek-expressing E. coli K-12 were in fact intracellular whereas only 50% of the cell-associated Tia-expressing bacteria were internalised. These data suggest that the Hek protein is in fact a more efficient invasin than Tia. Both Hek and Tia were also found to promote invasion of human T84 cells, although invasion by Tia-expressing E. coli K-12 was 4.5-fold greater than those expressing Hek. The disparity between invasion frequencies with the two cell lines perhaps indicate that Hek and Tia bind different receptors and that these receptors are differentially expressed on T84 and CHO-K1 cells. Furthermore, in T84 invasion assays, the numbers of cell-associated Hek- or Tia-expressing bacteria were considerably higher than the numbers of internalised bacteria, with less than 1 in 20 bacteria successfully gaining access to the host cytosol. It was estimated that overall only approximately 1 in 3,000 T84 cells contained an internalised bacterium. Also, in the qualitative adhesion assays described above, Hek-expressing E. coli K-12 appeared to cluster on individual T84 cells rather than a more diffuse pattern of adherence. Taken together these data suggest that the receptors for Hek and Tia are expressed at considerably higher levels on CHO-K1 cells than on T84 cells and may be only expressed or accessible on a small number of cells in a monolayer. Mutation of the hek gene in E. coli K1 strain RS218 was also found to reduce invasion of T84 cells and expression of the Hek protein from a low copy plasmid could complement this defect.

Proteoglycans are ubiquitous mammalian cell surface molecules consisting of glycosaminoglycans chains (repeats of alternating amino sugars and uronic acids) covalently linked to a protein core and have been identified as receptors for many bacterial and viral pathogens (85, 101, 184, 197, 213). It has been reported in the literature that the Tia protein binds to epithelial cells via interactions with heparinated proteoglycans (73). In order to examine if Hek-promoted adhesion and invasion utilised a similar class of receptor a CHO-K1-derivative cell line, pgsA-745, was used in cell-association and invasion assays. The pgsA-745 cell line has a mutation in the pgsA gene, encoding a xylotransferase enzyme, and is completely deficient in the
production of glycosaminoglycans (GAGs) (70). Invasion of Hek- or Tia-expressing *E. coli* K-12 was greatly reduced in pgsA-745 cells compared with CHO-K1 cells. However, invasion was not completely abolished suggesting that Hek and Tia may recognise some additional mammalian cell surface component or that some residual level of GAG production remains in pgsA-745 cells, although this has not been reported in the literature. The background level of invasion by *E. coli* K-12 harbouring the vector control was also reduced in pgsA-745 cells indicating that *E. coli* K-12 perhaps has some inherent GAG-binding ability. To establish if Hek bound a specific GAG moiety, a series of invasion assays were carried out using CHO-K1 cells in which purified heparin, heparan sulphate or chondroitin sulphate were included as potential inhibitors of invasion. Both chondroitin and heparin GAG moieties are commonly found on mammalian epithelial cell proteoglycans. The non-related polyanionic sugar dextran sulphate was included as a control as it has a similar mass:charge ratio as the highly-sulphated GAG heparin. Invasion of CHO-K1 cells by *E. coli* K-12 expressing Hek was inhibited by heparin in a sensitive and dose-dependent manner. Invasion could also be inhibited with heparan sulphate at high concentrations (100 μg/ml) but when inhibition assays were repeated with 10-fold less of each inhibitor only heparin was found to inhibit Hek-promoted invasion. This finding indicates that Hek binds more avidly to the more sulphated (and more negatively charged) heparin. The lack of inhibition with the highly charged dextran sulphate also rules out a non-specific charge-charge interaction between Hek and its receptor. These findings make Hek only the second heparin-binding protein to be identified in *E. coli*, the first being the Tia protein from ETEC (73).

Further work will be required to characterise the Hek receptor in greater detail. Using currently available techniques and reagents it should be possible to conclusively identify this receptor in time. Among the most likely candidates are the syndecan family of surface proteoglycans. Studies with syndecan-deficient cell lines or using knockdown techniques may be used to examine this possibility. Purified syndecans and monoclonal antibodies are also now available for each of the four members of the syndecan family allowing *in vitro* binding and inhibition assays.
As the four external loops of Hek are the only surface exposed portion of the protein it is a reasonable assumption that the biological activity of the protein resides in one or more of these surface structures. A series of four mutated *hek* genes were constructed on high copy number plasmids using inverse PCR, each lacking the coding sequence for a single surface exposed loop. The expression of each Hek loop deletion mutant was confirmed by western immunoblotting and surface exposure was detected by flow-cytometry. Deletion of the first loop was found to have no effect on any of the Hek promoted phenotypes. The loop 4 deletion gave intermediate haemagglutination and autoaggregation phenotypes but actually appeared to slightly enhance invasion of CHO-K1 cells. The loop 3 deletion mutant protein was poorly expressed and displayed almost undetectable haemagglutination and autoaggregation. However, despite its poor expression, bacteria expressing the loop 3 mutant protein still efficiently invaded CHO-K1 cells. Only mutation of loop 2 was found to completely abolish all of the Hek-promoted phenotypes. These data indicate that the interaction between Hek and its mammalian or bacterial surface receptors is mediated exclusively by the second surface exposed loop of the protein. The loop 2 mutation removed 87 bp (the coding sequence for 29 amino acids - FYARGKADSKYNYVDSWSSGYYWRDDLKN) and replaced it with CGCGCG (encoding the turn-promoting AR residues). The deleted region contained a large number of charged amino acids (6 positive and 5 negatively-charged residues), however each of the remaining surface exposed loops contained a similarly high proportion of charged residues. This observation, combined with the lack of inhibition with the highly charged dextran sulphate, appears to discount the possibility of charge-charge interactions between Hek and its receptor. Loop 2 may contain an as-yet undefined heparin-specific binding sequence. This specific interaction between Hek and its receptor may then be stabilised by non-specific interactions between the receptor and the highly charged surface of the Hek protein. A number of different mutagenesis techniques could be employed to further characterise the binding ability of Hek loop 2. A complete series of loop 2 single amino acid deletions or substitutions would be a first step in identifying important residues, perhaps starting with the 14 residues that are conserved between Hek and Tia in this region.

In order to implicate Hek in NMEC pathogenesis it was necessary to demonstrate that the *E. coli* K1 clinical isolate RS218 expressed the protein. Standard western
immunoblotting techniques failed to detect Hek expression in RS218 and even expression from a high copy number plasmid appeared to be sporadic. An enhanced sensitivity western immunoblot system succeeded in detecting chromosomally-expressed Hek protein but revealed a disparity in the level of expression from the chromosomal hek allele and a low copy number plasmid. The difference in the level of protein expression did not correlate with the difference in gene copy number. Expression from the plasmid-borne hek gene appeared greater than would be expected based on Hek expression in *E. coli* RS218. This observation lead to the cloning of the hek gene with additional upstream DNA that was found to restore Hek expression to levels which correlated with chromosomally expressed protein. The additional upstream DNA was initially thought to contain some regulatory sequences that served to repress Hek expression. This theory was partially supported by data from hek-phoA fusions with and without the additional upstream DNA. However, when the transcription of the hek gene was examined using a lacZ reporter plasmid the upstream DNA was found to have the exact opposite effect to that expected. Addition of the putative regulatory region actually resulted in increased transcription. These seemingly contradictory findings were reconciled in a model of Hek expression which proposes the existence of a second promoter upstream of the hek gene within the extended regulatory region (Fig. 5.14). In order to reconcile these results the second promoter, P_X, would be stronger than the promoter proximal to the hek gene, P_hec, but produce an mRNA transcript which was less well translated. This could be due to additional secondary structure in the longer mRNA, perhaps sequestering the hek Shine-Delgarno sequence, or even the binding of a regulatory protein to the longer transcript. The effect of IHF on hek transcription also appears to support this model. Examination of hek transcription in an ihfA mutant showed an inhibitory effect by IHF on hek transcription only in absence of the upstream regulatory region. According to the two-promoter model IHF could sterically inhibit transcription of hek from P_hec, by binding to sites adjacent to the promoter, but IHF-binding at these sites would not affect transcription from the second upstream promoter.

In addition to this seemingly complex control of Hek expression the expression of the protein also appeared to be slightly variable. Expression of Hek in *E. coli* K1 from a strong inducible promoter appeared to vary between two distinct levels and this finding was supported by experiments using a hek-phoA reporter fusion expressed from the
native or an inducible promoter. To examine if this variability was completely independent of the native promoter the transcription of the hek gene was also observed using the lacZ reporter system. Transcription was also found to be variable and this variability was apparent from plasmids containing either the putative P$_{hek}$ promoter alone or the extended upstream region. DNA sequencing of inserts in hek-lacZ fusion plasmids revealed no sequence changes which correlated with variability in transcription. It is conceivable that the observed variability is in fact due to variation in the expression of a regulatory protein. Direct examination of chromosomally encoded hek transcription and Hek protein expression may shed further light on this phenomenon. As Hek expression in *E. coli* RS218 is at the limit of detection for the previously used western immunoblot techniques a refined protocol or a novel technique would be required to accurately quantify exogenously expressed protein. Transcription of hek, however, would be considerably easier to monitor using RT-PCR technology.

The effects of mutations in known *E. coli* K1 virulence factors on human colonic epithelial cell invasion were also investigated. Mutations in either *ibeB* or *ompA* were found to reduce the invasion efficiency of *E. coli* RS218 and these mutations could be complemented. As both of these genes have been implicated in multiple stages of NMEC pathogenesis it was proposed that both could play indirect roles by impacting on the expression of additional genes in *E. coli* K1. The IbeB protein is a suspected copper transporter (175) and OmpA is one of the most abundant *E. coli* proteins and is the major protein constituent of the outer membrane (136, 183). A 2-D PAGE protocol was developed for *E. coli* K1 using an *E. coli* RS218 ihfA mutant which was expected to display altered expression for a large number of proteins. The proteomes of *E. coli* RS218 and isogenic *ibeB* or *ompA* mutants grown under BMEC invasion assay conditions were then compared by 2-D PAGE. Mutation of either *ibeB* or *ompA* was found to alter the expression of a large number of proteins (18 and 30 proteins respectively). Although the identities of these differentially expressed proteins is as yet unknown it is clear that such isogenic mutations have a surprisingly severe impact on *E. coli* protein expression. These findings highlight the necessity of proteomic techniques in any study of bacterial pathogenesis.
Figure 7.1 Model of GI colonisation and invasion by *E. coli* K1. Magnified view (A) and overview (B) of the proposed mechanism of interaction between *E. coli* K1 and the neonatal GI epithelium with the possible role of the Hek protein highlighted.
In conclusion, this work has identified a new *E. coli* K1 virulence factor, Hek, which is sufficient to promote invasion of cultured epithelial cells by non-invasive *E. coli* K-12. Hek is similar to the Hra1 protein in porcine *E. coli* and the Tia adhesin/invasin in ETEC and has homologues in UPEC and EAEC strains. The protein is predicted to adopt an 8-stranded β-barrel conformation in the *E. coli* outer membrane. Hek was found to interact with heparinated-proteoglycans on mammalian epithelial cells and this interaction requires the second surface-exposed loop of the Hek β-barrel. Based on these findings it is now possible to propose a model for the initial stages of NMEC pathogenesis which includes contributions by type 1 fimbriae, the K1 capsule and the Hek protein (Fig. 7.1). The K1 capsule surrounding *E. coli* K1 provides resistance to environmental stresses and the host immune response and thus facilitates colonisation of the neonatal GI tract. However, expression of the capsule is likely to mask short adhesins such as the Hek protein. Initial adherence of *E. coli* K1 to the GI epithelium may involve long-range interactions between the type 1 fimbrial tip, FimH, and mannose-containing receptors on the host cell surface. Binding of this fimbriae has been shown to result in down-regulation of the capsular assembly genes and reduction or even complete loss of the K1 capsule (230). The interaction between FimH and its receptor may also induce changes in the protein composition of the bacterial outer membrane (187) including a decrease in OmpA expression with possible subsequent knock-on effects on the expression of many other proteins. Initial attachment mediated by type 1 fimbriae and down-regulation of the capsule may then allow more intimate attachment to occur, mediated perhaps by Hek binding to heparinated-proteoglycans such as syndecan. Hek-promoted attachment could then induce signal-transduction events resulting in bacterial uptake. A considerable amount of additional data would be required to confirm aspects of this model, including conclusive identification of the Hek receptor and the elucidation of a molecular basis of Hek-mediated invasion. NMEC pathogenesis is a complex process and it is likely that several other bacterial factors also contribute to this colonisation and invasion process. However, the ability of Hek to promote invasion in non-pathogenic strains of *E. coli* and the presence of the *hek* gene in other pathogenic *E. coli* lineages, including EAEC and UPEC strains, indicates that the Hek protein is an important *E. coli* virulence factor.
References


151


